Printed in U.S.A.
Vol. 111, No. 5

Article

PUNISHING ON A CURVE

Adi Leibovitch

ABSTRACT—Does the punishment of one defendant depend on how she fares in comparison to the other defendants on the judge's docket? This Article demonstrates that the troubling answer is yes. Judges sentence a given offense more harshly when their caseloads contain relatively milder offenses and more leniently when their caseloads contain more serious crimes. I call this phenomenon "punishing on a curve."

Consequently, this Article shows how such relative sentencing patterns put into question the prevailing practice of establishing specialized courts and courts of limited jurisdiction. Because judges punish on a curve, a court's jurisdictional scope systematically shapes sentencing outcomes. Courts of limited jurisdiction usually specialize in relatively less serious crimes—such as misdemeanors, drug offenses, or juvenile cases. They treat the mild offenses on their docket more harshly than generalist courts that also see severe crimes. This leads to the disturbing effect of increasing punitive outcomes vis-à-vis these offenses, wholly contradictory to the missions of these courts. Such sentencing patterns undermine notions of justice and equitable treatment. They also undermine retributive principles and marginal deterrence across crimes of increasing severity.

In light of the profound normative and practical implications, this Article proposes a remedy to standardize sentences through "curving discretion." In addition to consulting the sentencing range recommended by the sentencing guidelines for a particular offense, a judge should see the distribution of sentences for the same offense across different courts. This Article illustrates the feasibility of this proposal empirically using sentencing data from neighboring judicial districts in Pennsylvania. It also explains how this proposal fits within the Supreme Court's jurisprudence following *United States v. Booker*, which rendered sentencing guidelines advisory, and its potential advantage to improve appellate review.

AUTHOR—Associate Professor, Hebrew University of Jerusalem, Faculty of Law. For helpful comments and discussions, I thank Ronen Avraham, Omri Ben-Shahar, Tony Casey, Zohar Goshen, Jamal Greene, Sasha Greenawalt, Bernard Harcourt, Bert Huang, William Hubbard, Avery Katz, Ryan Liss, Anna Lvovsky, Richard McAdams, Ed Morrison, Ariel Porat,

Eric Posner, J.J. Prescott, John Rappaport, Gabriel Rauterberg, Daniel Richman, Elizabeth Scott, Matthew Stephenson, Marah Stith McLeod, Eric Talley, and workshop participants at the University of Chicago Law School, Columbia Law School, George Mason University Antonin Scalia Law School, Georgetown University Law Center, Bar-Ilan University Faculty of Law, Haifa University Faculty of Law, the Hebrew University of Jerusalem Faculty of Law, and Tel Aviv University Buchmann Faculty of Law. I also thank the judges and court administrators at the first and second municipal district courts of Cook County, Illinois, for welcoming me into their courts.

INTRODUCTION

Assigning cases across courts and judges is commonly viewed as an administrative decision. Yet a comparative analysis of criminal courts reveals an astounding list of unintended consequences on the substance of legal decisions. Juvenile courts, for example, were established with an eye toward rehabilitation of adolescents, but from the Progressive Era through the twentieth century, these courts have often adopted more punitive sentencing and pretrial-release practices than those used by adult criminal courts.[1] When misdemeanor and felony cases are assigned to separate

[1] *See infra* Section II.B.2. *See generally* LAWRENCE BAUM, SPECIALIZING THE COURTS 106–09 (2011) (discussing the history of the juvenile courts); Jeffrey Fagan, *Separating the Men from the Boys:*

divisions, the sentencing levels for misdemeanors sometimes exceed those of felonies involving similar conduct.[2] With the division of geographic jurisdiction, not only do courts tend to develop different sentencing practices but a long line of research suggests they also follow a particular pattern: courts in rural and suburban areas tend to develop harsher sentencing practices than their counterparts in large urban locations.[3] These observations have so far been regarded as unrelated. This Article argues they are not.

This Article demonstrates that judicial decisions are based on a relative ranking of "blameworthiness"; judges compare the cases before them to others on their dockets. Judicial evaluation of blameworthiness, in other words, involves a relative judgment not an independent one. The cases on each judge's docket form a particular "punishment curve" that serves as a benchmark for the sentencing decision in an individual case. Consequently, judges handling different dockets end up using different comparison groups as the benchmarks for their decisions so that particular cases are evaluated against different punishment curves. This phenomenon leads not only to troubling sentencing disparities but also to predictable and systemic biases across courts of different jurisdictions.

The positive claim—which I support with quantitative and qualitative evidence—is that judges who are exposed only to relatively mild offenses (for example, misdemeanors) or offenders (for example, juveniles) tend to see those cases as relatively worse than they otherwise would. As a result, they may develop harsher sentencing practices than judges exposed to the full spectrum of criminal cases.

This means that the decision about the types of cases before particular courts or judges is not merely administrative; it can affect substantive case outcomes. Where a court of limited jurisdiction is established as part of a broader mission to alter the treatment of certain offenses or offenders, it can even undermine that mission by nudging courts in exactly the opposite

---

The Comparative Advantage of Juvenile Versus Criminal Court Sanctions on Recidivism Among Adolescent Felony Offenders, *in* SERIOUS, VIOLENT & CHRONIC JUVENILE OFFENDERS 238, 244 (James C. Howell et al. eds., 1995) (reviewing the competing findings comparing juveniles sentenced in juvenile and adult courts); James C. Howell, *Juvenile Transfers to the Criminal Justice System: State of the Art*, 18 LAW & POL'Y 17, 17 (1996) (same).

[2] This happened, for example, with the sentencing of domestic violence offenses in Chicago in 2010. *See infra* Section II.B.1.

[3] *See, e.g.*, JAMES EISENSTEIN ET AL., THE CONTOURS OF JUSTICE: COMMUNITIES AND THEIR COURTS 278 (1988) ("[S]entences in larger jurisdictions will be less severe."); Jeffery T. Ulmer & Brian Johnson, *Sentencing in Context: A Multilevel Analysis*, 42 CRIMINOLOGY 137, 168 (2004) ("[C]ourt size produces distinctive sentencing patterns, with large urban courts exhibiting the most lenient sentencing."); *see also infra* Section II.B.3.

of the intended direction. Courts that were designed to emphasize treatment and rehabilitation might be more punitive.[4] Attempts to get tougher on crime can lead to unexpected leniency.[5] Sometimes even more general policy goals, such as proportionality and marginal deterrence, are compromised.[6]

In making the positive claim about judges' "relative judgments," this Article builds on two previously unrelated strands of literature. It connects (1) the experimental findings of social psychologists and behavioral scholars who have analyzed the effects of context dependence on decisionmaking with (2) the qualitative and quantitative field evidence provided by sociologists, criminologists, and political scientists who have documented sentencing trends and disparities across courts of different jurisdictions. By bridging the two, this Article departs from the current discourse in an important way. While each strand of the literature, separately, has regarded the phenomena documented as arbitrary or ad hoc, this Article identifies the potential for a systemic bias in a predicted direction.

Behavioral law and economics scholars have repeatedly documented in experimental settings how, due to the contrast effect, "[t]he availability of comparison cases apparently makes a serious case appear more serious than it would on its own and makes a milder case appear milder."[7] But their focus has been on how such context dependence leads to "arbitrary," "erratic," and "unpredictable" judgments, mostly among jurors and laypersons.[8] Their underlying assumption is that the absence of context causes jurors' decisions to be erratic, which provides "a possible reason to favor judicial decisions over jury decisions, because judges are more likely to have a menu of cases before them."[9] Judges' wider perspective does

---

[4] *See infra* Section II.B.2.

[5] *See infra* Section II.B.1.

[6] *See infra* Sections II.B.1–2.

[7] Cass R. Sunstein et al., *Assessing Punitive Damages (with Notes on Cognition and Valuation in Law)*, 107 YALE L.J. 2071, 2104 (1998) [hereinafter Sunstein et al., *Assessing Punitive Damages*]; *see also* Daniel Kahneman et al., *Shared Outrage and Erratic Awards: The Psychology of Punitive Damages*, 16 J. RISK & UNCERTAINTY 49, 59 (1998); Jeffrey J. Rachlinski & Forest Jourden, *The Cognitive Components of Punishment*, 88 CORNELL L. REV. 457, 477 (2003); Cass R. Sunstein et al., *Predictably Incoherent Judgments*, 54 STAN. L. REV. 1153, 1176 (2002) [hereinafter Sunstein et al., *Predictably Incoherent Judgments*]; *infra* Section I.B.2.a.

[8] Kahneman et al., *supra* note 7, at 75; *see also* Sunstein et al., *Assessing Punitive Damages*, *supra* note 7, at 2142.

[9] Sunstein et al., *Predictably Incoherent Judgments*, *supra* note 7, at 1156–57; *see also* Sunstein et al., *Assessing Punitive Damages*, *supra* note 7, at 2079 (suggesting that "the legal system should provide a mechanism by which judges or administrators, instead of jurors, can translate the relevant moral judgments into dollar amounts").

mitigate the problem of erratic decisionmaking yet it suffers from a separate problem worthy of attention. Because of their wider perspective, judges' decisions can be tainted by comparison of a particular case to the other cases in their caseloads. In previous work, I demonstrated this claim empirically, showing that judges who were at first randomly exposed to milder offenses developed harsher sentencing practices than judges initially exposed to offenses that are more serious.[10] In this Article, I build on those findings to develop a theory of how the institutional design of courts and the rules for case assignment can lead to a systemic bias in legal decisions in a predicted direction.

Similarly, sociologists, criminologists, and political scientists have offered varying ad hoc explanations of documented differences across courts of differing jurisdictions. Differences in sentencing levels between specialized and generalist courts have been attributed to different perceptions by generalist judges of the uniqueness of cases from the specialized subject matter,[11] political capture of the specialized courts by high-stakes constituents,[12] or selection of judges with known policy and punitive preferences to staff such courts.[13] The divergence between the intended purposes and unintended consequences of some specialized courts have also been explained by such factors as well as by the oversight of

[10] Adi Leibovitch, *Relative Judgments*, 45 J. LEGAL STUD. 281 (2016). I discuss the empirical findings in greater detail in Section I.B.2.c.

[11] BAUM, *supra* note 1, at 35–36, explains how immersion in a subject matter can shape judicial policy through increased confidence—and subsequently greater willingness to overturn decisions and change policy—but also exposes judges to insularity, making them susceptible to biases and stereotypes about particular types of cases. The concentration of offenses in a specialized docket can also signal to judges "that the offense should be taken seriously" and lead to increased punishments. *Id.* at 102.

[12] Interests groups can shape the policy developed in specialized courts in several ways: they can promote the establishment of specialized courts emphasizing a particular policy orientation, *id.* at 133, they can affect the selection of judges to such courts, and they can influence sitting judges to develop the law a certain way thanks to their advantage as repeat players in the courts, *id.* at 37–39.

[13] Judges who staff juvenile, drug, and mental health courts, for example, are often chosen based on their sympathy toward and enthusiasm about the court's mission. *Id.* at 37–38, 40, 221. When policymakers aim to get tougher on crime, other examples are more flagrant. In Philadelphia, the first city to establish specialized dockets for homicide cases in the 1970s, "[p]rosecutors have been able to knock lenient judges out of the program and steer the most serious cases to a handful of those most sympathetic to their charges." Tina Rosenberg, *The Deadliest D.A.*, N.Y. TIMES MAG. (July 16, 1995), http://www.nytimes.com/1995/07/16/magazine/the-deadliest-da.html? [https://perma.cc/S2F4-MJRL]. Chicago's Speeders' Court "was assigned a judge who did not own an automobile and who was not owned by the owner of any automobile." Herbert Harley, *Business Management for the Courts*, 5 VA. L. REV. 1, 11 (1917). The first gun court, established in Providence in 1994, was "staffed by a judge who had once sentenced a defendant to death despite the absence of capital punishment in Rhode Island," BAUM, *supra* note 1, at 220, and received the nickname "Maximum John," *Gun Court Triggers Delight in Rhode Island*, DESERET NEWS (Dec. 9, 1994), http://www.deseretnews.com/article/392266/GUN-COURT-TRIGGERS-DELIGHT-IN-RHODE-ISLAND.html [https://perma.cc/XJX5-LWJU].

policymakers—designing such courts "on the basis of . . . folk theories, commonsense notions that do not fully accord with reality."[14] Across district courthouses, sentencing patterns have been explained by a long list of factors,[15] including caseload pressure, attitudes toward plea bargaining, local perceptions of crime, courtroom organization and resources, and court communities' ties and norms.[16] This Article highlights the pattern that emerges from these studies—a negative relationship between sentencing outcomes and the scope of criminal severity within a court's jurisdiction—as well as how relative-judgments bias therefore offers an additional, unifying explanation to their findings. An explanation that, through its greater generalizability, entails not only descriptive but predictive force.

From the institutional design perspective, the anomalous sentencing patterns that result from relative judgments highlight the importance of accounting for the potential unintended consequences of caseload exposure when assigning cases and responsibilities across courts. More generally, they call attention to the role of sentencing guidelines in promoting sentencing uniformity and to the limitations that impede them from achieving that goal. For more than three decades, sentencing guidelines across jurisdictions have not substantially changed in form—classifying offenses by recommending ex ante ranges for sentences. Yet with the advancement of and increased reliance on computerized systems in the courts to calculate sentencing recommendations and with the collection of sentencing information by sentencing commissions statewide and nationwide, the time is ripe to rethink how sentencing matrices and recommendations are developed.

This Article proposes to standardize sentences by complementing existing sentencing guidelines with "statistical curving": in addition to consulting the sentencing range recommended by the sentencing guidelines for a particular offense, a judge should see the distribution of sentences for the same offense across different courts. Displaying the overall punishment

---

[14] BAUM, *supra* note 1, at 5.

[15] For an overview of the literature, see Brian D. Johnson, *Contextual Disparities in Guidelines Departures: Courtroom Social Contexts, Guidelines Compliance, and Extralegal Disparities in Criminal Sentencing*, 43 CRIMINOLOGY 761 (2005). *See also infra* Section II.B.3.

[16] EISENSTEIN ET AL., *supra* note 3, at 768 (describing "court communities" as manifested in the interdependencies of court personnel; power and status relations among judges on the court; local customs; and the characteristics of the court environment, including perceptions in the county, type of legal representation, demographics, and socioeconomic parameters); JEFFERY T. ULMER, SOCIAL WORLDS OF SENTENCING: COURT COMMUNITIES UNDER SENTENCING GUIDELINES (1997) (defining the "distinctive social worlds" of different courts based on participants' shared workplace, interdependent working relations between key sponsoring agencies—the prosecutor's office, judges, and the defense bar—and distinctive legal and organizational cultures, such as norms regarding guidelines compliance or rates of substantial assistance departures).

curve would help judges conform to general sentencing practices in a way that would increase consistency, granularity, and uniformity in sentencing. This kind of statistical curving of discretion also offers several additional benefits. Because such statistical curving would serve as advisory guidance based on judicial decisions themselves, it would not impose external rigidity. Because it is informative about actual sentencing practices, it would facilitate appellate review. Because it is based on actual sentences imposed by judges, it may also answer the concerns that are sometimes raised about the institutional roles of legislators, judges, and prosecutors in shaping sentencing levels.

By developing a theory of relative judgments based on different punishment curves, this Article makes three main contributions. First, it identifies a neglected mechanism for why and how sentencing practices are formed through the composition of judicial dockets. Second, by connecting the findings of court scholars under one unifying conceptual framework, this Article develops a prescriptive account of the connection between institutional capacity and policy outcomes in the criminal justice system. Lastly, this Article identifies the limited exposure from caseloads as the source of the bias and offers a solution to standardize sentencing over a joint curve.

This Article also contributes to the broader sentencing guidelines discourse by refocusing on the commitment to uniformity of sentencing. For more than three decades, debates around sentencing guidelines have revolved mainly around the guidelines' rigidity and the resulting harshness of sentencing outcomes.[17] Under the classic account, relaxing the rigidity of sentencing guidelines and increasing judicial discretion sacrifices some uniformity for the sake of a greater emphasis on individuality in

[17] *Compare, e.g.*, Frank O. Bowman, III, *Dead Law Walking: The Surprising Tenacity of the Federal Sentencing Guidelines*, 51 HOUS. L. REV. 1227, 1230 (2014) [hereinafter Bowman, *Dead Law Walking*] (describing how the federal sentencing regime founded by the Sentencing Reform Act of 1984 was "widely criticized by the defense bar and many in the judiciary and the academy for being unduly inflexible and unremittingly harsh"), Frank O. Bowman, III, *The Failure of the Federal Sentencing Guidelines: A Structural Analysis*, 105 COLUM. L. REV. 1315, 1328 (2005) [hereinafter Bowman, *Failure of the Federal Sentencing Guidelines*] ("At or near the root of virtually every serious criticism of the guidelines is the concern that they are too harsh . . . ."), *and* David Yellen, *Reforming the Federal Sentencing Guidelines' Misguided Approach to Real-Offense Sentencing*, 58 STAN. L. REV. 267, 268 (2005) (describing the Federal Sentencing Guidelines as "overly rigid and complex"), *with* Paul G. Cassell, *Too Severe?: A Defense of the Federal Sentencing Guidelines (and a Critique of Federal Mandatory Minimums)*, 56 STAN. L. REV. 1017, 1017–18 (2004) (arguing that the Guidelines themselves are not too severe but rather that the problem is with statutes prescribing mandatory minimum sentences and preventing downward departures), *and* Thomas N. Whiteside, *Reality of Federal Sentencing: Beyond the Criticism*, 91 NW. U. L. REV. 1574, 1577, 1591 (1996) (arguing that the severity of the Guidelines mostly results from federal statutes and that the Guidelines have "maintained significant judicial discretion over sentences").

sentencing.[18] The bias identified by this Article, however, undermines such a dichotomous account. When the judicial decision in a particular case is affected by the characteristics of the *other* cases in a judge's caseload, the outcome reflects neither uniformity nor individuality in sentencing.

This concern is intensified by a particular design feature of specialized courts that has mostly escaped scholarly attention: courts of limited jurisdiction (such as drug courts, veterans courts, or juvenile courts) are usually specialized around categories of relatively *less* serious crimes and their jurisdiction is often restricted to misdemeanors or low-class felonies.[19] If exposure to less serious offenses leads judges to view particular cases as relatively worse and sentence them more harshly, then lack of uniformity is also inevitably tied to increased punitiveness by the very design of the compartmentalized criminal justice system.

The theory presented in this Article may be applied beyond criminal sentencing decisions. It is relevant to the decisionmaking processes of other actors in the criminal justice system—from police officers to prosecutors to parole boards. It is relevant to other judicial decisions in criminal cases such as those about pretrial confinement and setting bail. It is relevant to judicial decisions in civil cases as well as decisions by administrative

---

[18] *See, e.g.*, Albert W. Alschuler, *The Failure of Sentencing Guidelines: A Plea for Less Aggregation*, 58 U. CHI. L. REV. 901 (1991); Daniel J. Freed, *Federal Sentencing in the Wake of Guidelines: Unacceptable Limits on the Discretion of Sentencers*, 101 YALE L.J. 1681, 1703 (1992); Stephen J. Schulhofer, *Assessing the Federal Sentencing Process: The Problem is Uniformity, Not Disparity*, 29 AM. CRIM. L. REV. 833 (1992).

[19] Many states, for example, assign misdemeanor cases to particular judges or to separate municipal or criminal courts of limited jurisdiction. The National Center for State Courts (NCSC) offers comprehensive charts summarizing the court structures in all U.S. states. COURT STATISTICS PROJECT, NAT'L CTR. FOR STATE COURTS, STATE COURT STRUCTURE CHARTS (2013), http://www.courtstatistics.org/Other-Pages/State_Court_Structure_Charts.aspx [https://perma.cc/2YPM-NCDX]; *see also infra* notes 106–09 and accompanying text. Domestic violence courts are often limited to hearing only misdemeanors or Class-D felonies. *See, e.g.*, GENERAL ORDERS OF THE CIRCUIT COURT OF COOK COUNTY, ILLINOIS, Nos. 1.2, 2.1(g) (Sept. 29, 2014), http://www.cookcountycourt.org/Manage/DivisionOrders/ViewDivisionOrder/tabid/298/ArticleId/188/GENERAL-ORDER-NO-1-2-2-1-County-Department.aspx [https://perma.cc/PK9X-3LDZ]; *see also infra* notes 120, 124–25 and accompanying text (discussing Illinois). Juvenile courts handle adolescents who are viewed as lesser offenders usually with no prior criminal record, and the most serious felonies are commonly excluded, either mandatorily or presumptively, from the jurisdiction of the juvenile court and transferred to the adult criminal court. PATRICK GRIFFIN ET AL., U.S. DEP'T OF JUSTICE, OFFICE OF JUVENILE JUSTICE & DELINQUENCY PREVENTION, TRYING JUVENILES AS ADULTS (2011), https://www.ncjrs.gov/pdffiles1/ojjdp/232434.pdf [https://perma.cc/2D49-E4XJ]. Veterans courts focus mostly on nonviolent and largely misdemeanor offenses. Michael Daly Hawkins, *Coming Home: Accommodating the Special Needs of Military Veterans to the Criminal Justice System*, 7 OHIO ST. J. CRIM. L. 563, 564, 566, 571 (2010). Many drug courts exclude defendants charged with felonies or violent crimes. NAT'L ASS'N OF CRIMINAL DEF. LAWYERS, AMERICA'S PROBLEM-SOLVING COURTS: THE CRIMINAL COSTS OF TREATMENT AND THE CASE FOR REFORM 22 (2009), https://www.nacdl.org/WorkArea/DownloadAsset.aspx?id=20217 [https://perma.cc/NF2W-VDVD].

agencies. In some of these realms, there is qualitative evidence for such patterns, and I reference that evidence when relevant. But for the coherence of the argument, and because certain features of the different processes are worthy of particular attention, I focus this Article on judicial decisions of sentences.

This Article proceeds in four parts. Part I builds on psychological theory to develop a theory of why judges sentence cases relative to their other cases. Part II demonstrates how the problem of relative judgments can lead to unintended consequences on sentencing outcomes and offers evidence of such trends in sentencing from three types of jurisdictional assignments: (1) the assignment of misdemeanor and felony cases under a unified or a two-tiered model, (2) the institution of specialized courts dealing with particular types of offenses (such as domestic violence courts) or offenders (such as juvenile courts), and (3) the division of geographic jurisdiction across district courthouses.

Parts III and IV then discuss ways to address the problem of punishing on a curve. Part III draws lessons for the institutional design of courts of general, limited, and hybrid jurisdictions. It explains how the concern of relative-judgments bias can tilt the scale toward preferring generalist court models where the underlying reasons for case assignments are administrative—as is for the division between misdemeanor and felony cases. But it also acknowledges that unification is often impractical for other forms of limited jurisdiction, like geographic jurisdiction, and can sometimes contradict the substantive benefits that specialized courts seek to achieve—especially in courts that are focused around a particular population of offenders or victims. Part IV offers a more general remedy: it proposes to create a general punishment curve by using information about the actual sentences imposed for offenses across judges and courts. Judges could then standardize the particular sentences they impose in accordance with the overall punishment curve.

I. THE THEORY OF RELATIVE JUDGMENTS

A. The Normative Role of Relative Severity

Judicial decisions are prone to contextual influences because of the inherent relativity in the assessment of cases. Judicial decisionmaking requires comparing cases based on the gravity of the behavior, the circumstances surrounding it, the harm created by the illicit act, and the characteristics of the offender. The evaluation of all of these aspects can be affected by their existence and magnitudes in the relevant comparison group.

For criminal law in particular, relative severity has a central doctrinal role—the principle of proportionality. In its most basic form, it is the retributive command to let the punishment fit the crime by punishing criminals in proportion to the gravity of their acts. Punitive severity must accord with the severity of the crime.[20] Deterrence theory focuses on the magnitude of social harm based on utilitarian justifications.[21] Efficient deterrence requires specific and general deterrence relative to the costs and benefits associated with the act[22] as well as marginal deterrence between acts of increasing severity.[23] The higher the social harm imposed by the offense, the higher the penalty should be.[24]

When the sanction involves incarceration, incapacitation is not free of comparative assessments. Incapacitation in the real world, as the United States' mass incarceration crisis illuminates, is costly to administer and is bounded by the capacity of jail and prison facilities.[25] Under a given resource constraint, even an incapacitation rationale must differentiate across offenders based on the relative danger they pose, the harm inflicted by their offenses, and their likelihood of reoffending—or at least based on some criterion for sorting.

A just legal system must also treat cases with similar characteristics alike and those that differ differently.[26] The desire for equitable treatment requires an assessment of when the circumstances of a particular case justify harsher or more lenient punishment and how to identify whether two cases are similar enough that they should be treated the same. Such comparisons are the means to guard against arbitrariness and to assess the reasonableness of a particular sentence. Relative severity, therefore, is a tenet of penal theory—a key element of both justifications for punishment and the way punishment is implemented in the courts. The crucial underlying question is: Relative to what?

---

[20] 18 U.S.C. § 3553(a)(2)(A) (2012); United States v. Bergman, 416 F. Supp. 496, 500 (S.D.N.Y. 1976).

[21] § 3553(a)(2)(B); Pell v. Procunier, 417 U.S. 817, 822 (1974); Hopt v. Utah, 110 U.S. 574, 579 (1884).

[22] CESARE BECCARIA, ON CRIMES AND PUNISHMENTS AND OTHER WRITINGS 10, 31 (Richard Bellamy ed., Richard Davies et al. trans., Cambridge Univ. Press 1995); Gary S. Becker, *Crime and Punishment: An Economic Approach*, 76 J. POL. ECON. 169, 176–79 (1968).

[23] BECCARIA, *supra* note 22, at 21; Becker, *supra* note 22, at 179–80, 185–87.

[24] BECCARIA, *supra* note 22, at 19; Becker, *supra* note 22, at 185–86.

[25] *See, e.g.*, Brown v. Plata, 563 U.S. 493, 537–44 (2011); JEREMY TRAVIS ET AL., NAT'L RESEARCH COUNCIL OF THE NAT'L ACADEMS., THE GROWTH OF INCARCERATION IN THE UNITED STATES (2014), http://www.nap.edu/catalog.php?record_id=18613 [https://perma.cc/3C7X-Y5FD].

[26] *See* U.S. SENTENCING GUIDELINES MANUAL § 1A1.3, policy statement (U.S. SENTENCING COMM'N 2016).

B. A Positive Psychological Theory of Relative Severity

Since there is no objective metric to relative severity, what emerges is an experience-based methodology. Judicial evaluations of severity are done not in relation to the entire possible spectrum of human behavior but to the limited spectrum that judges view on a regular basis. Judges compare a defendant in one case to defendants in other cases before them.

The evaluation of severity relative to the contours of cases in judges' caseloads exposes judges to the same psychological biases of context dependence that affect human decisionmaking more generally. Emphasizing the relative comparison *within* the sample can distort the relationship between cases in the sample and cases *outside* the sample.

1. Evidence from Psychology.

a. Extracting similarities.—Similarity plays a fundamental role in psychological theories of human knowledge and behavior.[27] Comparing similarities and differences is how people categorize and classify objects, analyze meaning, form concepts, and make inferences and generalizations.[28] It is how people evaluate the desirability of a particular choice or the merit of a particular issue.[29] It is how judges evaluate the circumstances of a particular case. And like other judgments, similarity depends on the context and frame of reference.[30]

Since objects or cases involve numerous potentially relevant dimensions for comparison, the outcome can depend on which dimensions are used and how they are weighed. One factor that affects the cognitive choice to emphasize certain factors over others is their diagnostic value—How relevant is the particular feature for the task at hand? To quote a classic example:

> [T]he feature "real" has no diagnostic value in the set of actual animals since it is shared by all actual animals and hence cannot be used to classify them. This feature, however, acquires considerable diagnostic value if the object set is extended to include legendary animals, such as a centaur, a mermaid, or a phoenix.[31]

In other words, the features of the members in the group can affect which dimensions of similarity will be used. Certain dimensions become more salient because they are useful to assess similarities and differences in

---

[27] *See, e.g.*, Amos Tversky, *Features of Similarity*, 84 PSYCHOL. REV. 327, 327 (1977).

[28] *Id.*

[29] For some prominent examples, see *infra* notes 47–51 and accompanying text.

[30] Tversky, *supra* note 27, at 340.

[31] *Id.* at 342.

comparison to other cases, and other dimensions are less salient if they are less useful.[32] Features that are shared by all objects in the set cannot be used to classify objects and, therefore, become less salient, while the quality that is not shared across all members becomes salient because it is a key characteristic in sorting the group members into different categories.[33]

The human tendency to group objects in a set around clusters also determines the diagnostic value of certain features. Such clusters are created in a way that maximizes the similarity of objects within a cluster and the dissimilarity of objects in different clusters.[34] Consider another example using the animal kingdom: a whale may be clustered with a fish in comparison to land animals, but as a mammal it would be different from a fish in a comparison among marine creatures. Changing the composition of a set can therefore alter the clustering of objects and change the diagnostic value of associated features.

The level of heterogeneity or homogeneity in the set of objects will affect the dimensions used for classification. When cases are heterogeneous, the analysis is based on more general features at a higher level of organization (for example, mammals, fish, and birds). Consequently, objects that can be grouped around a common dimension appear more similar to one another. When cases are more homogenous, the analysis becomes more granular and people apply more concrete distinctions within groups (for example, singing birds and poultry). The more homogenous caseloads are, the more dissimilar cases appear to be.[35] This means that similarity between two (or more) objects is both causal and derivative. It is causal in the sense that the similarity or difference across objects serves as the basis for classification and judgment. But it is also derived from the adopted classification that then influences the finding of similarity or lack thereof.[36]

As a result, when people face different sets, their answers to the same question of evaluation can be different. The problem is that people do not properly discount how much their evaluations depend on the particular set under examination. In a long line of experiments subjects were simply asked—How similar are X and Y? Whether discussing geometric shapes,[37]

---

[32] *Id.*

[33] *Id.* at 343–44.

[34] *Id.* at 342.

[35] *Id.* at 344.

[36] *Id.*

[37] Douglas L. Medin et al., *Respects for Similarity*, 100 PSYCHOL. REV. 254, 261–64 (1993) (finding that descriptions of an ambiguous shape were affected by the features of the other shape with which it was presented).

drawings of faces,[38] countries,[39] animals,[40] or musical instruments,[41] the answers reflected a contextual evaluation rather than an independent one.

b. Ranking severity.—Because human decisions are sensitive to the context in which they are made, the composition of object sets distorts the difference between relative and absolute quality. People are affected by the sample of cases before them to such an extent that they do not just rate the qualities of different features; they rank them relative to those of the other cases in the encountered sample. Even when asked to assign absolute values, people's estimations of any particular trait are affected by the location of the case relative to the range and skewness of the distribution of other cases with which it is presented.[42]

When a case is presented together with relatively milder cases, it can be evaluated as *absolutely* worse and vice versa. In psychological studies, such effects have been documented over a wide array of subjects and decisionmakers. They have been documented to affect the estimation of actual physical magnitudes: when encountered together with milder stimuli, the same volume was evaluated as louder,[43] the same object was found to weigh more,[44] the same line appeared longer,[45] and the same pattern

[38] Tversky, *supra* note 27, at 342 (showing that the features of a third face presented to the group largely determined which of two faces was chosen as most similar to the target).

[39] *Id.* at 344 (asking subjects which of two countries is most similar to a target country and finding that subjects' choices changed based on what the third country in the set was); Amos Tversky & Itamar Gati, *Studies of Similarity*, *in* COGNITION & CATEGORIZATION 79, 90–95 (Eleanor Rosch & Barbara B. Lloyd eds., 1978) (finding that the same pairs of neighboring countries were evaluated as more similar in a heterogeneous sample that included American and European pairs—where the feature of continent had diagnostic value—than in a homogenous sample that included only American or European pairs).

[40] Lennart Sjöberg & Christer Thorslund, *A Classificatory Theory of Similarity*, 40 PSYCHOL. RES. 223, 231–33 (1979) (finding that the similarity of pairs of mammals and birds increased when a more heterogeneous animal (i.e., a wasp) was added to the pool of pairs).

[41] *Id.* at 233–36 (demonstrating that the similarity of homogenous pairs of string instruments increased when heterogeneous pairs, including a string instrument and a wind instrument, were introduced into the list).

[42] Allen Parducci, *Category Judgment: A Range-Frequency Model*, 72 PSYCHOL. REV. 407, 408 (1965); Allen Parducci, *Contextual Effects: A Range-Frequency Analysis*, *in* 2 HANDBOOK OF PERCEPTION: PSYCHOPHYSICAL JUDGMENT AND MEASUREMENT 127 (Edward C. Carterette & Morton P. Friedman eds., 1974); Allen Parducci, *The Relativism of Absolute Judgments*, SCI. AM., Dec. 1968, at 84 [hereinafter Parducci, *Relativism of Absolute Judgments*].

[43] W.R. Garner, *Context Effects and the Validity of Loudness Scales*, 48 J. EXPERIMENTAL PSYCHOL. 218, 220–21 (1954); Allen Parducci & Arthur J. Sandusky, *Limits on the Applicability of Signal Detection Theories*, 7 PERCEPTION & PSYCHOPHYSICS 63 (1970).

[44] Vincent Di Lollo, *Contrast Effects in the Judgment of Lifted Weights*, 68 J. EXPERIMENTAL PSYCHOL. 383 (1964).

[45] Michael H. Birnbaum et al., *Contextual Effects in Information Integration*, 88 J. EXPERIMENTAL PSYCHOL. 158 (1971); David L. Krantz & Donald T. Campbell, *Separating Perceptual and Linguistic Effects of Context Shifts upon Absolute Judgments*, 62 J. EXPERIMENTAL PSYCHOL. 35 (1961).

appeared darker.[46] Context affects not only physical attributes but also feelings, such as happiness and satisfaction,[47] and evaluations of traits and normative opinions, such as judgments of merit[48] or antisocial behavior.[49] These effects have been repeatedly documented not only among laypersons but also among experts such as psychiatrists[50] and physicians[51] in their medical diagnoses.

2. Legal Applications.

a. Experimental evidence in the laboratory.—Both types of cognitive effects have been applied to legal decisionmaking in experimental settings to study the evaluation of behaviors and harms as well as the assessment of punitive damages awards and criminal sentences.

Parducci was the first to test the theory's effect on the evaluation of moral or immoral behaviors. In his experiment, two groups of students were asked to evaluate the moral blameworthiness of different behaviors ranging from "[f]ailing to put back in the water lobsters which are shorter than the legal limit" to "[p]oisoning a neighbor's dog whose barking bothers you."[52] The students were asked to judge the moral value of eighteen such acts.[53] Six of the acts were common to both groups of students; the other twelve acts differed.[54] One list contained relatively mild

---

[46] Barbara A. Mellers & Michael H. Birnbaum, *Loci of Contextual Effects in Judgment*, 8 J. EXPERIMENTAL PSYCHOL.: HUM. PERCEPTION & PERFORMANCE 582 (1982).

[47] Parducci, *Relativism of Absolute Judgments*, *supra* note 42, at 84–86.

[48] Barbara A. Mellers, *Equity Judgment: A Revision of Aristotelian Views*, 111 J. EXPERIMENTAL PSYCHOL. 242 (1982); Barbara A. Mellers, *"Fair" Allocations of Salaries and Taxes*, 12 J. EXPERIMENTAL PSYCHOL.: HUM. PERCEPTION & PERFORMANCE 80 (1986).

[49] Parducci, *Relativism of Absolute Judgments*, *supra* note 42, at 84–87.

[50] Douglas H. Wedell et al., *Reducing the Dependence of Clinical Judgment on the Immediate Context: Effects of Number of Categories and Type of Anchors*, 58 J. PERSONALITY & SOC. PSYCHOL. 319 (1990) (describing experiments related to diagnoses of psychopathology).

[51] To name one astonishing example, in a classic study, 389 children were examined for a tonsillectomy. Each time, the group that a doctor found not to need an operation was sent to another doctor. Across three different physicians in this chain, each doctor always found about half of the children in a set—cleared by a previous doctor—in need of surgery. AM. CHILD HEALTH ASS'N, PHYSICAL DEFECTS: THE PATHWAY TO CORRECTION 80–96 (1934); *see also* ELIOT FREIDSON, PROFESSION OF MEDICINE: A STUDY OF THE SOCIOLOGY OF APPLIED KNOWLEDGE 256–57 (1970); Harry Bakwin, *Pseudodoxia Pediatrica*, 232 NEW ENG. J. MED. 691, 691–92 (1945); Brian H. Bornstein & A. Christine Emler, *Rationality in Medical Decision Making: A Review of the Literature on Doctors' Decision-Making Biases*, 7 J. EVALUATION CLINICAL PRAC. 97, 100 (2001). Similar findings have been replicated for pediatricians' likelihood to recommend tympanostomy tube placement and to order ambulatory radiography. John Z. Ayanian & Donald M. Berwick, *Do Physicans* [sic] *Have a Bias Toward Action?: A Classic Study Revisited*, 11 MED. DECISION MAKING 154 (1991).

[52] Parducci, *Relativism of Absolute Judgments*, *supra* note 42, at 87.

[53] *Id.*

[54] *Id.*

behaviors, such as "[c]heating at solitaire," while the behaviors on the other list were much more troubling, such as "[u]sing guns on striking workers."[55] The students were instructed to rate each act based solely on their own moral judgment and to rate each act's moral value independently as if it were the only one they were evaluating. Still, Parducci found that the same acts "were rated more leniently by students who judged them in the context of the nasty list than they were by those who encountered them in the context of relatively mild wrongdoing."[56]

A series of studies by Sunstein, Kahneman, Schade, and Ritov has demonstrated that the evaluation of severity for different harms and subsequent decisions about damage awards are also sensitive to the group of cases being evaluated. One experiment looked at perceptions of different categories of harm: physical injury and financial loss.[57] When each case was evaluated in isolation, subjects emphasized within-category considerations[58] and awarded similar damages in both cases.[59] But when the two cases were evaluated together, subjects were reminded that personal injuries are much worse than financial losses, and the differences across categories dominated the decisions. While the damage awards in the financial-fraud case remained similar, damage awards for personal injury more than doubled from $1 million to $2.25 million.[60]

In a later paper, Eisenberg, Rachlinski, and Wells showed that these findings are also supported by patterns from real cases.[61] Judges ordinarily view cases involving varying types of harms while jurors view one case in isolation. And their decisions followed the predicted pattern: judges awarded higher damages than juries for personal injury cases and lower damages than juries for financial-loss cases.[62]

Another experiment examined awards of punitive damages in personal injury cases that involved a plaintiff suing a firm.[63] The experiment's

---

[55] *Id.*

[56] *Id.* at 84.

[57] Sunstein et al., *Predictably Incoherent Judgments*, *supra* note 7, at 1174–75.

[58] *Id.* at 1174–76.

[59] *Id.* at 1176. The average median award was $1 million for personal injury cases and $800,000 for cases involving financial loss.

[60] *Id.* at 1177. In that setting, the fact that it was the evaluation of physical injury that was mainly affected was expected. Because personal injury cases are more readily available in people's minds, the effect of the immediate comparison was mostly to increase the salience of the harm from the less prevalent category—the financial harm.

[61] Theodore Eisenberg et al., *Reconciling Experimental Incoherence with Real-World Coherence in Punitive Damages*, 54 STAN. L. REV. 1239 (2002).

[62] *Id.* at 1252–56.

[63] Sunstein et al., *Assessing Punitive Damages*, *supra* note 7.

vignettes differed in the harm the plaintiff was said to have suffered and in the size of the firm.[64] When a particular case was viewed in isolation, there was a tendency toward diminished differentiation in judgments about different scenarios. When several cases were evaluated in quick succession, however, the availability of context improved people's ability to discriminate among the cases and the range of awards increased.[65]

Similar results have been reported for sentencing decisions. Rachlinski and Jourden conducted an experiment using two hypothetical cases: an outrageous example of a fraud case and a mild example of an armed robbery with an unloaded handgun.[66] In each case, subjects were informed that a conviction carried a sentence of between two and fifteen years.[67] When evaluated separately, subjects ordered sentences that were closer to the upper limit of the available sentencing range for the outrageous fraud case, an average sentence of 12.27 years.[68] The average sentence ordered for the mild robbery case was 6.83 years.[69] When the two cases were viewed together, however, the comparative severity of the cases became more salient. The comparison across cases mitigated the sentence of the defrauder by approximately 2 years from 12.27 to 10.39 years and increased the sentence of the robber by approximately 1 year from 6.83 to 7.63 years.[70]

In a study by Rodríguez and Blanco, students were presented with five cases sequentially.[71] The last case was always the same robbery case, and it was preceded by either four relatively mild offenses or four relatively serious offenses.[72] The students were asked to choose one of seven possible sentences for the robbery case ranging from twelve to thirty-six months in four-month increments.[73] The authors found that the students exposed to a set of more severe crimes were more prone to impose lighter sentences (more often choosing twelve- or sixteen-month sentences) and less prone to impose harsher sentences (less often choosing thirty-two- or thirty-six-month sentences) than the students exposed to milder cases.[74]

---

[64] *Id.* at 2095.

[65] *Id.* at 2104.

[66] Rachlinski & Jourden, *supra* note 7, at 467–68.

[67] *Id.* at 468.

[68] *Id.* at 477.

[69] *Id.*

[70] *Id.*

[71] Gabriel Rodríguez & Sara Blanco, *Contrast Effect on the Perception of the Severity of a Criminal Offence*, 26 ANUARIO DE PSICOLOGÍA JURÍDICA 107 (2016).

[72] *Id.* at 109.

[73] *Id.* Because the experiment was conducted in Spain, the vignette followed Spanish law.

[74] *Id.* at 110.

The experimental findings focused on the immediate comparison across cases as a problem that can lead to arbitrariness in judgments. But in the same way contrast effect can change the relative evaluation of a case viewed against a more or less serious case, contrast effect in judicial decisionmaking can arise from the comparison between a particular case and the general caseload before the judge.

b. Qualitative evidence from the field.—Legal scholars have not researched a possible "caseload effect" in courts—where decisions in individual cases can be affected by the contours of the other cases in the judge's caseload. But criminologists and sociologists have qualitatively documented such a phenomenon among other agencies in the criminal justice system.[75]

Sociologists have suggested, for example, that police officers' exposures to crime transform their evaluations of the severity of different crimes. Police officers working in higher crime areas view certain crimes, such as prostitution or drug violations, as less serious and more tolerable, while police officers in lower crime areas control these crimes vigilantly.[76] As Klinger explains: "The more criminals and criminal types that officers observe . . . [and] the more apparent criminal conduct . . . that officers see in public, the higher the perceived level of deviance."[77]

Police detectives similarly differed in which cases they investigated: "[I]n the context of the cases a detective or detail typically received for possible investigative action, what was considered a 'big case' for one detective was a 'little case' for another."[78] Battery cases were regarded as "little" cases and rarely investigated by police detectives from a major crimes detail, but they were viewed as "fairly important" cases at the juvenile detail.[79] Urban police detectives regularly dealing with homicides "not only become[] 'familiar' with, hardened to, and less affected by such cases but also make[] finer, more varied distinctions between types of homicides . . . . [W]here killings make up the routine and regular work of

---

[75] For an extensive review of the literature, see Robert M. Emerson, *Holistic Effects in Social Control Decision-Making*, 17 LAW & SOC'Y REV. 425 (1983).

[76] David A. Klinger, *Negotiating Order in Patrol Work: An Ecological Theory of Police Response to Deviance*, 35 CRIMINOLOGY 277 (1997).

[77] *Id.* at 289.

[78] WILLIAM B. SANDERS, DETECTIVE WORK 95 (1977).

[79] *Id.* These findings are also in line with a more rational model of police enforcement, and police officers in both examples may well be engaging in efficient allocation of limited resources across crimes. The qualitative studies, *see supra* notes 75–78, however, suggest that police perceptions of the inherent severity of such crimes has changed as well.

detectives, not just any killing, but only some killings, are seen as serious.'"[80]

Increased exposure to criminal behavior over time may have a similar impact. Multiple studies have found that junior prosecutors tend to be harsher and more adversarial than their more seasoned colleagues, exposed to more frequent and more serious criminal cases. Junior prosecutors are more likely to insist on sticking with the most serious charge during plea negotiations and to offer harsher sentences.[81] Experienced prosecutors, on the other hand, view as routine circumstances that younger prosecutors initially view as outrageous and offer more favorable plea bargains as a result.[82] Prosecutors themselves describe in interviews how they soften over time especially as they move to prosecute more serious felonies.[83]

Across different agencies that encounter different populations of offenders, similar differences appeared. Juvenile delinquents who were identified by social welfare agencies as serious delinquents, for example, were often regarded as "essentially 'good kids'" by the juvenile court and parole boards that encountered much graver cases.[84] By routinely encountering a wide range of youthful misconduct, court personnel "have a higher tolerance of 'delinquency.'"[85]

Such different perceptions of offenders have also led to divergence between courts and parole boards. When courts in the 1960s sentenced the first marijuana offenders to imprisonment, judges viewed the required sentences as extremely harsh and recommended that offenders be released after the minimum sentence was served.[86] The parole board in California, however, was dealing with a population of offenders mostly serving much longer sentences of at least thirty months.[87] Compared to the general prison population, the lower sentences imposed in marijuana cases seemed too

---

[80] Emerson, *supra* note 75, at 434–35.

[81] Ronald F. Wright & Kay L. Levine, *The Cure for Young Prosecutors' Syndrome*, 56 ARIZ. L. REV. 1065, 1087–88, 1107–08 (2014).

[82] MILTON HEUMANN, PLEA BARGAINING: THE EXPERIENCES OF PROSECUTORS, JUDGES, AND DEFENSE ATTORNEYS 98–99 (1978).

[83] Wright & Levine, *supra* note 81, at 1107–08 (providing a long list of quotes from interviews with prosecutors); *see also* HEUMANN, *supra* note 82, at 99; Alafair S. Burke, *Prosecutorial Passion, Cognitive Bias, and Plea Bargaining*, 91 MARQ. L. REV. 183, 190 (2007).

[84] Emerson, *supra* note 75, at 431–32.

[85] ROBERT M. EMERSON, JUDGING DELINQUENTS 84 (1969).

[86] Caleb Foote, Annual Chief Justice Earl Warren Conference on Advocacy in the U.S., *The Sentencing Function*, *in* A PROGRAM FOR PRISON REFORM 17, 29 (1972).

[87] *Id.* at 29.

lenient to warrant early release, not only after the minimum sentence of usually six to twelve months was served but even after eighteen months.[88]

Notice that the diverging perspectives of the court and the parole board on sentences were driven by the difference in the dimensions that serve for the comparison across cases. Courts compare offenders based on behaviors or crimes in order to make sentencing decisions. Parole boards, it seems, compare offenders based on the length of sentence served when making release decisions. What is remarkable is how this led to increasingly punitive responses as each determination in the chain of criminal justice was made relative to a different contextual background.

These findings are qualitative and cannot isolate a particular causal effect from the other local and organizational environments agencies operate in. But the cumulative body of research suggests a pattern that is in line with the psychological theory. It demonstrates how each agency always has some crimes viewed as the most severe or the most trivial as defined in relation to its entire caseload.[89]

c. Evidence from a natural experiment.—Do judges' caseloads play a similar role in sentencing decisions? Taken together, the experimental findings and the qualitative field evidence suggest that they could. Yet not everyone will be willing to allow each line of research to compensate for the limitations of the other. The experimental settings allow for a causal inference, but they are limited to the lab and to measuring short-term effects, and they mostly rely on lay subjects. The field studies look at real cases but can only provide evidence of observed trends, not of their causes, and their focus is nonjudicial bodies.

Empirically testing for relative judgments in actual judicial decisions is technically complicated, and I, therefore, devoted a separate Article to developing the empirical methodology and reporting the findings.[90] In sum, the difficulty stems from the fact that a simple comparison across judges sitting in different courts or handling different dockets does not suffice. When judges have different caseloads, one cannot tell, for example, whether the differences observed in sentencing patterns are evidence of judges treating *different* cases differently or of judges treating *similar* cases differently. Another concern is that judges might be elected or appointed to different courts based on some prior underlying characteristics or propensities to punish, potentially creating reverse causation. More broadly, the communities across different courthouses can differ in many

[88] *Id*.

[89] Emerson, *supra* note 75, at 428.

[90] Leibovitch, *supra* note 10.

ways: there might be differences in the methods for processing cases, in prosecutorial policies, in local norms, and so on. Such confounding factors restrict the ability to infer that any differences observed are causally linked to caseload exposure among all other factors.[91]

To circumvent these difficulties, I developed an identification strategy to measure exogenous variation of a judge's exposure to criminal gravity. The methodology builds on a natural experiment enabled by the fact that cases are randomly distributed. Under random assignment, judicial caseloads should be balanced on average only when they contain a large-enough number of cases. But in the short term, with small numbers, we can observe a wider distribution of caseload compositions across judges. Tracking caseload composition over time, therefore, allows one to identify judges who randomly began with unbalanced caseloads before they balanced out in the long term and to compare their sentencing outcomes only during the period for which their caseloads are balanced.[92]

With this insight at hand, I created a matched sample of newly elected judges from the Pennsylvania Courts of Common Pleas.[93] In Pennsylvania, each offense is assigned by statute an offense gravity score (ranging from one to fourteen) that is used as a measure of offense severity to calculate the recommended sentencing range under the state sentencing guidelines.[94] I used the offense gravity score as an exogenous measure for case gravity, and for each judge I calculated the cumulative gravity exposure through caseloads for each day on the bench. Each pair in the matched sample included two judges who started hearing criminal cases in the same year and in the same district. In each pair, one judge randomly drew harsher cases during her first months on the bench and the other randomly drew milder cases. On average across the different matches, it took approximately three months until the initial differences decreased and judicial caseloads were balanced with similar exposure to criminal gravity. I then compared how the different initial exposure affected judges' sentencing decisions during the *following* period when their caseloads were similar, i.e., the "comparison period."[95]

I verified that during the comparison period, judicial caseloads were balanced with regard to a long list of case characteristics: the caseloads of the two groups contained cases with similar average offense gravity score, defendant prior record score, and mandatory minimum sentences; a similar

[91] *Id.* at 284–85.
[92] *Id.* at 293–95.
[93] *Id.* at 295–99.
[94] 204 PA. CODE § 303.15 (2015).
[95] Leibovitch, *supra* note 10, at 295–300.

gender and racial representation; and a similar proportion of cases going to trial.[96] I also tested for the individual characteristics of judges in the "high" and "low" gravity-exposure groups: both groups had the same proportion of judges with prior experience as prosecutors or defense attorneys, similar gender representation, and similar representations of political affiliations.[97] The two groups, in other words, had overall similar judges deciding randomly assigned similar cases. But sentences across the two groups widely differed depending on the caseloads encountered during the earlier period.

The results demonstrated the relativism in judges' sentencing decisions. The short-term impact on judicial sentencing patterns was substantial. Judges with initial exposure to low-gravity caseloads imposed sentences that were on average two months longer than those of judges initially exposed to more serious caseloads.[98] This represents a 25% increase relative to the average sentence (eight months) in the sample.[99] Judges who initially had caseloads of milder gravity were also significantly more likely either to sentence a defendant in the aggravated sentencing range or to depart above the guidelines' recommendations.[100] Although these effects decreased as judges were exposed to additional cases, it took approximately six months for the differences in sentencing practices to completely decay.[101]

These findings demonstrate that judges too are susceptible to relative judgments. The setting of the natural experiment focused on judges sitting in the same court who over time saw similar dockets, in order to isolate the causal effect of caseload exposure on sentencing outcomes. In many courthouses, however, differences across judicial docket compositions are systematic and persistent. When differences in judicial exposure to gravity are the result of a systematic diversion in the assignment of cases to judges and are not corrected over time, the resulting differences in the evaluation of criminal cases may persist and pose a greater concern for sentencing disparities in the justice system. The next Part offers such examples.

---

[96] *Id.* at 301.

[97] *Id.*

[98] *Id.* at 304–07.

[99] *Id.*

[100] *Id.* at 306, 308–13. Judges with initially milder caseloads were 1.8 percentage points more likely to depart above the sentencing guidelines recommendations and 6.4 percentage points more likely to sentence a defendant either above the guidelines or in the aggravated sentencing range. *Id.*

[101] *Id.* at 315–16.

II. JUDICIAL DOCKETS AND PUNISHMENT CURVES

A. The Caseload Effect

Judicial caseloads provide the context against which particular cases are evaluated. This caseload effect can change the substance of judicial decisions because of two cognitive processes: extracting similarities and relative comparison.[102]

Assessing similarities is a crucial part of legal decisions from the initial evaluation of offense severity to the detailed investigation into the existence of mitigating and aggravating circumstances. In criminal law, offenses cluster with different levels of generality. Dimensions of similarity can divide offenses according to their severity—for example, an infraction, a misdemeanor, or a felony. They can divide cases based on the type of offense—such as those involving drugs, property, or violence. They can be employed more granularly within a category of cases—to distinguish among, for example, drug cases involving marijuana, amphetamines, or cocaine. And they are demonstrated in clustering around aggravating or mitigating circumstances, as well as in the weight granted in the sentencing decision to personal circumstances or circumstances of the offenses.

The contours of cases in a judge's caseload affect which comparisons will be drawn. When caseloads are more heterogeneous, general distinctions will be most salient: between misdemeanors and felonies, for example, or violent and nonviolent crimes. The more homogenous caseloads become, the less relevant those general dimensions are. For a judge who only hears misdemeanors—or only drug cases or only cases involving juvenile offenders—such dimensions across categories lose their diagnostic value. Judgments of relative severity can be more directly affected by an evaluation of the case's severity relative to the other cases before the judge. The same case will be evaluated as relatively worse compared to more serious cases and as relatively less serious compared to more egregious crimes. The same defendant will be regarded as less culpable compared to recidivist felons but as a more serious criminal compared to first-time offenders.

The important question is: How representative are judicial caseloads of the universe of relevant criminal conduct? If caseloads are fairly representative, the relative severity of a case in the caseload—its location on the individual judge's punishment curve—will roughly correspond to its absolute severity. But judicial caseloads are often not representative.[103]

---

[102] *See supra* Section I.B.1.

[103] *See infra* Section II.B.

When the sentencing decision in a particular case is affected by the cases previously encountered—perhaps on the same day or several days before—relative judgments inject arbitrariness into sentencing. When the effect is driven by the overall composition of cases in the judge's docket, this can also introduce a systematic bias.

B. Limited Jurisdiction, Limited Perspectives

The idea that courts of limited jurisdiction are better tailored to serve particular policy goals has shaped the movement toward increased specialization of courts in recent decades.[104] At the same time, because of their focus on nonrepresentative caseloads, courts of limited jurisdiction are an institutional design that can unintentionally and systematically bias the punishment curve.

This Section discusses three categories that distinguish the various court structures currently in place across the United States: (1) unified and two-tiered courts for the assignment of misdemeanor and felony cases, (2) specialized courts focused on a particular type of offense (e.g., drug courts) or offender (e.g., juvenile or veteran courts), and (3) local courts focused on a particular geographic area because courts in different localities often encounter different crime levels.

What is common to these three categories of limited jurisdiction is that the cases in the jurisdiction of a court (or in the caseload of a judge) are more homogenous than the full spectrum of cases encountered by a judge in a generalist court. The added similarity makes dimensions that would otherwise distinguish these cases from other (unobserved) cases less salient and, therefore, distorts the comparison between the observed and unobserved cases leading to displacement of the punishment scale for the observed cases.[105] In all three structures, the evidence provides support for relative judgments as a model for judicial decisionmaking.

1. Unified and Two-Tiered Courts.—Some states use a two-tiered model of criminal jurisdiction in which a lower level court hears misdemeanor cases and a higher court sits in felony cases;[106] other states have a unified court system in which one court has general jurisdiction over

---

[104] *See generally* BAUM, *supra* note 1 (reviewing the history, policy aims, and varied implementation of specialized courts in the United States).

[105] *See supra* Section I.B.

[106] States in which there are two-tiered courts include, for example, Florida, Georgia, Kentucky, Michigan, New York, Ohio, and Virginia. For court structure charts of all U.S. states, see COURT STATISTICS PROJECT, *supra* note 19.

all criminal offenses.[107] States may also change, at times, the court models they implement, usually for efficiency reasons.[108]

In practice, even in states with a unified court system, the allocation of cases across particular judges of the court may nevertheless result in some judges having caseloads that are mostly misdemeanors, mostly felonies, or a mix.[109] For decisions about docket assignments, "the most meaningful level is the county or other unit of trial-court jurisdiction rather than the state."[110] Exact statistics are hard to come by, but from a survey of judges on courts of general jurisdiction conducted in 1977, 16% were actually serving in specialized courtrooms, and in 2011 Baum observed that "that level almost surely is higher today."[111] Presiding judges can be especially influential in metropolises where both caseload pressure and the number of available judges are large.[112]

The structuring of such court divisions is often based on administrative reasons—to save operating costs,[113] to address backlogs in

---

[107] States characterized by unified criminal court systems include, for example, Alabama, California, Connecticut, Illinois, Indiana, Maryland, Massachusetts, Missouri, New Jersey, Tennessee, and Wisconsin. *Id*.

[108] In 1998, California changed the entire structure of its state court system from a two-tiered to a unified model in an effort to save on operating costs. CAL. ATT'Y GEN., PROPOSITION 220: COURTS. SUPERIOR AND MUNICIPAL COURT CONSOLIDATION (1998) http://www.lao.ca.gov/ballot/1998/220_06_1998.htm [https://perma.cc/4FJV-8WB6]. In 2004, the criminal court (previously hearing only misdemeanors) and the supreme court (previously hearing only felonies) in the Bronx, NY, were merged to become a single supreme court hearing misdemeanor and felony cases together, with the hope of expediting the disposition of the large backlog of misdemeanors that accumulated in the criminal court. N.Y. COMP. CODES R. & REGS. tit. 22 § 42.1 (2004). In 2012, in light of the creation of large felony backlogs, the Bronx courts reverted to the two-tiered model. COMM. ON CRIM. COURTS & COMM. ON CRIM. JUST. OPERATIONS, N.Y.C. BAR, REPORT ON THE MERGER OF THE BRONX SUPREME AND CRIMINAL COURTS (2009), http://www.nycbar.org/pdf/report/20071735Merger_BronxSupreme_CriminalCourts.pdf [https://perma.cc/X99Y-VXJX]; LAWERENCE K. MARKS ET AL., BRONX MERGER STUDY GRP., THE BRONX CRIMINAL DIVISION: MERGER AFTER FIVE YEARS (2009), https://www.nycourts.gov/publications /pdfs/BronxReport11-09.pdf [https://perma.cc/SJ7V-9P29].

[109] Illinois, for example, has the second-largest unified court system in the United States. Yet my visits to the courts revealed how the division of felony and misdemeanor cases (as well as other forms of specialized dockets) varies considerably across courts. Even within Cook County, at the first municipal district (the City of Chicago), new judges are assigned solely to the misdemeanor docket and more senior judges to the felony docket. At the second municipal district (Skokie), the presiding judge assigns some judges only to misdemeanors, some judges to misdemeanors together with another specialization (usually either traffic or domestic violence cases), and some judges to both misdemeanor and felony cases.

[110] BAUM, *supra* note 1, at 21 (depicting the great differences between state courts' organizational charts and their operation in practice).

[111] *Id.*

[112] *Id.*

[113] The unification of California state courts, for example, was meant to save operating costs by merging the misdemeanor and felony courts into one building with joint staff. CAL. ATT'Y GEN., *supra* note 108.

the processing of certain offenses,[114] or to create promotion tracks for judges within the judicial system.[115] The division of misdemeanor and felony cases, however, can affect not only the processing of cases but the evaluation and substantive outcomes of cases as well. The case distribution observed by a misdemeanor judge, for example, is capped by the worst case charged as a misdemeanor, and it contains no representation of any harsher cases charged as felonies. This can pose a problem for the misdemeanor judge in evaluating the relation between the misdemeanors she observes and all other cases, and it may displace the sentencing scale by making such cases look worse than they otherwise would.

My own qualitative research at the courts of Cook County, Illinois, revealed that this is what happened in 2010 with domestic violence offenses in Chicago. Cook County, Illinois, is the second largest unified court system in the United States. The name "unified," however, as previously mentioned, rarely describes only one unified court. In fact, the City of Chicago has historically maintained one of the most specialized judicial systems in the country.[116]

Over the last decade, courts in Cook County gradually implemented specialized divisions for domestic violence cases. The first domestic violence division was established in October 2005 in the first municipal district—the City of Chicago.[117] The domestic violence division sits at a separate location from the criminal court, and ten judges are assigned solely to that division.[118] At the start of 2010, only the first municipal district had such a specialized division,[119] and its jurisdiction included all

---

[114] The Bronx courts were merged to overcome the large misdemeanor backlogs by having felony judges assist with hearing misdemeanor cases, and those courts demerged when felony backlogs were instead created. N.Y.C. BAR, *supra* note 108; MARKS ET AL., *supra* note 108.

[115] In Chicago, my conversations with judges revealed that the assignment of judges to misdemeanor or felony dockets is used to create internal "promotion tracks" within the unified court system. *See supra* note 109.

[116] *See generally* BAUM, *supra* note 1, ch. 4; ALBERT LEPAWSKY, THE JUDICIAL SYSTEM OF METROPOLITAN CHICAGO (1932); MICHAEL WILLRICH, CITY OF COURTS (2003); Harley, *supra* note 13, at 9–14.

[117] Press Release, Circuit Court of Cook Cty., Chief Judge Evans Announces Circuit Court of Cook County Committee on Domestic Violence Court (Aug. 21, 2008), http://www.cookcountycourt.org/MEDIA/ViewPressRelease/tabid/338/ArticleId/487/Chief-Judge-Evans-announces-Circuit-Court-of-Cook-County-Committee-on-Domestic-Violence-Court.aspx [https://perma.cc/VG7N-HZ2R].

[118] LESLIE LANDIS, CITY OF CHI. MAYOR'S OFFICE ON DOMESTIC VIOLENCE, ASSESSMENT OF THE CURRENT RESPONSE TO DOMESTIC VIOLENCE IN CHICAGO 129 (2007).

[119] Domestic violence courtrooms were later established at the fourth (Maywood) and fifth (Bridgeview) municipal districts in late September 2010, at the third municipal district (Rolling Meadows) in December 2010, and at the second (Skokie) and sixth (Markham) municipal districts in 2011. However, unlike the Chicago division, some suburban domestic violence courtrooms encompass one or two judges assigned to the domestic violence docket while still sitting in the district courthouse

domestic violence misdemeanors.[120] Felony domestic violence offenses remained under the jurisdiction of the felony trial division in the criminal court.

In January 2010, Illinois amended its domestic violence laws to increase punishment for domestic violence recidivists by permitting prosecutors to treat domestic battery as a Class-4 felony instead of a Class-A misdemeanor for those with a prior record.[121] Because, at that time, the domestic violence division in Chicago had jurisdiction over only misdemeanor domestic violence cases, recidivist domestic violence offenders who were charged with felonies had their cases heard in the criminal court. Incredibly, following the amendment, the felony charges in the criminal court received shorter sentences than those that would have been ordered for the same offense if charged as a misdemeanor in the domestic violence division.[122]

Despite the intended purpose of the amendment, prosecutors' requests for harsher punishments did not succeed in increasing sentencing levels in the criminal court. The court administration in Chicago held "stakeholders meetings" that included the presiding judges, court administrators, prosecutors, public defenders, and representatives from community organizations dealing with domestic violence; all were aware of the problem, and yet the sentencing differentials persisted.[123] A year later, the problem was solved with an institutional change: jurisdiction for Class-4 felonies was transferred from the criminal court to the domestic violence division.[124] When additional domestic violence courtrooms were established in late 2010 and 2011, their jurisdiction included, by design, Class-4 felonies as well.[125]

---

and sometimes assigned additional dockets as well. *See* Press Release, Circuit Court of Cook Cty., Court's New Domestic Violence Division Expands to Two Suburban Districts (Oct. 28, 2010), http://www.cookcountycourt.org/MEDIA/ViewPressRelease/tabid/338/ArticleId/459/Courts-new-Domestic-Violence-Division-expands-to-two-suburban-districts.aspx [https://perma.cc/B6DN-HP4G].

[120] LANDIS, *supra* note 118, at 129.

[121] 720 ILL. COMP. STAT. 5/12-3.2(b) (2012).

[122] Based on the author's conversations with judges, court administrators, prosecutors, defense attorneys, and members of community organizations dealing with domestic violence in the Chicago area.

[123] *Id*. This awareness led to a change in the jurisdiction of the Chicago domestic violence division, *infra* note 124, and to the different design of the jurisdiction of future domestic violence courts, Circuit Court of Cook Cty., *supra* note 119.

[124] CIRCUIT COURT OF COOK CTY., GENERAL ORDER NOS. 1.2, 2.1(g) (2014); *see also* Circuit Court of Cook Cty., *supra* note 119.

[125] Circuit Court of Cook Cty., *supra* note 119.

2. Specialized Courts.—In addition to the misdemeanor–felony distinction, the proliferation of specialization in courts over the last decades has led to increased use of separate divisions focused on certain categories of offenses or offenders.[126] Specialized courts often aim to offer separate treatment for identified populations—to emphasize the rehabilitation and social integration of less culpable offenders (as in juvenile or drug courts)[127] or to offer enhanced protection to more fragile victims (as in domestic violence or elderly courts).[128] In the last decade, we have witnessed a movement calling to extend the use of specialized courts to a growing list of offenders such as veterans[129] and, more recently, "young adults."[130]

Specialized courts can achieve their unique policy goals through a myriad of alternative procedures,[131] dispositions,[132] penalties,[133] and physical

---

[126] *See generally* BAUM, *supra* note 1.

[127] Juvenile courts are discussed in detail in this Section. For drug and treatment-oriented courts, see *infra* Section II.C.3.

[128] For a discussion of domestic violence courts—which in some jurisdictions, such as Chicago, are combined with the elderly court—see *infra* notes 200–06 and accompanying text.

[129] *See, e.g.*, Hawkins, *supra* note 19, at 565–67, 569; William H. McMichael, *The Battle on the Home Front: Special Courts Turn to Those Who Served to Help Troubled Vets Regain Discipline, Camaraderie*, 97 A.B.A. J. 42 (2011).

[130] Recent voices have called to either extend the juvenile courts' jurisdiction to include "young adults" between the ages of eighteen and twenty-four or to establish separate courts for them. *See, e.g.*, ROLF LOEBER ET AL., NAT'L INST. OF JUSTICE, FROM JUVENILE DELINQUENCY TO YOUNG ADULT OFFENDING 20–21 (2013); VINCENT SCHIRALDI ET AL., NAT'L INST. OF JUSTICE, COMMUNITY-BASED RESPONSES TO JUSTICE-INVOLVED YOUNG ADULTS 2–3, 10 (2015).

[131] In integrated domestic violence courts—such as those implemented in New York, Florida, and Oklahoma—the courts' focus extends beyond just criminal cases, with one judge handling all criminal domestic violence cases and related family issues such as custody, visitation, civil-protection orders, and matrimonial actions. LABRIOLA ET AL., A NATIONAL PORTRAIT OF DOMESTIC VIOLENCE COURTS 5 (2009). For different models of implementation of integrated domestic violence courtrooms, see, for example, *Integrated Domestic Violence Courts*, N.Y. STATE UNIFIED COURT SYS., https://www.nycourts.gov/ip/domesticviolence/ [https://perma.cc/KL4D-RDF7]; *Integrated Domestic Violence Court,* TULSA CTY. DIST. COURT, http://www.tulsacountydistrictcourt.org/accountability_courts.html [https://perma.cc/XA43-TGNH]. Other domestic violence courts, as well as some sex-offender courts, rely on extensive monitoring of defendants for prolonged periods. Allegra M. McLeod, *Decarceration Courts: Possibilities and Perils of a Shifting Criminal Law*, 100 GEO. L.J. 1587, 1621 (2012). Therapeutic mental-health or drug courts implement less adversarial procedures where "the specialized criminal court judge . . . engages in a direct, emotional, and frequently effusive manner with defendants, who are often referred to as the courts' 'clients.'" *Id.* at 1613.

[132] Proceedings at juvenile courts, for example, are usually sealed or expunged so that juveniles do not face the burden of a criminal record later on. INST. OF JUDICIAL ADMIN., JUVENILE JUSTICE STANDARDS, Part XVII (1979). In drug courts, the charges against the defendant are stayed while in the program and can be dismissed upon successful completion of treatment, while failure to complete the program results in initiation of traditional criminal sanctions (either decided at that stage or based on a pre-entered deferred plea bargain). NAT'L ASS'N OF CRIMINAL DEF. LAWYERS, *supra* note 19, at 17.

[133] Problem-solving or treatment-oriented courts—including drug courts, mental-health courts, and veterans' courts—emphasize rehabilitative alternatives in lieu of incarceration. McLeod, *supra* note 131, at 1590–91. They can also implement escalated intermediate sanctions in response to offender

and professional resources[134]—all assisting the courts to serve successfully the unique populations they target. Such specialization also enables judicial training and expertise where such are required for the appropriate treatment of identified populations or subtypes of cases.

But there is also evidence of unintended consequences—where courts that were meant to emphasize treatment and rehabilitation have actually developed more punitive sentencing practices. Especially for low-level offenses, the evidence reflects how generalist courts are sometimes more lenient than specialized courts exposed only to a bounded scope of severity.

That was the story of the history of the juvenile courts. The Progressive Era juvenile courts were established with an eye toward rehabilitation of adolescents, viewed as lesser offenders usually with no prior criminal record.[135] Yet already by the mid-twentieth century, although created with goals of rehabilitation and treatment, in practice many such courts took a punitive approach and "turned out to be a bad bargain for juvenile defendants."[136]

Juvenile courts traditionally did not sentence defendants to time in jail or prison but rather to time at a juvenile facility.[137] Usually the fact that defendants cannot stay in such institutions past the age of twenty-one capped the length of the sentence; the court imposed indeterminate terms, while the agencies to which the child was committed had discretion over how much of the sentence was actually served.[138]

Viewing the sentence imposed by juvenile courts as completely incomparable to incarceration,[139] however, has led to the chasm between the punishments of juvenile and adult defendants. In the landmark case *In re*

relapse or offer rewards for progress. NAT'L ASS'N OF CRIMINAL DEF. LAWYERS, *supra* note 19, at 18. Some juvenile courts implement separate sentencing schedules that are more lenient than the sentencing guidelines used for adults. *See, e.g.*, WASH. REV. CODE § 13.40.0357 (2016). In other states, juvenile courts still rely on indeterminate sentencing regimes—where the agency to which the juvenile is committed later determines when the juvenile is rehabilitated and how much of the sentence is served. *E.g.*, CONN. GEN STAT § 46b-141 (2012); IDAHO CODE § 20-520(1)(r) (2016); TENN. CODE ANN. § 37-1-137(a)(1)(A) (2010). In many states, such indeterminate commitment of juveniles is also limited to a maximum period of no more than two years, and a new hearing is required to determine whether additional commitment periods are necessary. *E.g.*, N.D. CENT. CODE § 27-20-36(2) (2011) (twelve months); CONN. GEN STAT § 46b-141 (2012) (eighteen months); ALASKA STAT. § 47.12.120(b)(1) (2008) (two years); GA. CODE ANN. § 15-11-607 (2014) (two years).

[134] *See infra* Section III.B.

[135] BAUM, *supra* note 1, at 106–08.

[136] *Id.* at 109.

[137] SAMUEL M. DAVIS ET AL., CHILDREN IN THE LEGAL SYSTEM 1278 (5th ed. 2014).

[138] *Id.* at 1278, 1286.

[139] *See*, *e.g.*, *In re* Eric J., 601 P.2d 549 (Cal. 1979); *In re* Aline D., 536 P.2d 65, 70 (Cal. 1975); *In re* J.L.P., 25 Cal. App. 3d 86, 89 (Cal. Ct. App. 1972).

Gault, a fifteen-year-old boy was sentenced to a maximum of six years at a juvenile facility for the offense of making lewd phone calls to his neighbor.[140] By contrast, for adults the statutory maximum penalty for such an offense was a $50 fine or two months in jail.[141] As the Supreme Court noted, "It is of no constitutional consequence—and of limited practical meaning—that the institution to which [the juvenile] is committed is called an Industrial School . . . . [H]owever euphemistic the title . . . an 'industrial school' for juveniles is an institution of confinement in which the child is incarcerated."[142] The historical shift toward implementing constitutional due process protections for juveniles[143] was similarly fueled by the recognition that, despite its rehabilitative and disciplining roots, "the modern juvenile court—with its emphasis on accountability and 'just desert' sentencing—would be unrecognizable to the late-19th-century proponents of the juvenile court model."[144]

---

[140] 387 U.S. 1, 7 (1967). This was not as unusual a sentence as one might think. In the case of *In re Winship*, a twelve-year-old child convicted of stealing $112 from a woman's purse in a locker was sentenced to "training school" for a minimum of 18 months and "subject to annual extensions of his commitment until his 18th birthday—six years in appellant's case." 397 U.S. 358, 360 (1970).

[141] *Gault*, 387 U.S. at 8–9.

[142] *Id.* at 27. In *McKeiver v. Pennsylvania*, Justice William Douglas in dissent compared the Pennsylvania correctional institution for juveniles—"a brick building with barred windows, locked steel doors, a cyclone fence topped with barbed wire, and guard towers"—to "a maximum security prison for adjudged delinquents." 403 U.S. 528, 560 (1971) (Douglas, J., dissenting) (quoting *In re* Bethea, 257 A.2d 368, 369 (Pa. Super. Ct. 1969)); *see also* Barry C. Feld, *The Punitive Juvenile Court and the Quality of Procedural Justice: Disjunctions Between Rhetoric and Reality*, 36 CRIME & DELINQ. 443, 453–54 (1990) (reviewing studies of juvenile correctional facilities across different states and multiple lawsuits challenging the conditions of confinement and concluding that during the 1970s "the daily reality of juveniles confined in many 'treatment' facilities is one of staff and inmate violence, predatory behavior, and punitive, custodial incarceration"). *But see* Martin Forst et al., *Youth in Prisons and Training Schools: Perceptions and Consequences of the Treatment-Custody Dichotomy*, 40 JUV. & FAM. CT. J. 1 (1989) (regarding conditions in the 1980s, Forst and others interviewed youth in juvenile and adult correctional programs and found that youth at juvenile facilities report greater satisfaction with counseling, treatment, and training programs and experience lower victimization rates than youth incarcerated at adult prisons).

[143] Historically, juveniles were not awarded the same due process rights that apply to adults at criminal courts based on classifying the proceedings as civil, rather than criminal, with the state acting as parens patriae to discipline the child. Starting from the 1960s, and following the punitive turn at the juvenile courts, the Court has gradually extended criminal procedural safeguards to juveniles as well. *See* Gallegos v. Colorado, 370 U.S. 49 (1962) (discussing investigation due process and the privilege against self-incrimination); Kent v. United States, 383 U.S. 541 (1966) (discussing right to counsel and full investigation required for waiver of jurisdiction); *Gault*, 387 U.S. 1 (discussing notice of charges, right to counsel, the rights of confrontation and examination, and the privilege against self-incrimination); *Winship,* 397 U.S. 358 (discussing observance of the standard of proof beyond a reasonable doubt); Breed v. Jones, 421 U.S. 519 (1975) (discussing the ban on double jeopardy).

[144] Janet E. Ainsworth, *Children and Criminal Procedure*, *in* THE CHILD: AN ENCYCLOPEDIC COMPANION 221, 222 (Richard A. Shweder et al. eds., 2009). In *Kent*, the Court explained:

Reflecting the shift from a model of parens patriae to just-deserts sentencing, since the late 1970s a practice of waiver of jurisdiction became increasingly prevalent—allowing the transfer of juvenile delinquents from the juvenile court to an adult criminal court in graver cases to allow for harsher sentencing.[145] Yet that practice also had its share of unintended consequences. From the 1970s to the 1990s, researchers were divided over the effectiveness of such transfers especially for less serious offenses. Some scholars identified a "leniency gap": juvenile offenders received more severe sentences in adult courts (relative to juvenile courts) when they were convicted for serious violent crimes but less severe sentences when they were convicted for property crimes.[146] Some studies have found that juveniles transferred to adult court were more likely to have their cases dismissed[147] and that juvenile defendants charged as adults with violent crimes were more likely than those retained in juvenile court to be released pending adjudication.[148] Young offenders often also received more lenient sentences in criminal court than they would have received in juvenile

---

> While there can be no doubt of the original laudable purpose of juvenile courts, studies and critiques in recent years raise serious questions as to whether actual performance measures well enough against theoretical purpose to make tolerable the immunity of the process from the reach of constitutional guaranties applicable to adults . . . . There is evidence, in fact, that there may be grounds for concern that the child receives the worst of both worlds: that he gets neither the protections accorded to adults nor the solicitous care and regenerative treatment postulated for children.

383 U.S. at 555–56 (footnote omitted). In *Breed*, the Court reiterated: "[O]ur decisions in recent years have recognized that there is a gap between the originally benign conception of the [juvenile court] system and its realities." 421 U.S. at 528.

[145] Fagan, *supra* note 1, at 243; Forst et al., *supra* note 142, at 2–3.

[146] *See, e.g.*, Carole Wolff Barnes & Randal S. Franz, *Questionably Adult: Determinants and Effects of the Juvenile Waiver Decision*, 6 JUST. Q. 117, 131 (1989); Marcy Rasmussen Podkopacz & Barry C. Feld, *Judicial Waiver Policy and Practice: Persistence, Seriousness and Race*, 14 LAW & INEQ. 73, 164 (1995). For a general overview of the literature, see, for example, Fagan, *supra* note 1, at 244; Howell, *supra* note 1.

[147] Marshall Young, *Waiver from a Judge's Standpoint*, *in* MAJOR ISSUES IN JUVENILE JUSTICE INFORMATION AND TRAINING 309, 313–16 (John C. Hall et al. eds., 1981) (discussing a study by the National Center for Juvenile Justice that took advantage of the difference in the age of adulthood between New York, where sixteen-year-olds were adjudicated as adults, and Philadelphia, where sixteen-year-olds were handled in the juvenile court, and finding that the adult court dismissed 74% of the offenders in comparison to only 48% of cases dismissed at the juvenile court); Dean J. Champion, *Teenage Felons and Waiver Hearings: Some Recent Trends, 1980-1988*, 35 CRIME & DELINQ. 577, 584 (1989) ("[A]ccording to several juvenile court prosecutors, it is likely that many of those juveniles who received probation or had their cases dismissed in criminal courts would have been adjudicated as a delinquent by juvenile judges and sentenced to secure confinement in one of several secure state detention facilities.").

[148] David L. Myers & Kraig Kiehl, *The Predisposition Status of Violent Youthful Offenders: Is There a "Custody Gap" in Adult Criminal Court?*, 3 JUST. RES. & POL'Y 115 (2001).

court,[149] especially property offenders when they appeared in criminal court as first-time adult offenders.[150] This disparity across the courts was contradictory to the purpose of transferring juveniles to the adult court. As Podkopacz and Feld observed: "It makes no penological sense for juvenile court judges to send youths to adult court to receive longer sentences, and then to have those criminal court judges impose shorter sentences than those meted out in juvenile court."[151]

Relative evaluation of case severity can explain such an anomalous outcome if the adolescents in criminal court appear "less severe" in comparison to older offenders.[152] In addition, sweeping legal reforms in many states during the 1980s and 1990s expanded the transfer of juveniles to adult courts—especially for violent offenses but also in drug and property cases. These reforms excluded many serious offenses from juvenile court jurisdiction. Thereafter, as juvenile courts increasingly tended to deal with less serious offenses, relative judgments can explain why these cases appeared relatively worse. The Office of Juvenile Justice and Delinquency Prevention at the U.S. Department of Justice acknowledged the competing findings and suggested a similar explanation:

> [I]t is likely that juvenile-criminal sentencing differences are largest in states that criminally prosecute only the most serious juvenile offenders. In states with transfer laws that apply to a broader range of less serious offenses, one would expect the adult system to regard transferred youth more lightly—and perhaps more lightly than the juvenile system would.[153]

3. Geographic Jurisdiction.—Another dimension for the assignment of jurisdiction across courts is geographic jurisdiction. Geographic disparities in sentencing were one of the main concerns of the Sentencing Commission and a central justification for the adoption of the Sentencing

---

[149] Martin Roysher & Peter Edelman, *Treating Juveniles as Adults in New York: What Does It Mean and How Is It Working?*, *in* MAJOR ISSUES IN JUVENILE JUSTICE INFORMATION AND TRAINING, *supra* note 147, at 265, 267 (examining case outcomes following the New York Juvenile Justice Reform Act of 1976 and concluding that sanctions at the criminal court were not more severe and often were less harsh); Young, *supra* note 147, at 314 (also reporting that "[t]he [Pittsburgh] juvenile court was twice as likely to commit a violent offender to a facility when the offense involved injury to person as was the adult court in New York").

[150] Barnes & Franz, *supra* note 146, at 133 (In California "[p]roperty offenders with a long history of property offenses tend to receive a substantially lighter sentence in adult court than they would have received when moving up the ladder in juvenile court."); Podkopacz & Feld, *supra* note 146, at 165 (In Minnesota "[t]he juvenile court imposed significantly longer sentences on less serious property offenders than did adult criminal court. This anomalous disparity between juvenile and criminal court sentencing practice perpetuates the 'punishment gap.'").

[151] Podkopacz & Feld, *supra* note 146, at 165.

[152] EMERSON, *supra* note 85; Fagan, *supra* note 1, at 239.

[153] OFFICE OF JUVENILE JUSTICE & DELINQUENCY PREVENTION, *supra* note 19, at 24.

Guidelines. By now, however, there is wide agreement that this battle is lost.[154] Many factors contribute to the variance in sentencing practices across courts, including different norms and needs across local communities and the courts that serve them, and they have been the subject of intensive study.[155] What this Article adds to that discourse is that, although not generally thought of in such a way, district courthouses are courts of limited geographic jurisdiction: when local caseloads differ, that leads to different relative judgments.

In particular, the division of geographic jurisdiction can be correlated with disparate exposure to criminal gravity. Criminologists analyzing the differences across local state courts found not only that their sentencing practices are different but also that such differences follow a particular pattern: judges in large urban counties are more lenient than judges in smaller rural or suburban courts.[156] Repeated research has demonstrated that in larger counties sentences of incarceration are shorter and less frequent.[157] In urban counties, judges are also more likely to sentence defendants below the guidelines range.[158] Downward departures from sentencing guidelines ranges for serious violent offenses are more likely in medium and large courts than in small courts.[159] And in courts with larger caseloads, downward departures from sentencing guidelines are also more likely and upward departures are less likely.[160]

Criminologists have also documented how different levels of violent crime across district courts affect sentencing decisions in particular cases. When violent crimes are a smaller portion of a court's caseload, they are treated more seriously.[161] As the proportion of violent crime in the court

---

[154] *See*, *e.g.*, Albert W. Alschuler, *Disparity: The Normative and Empirical Failure of the Federal Guidelines*, 58 STAN. L. REV. 85, 95, 100–02 (2005) (discussing the empirical findings by the Sentencing Commission's research staff and noting the possible interpretation that "the reported increase in geographic disparity swamped the reported reduction in interjudge disparity so that the Guidelines were an even bigger bust"); Bowman, *Failure of the Federal Sentencing Guidelines*, *supra* note 17, at 1327 (observing that "interdistrict disparities appear to have grown larger in the guidelines era, particularly in drug cases"); Daniel Richman, *Federal Sentencing in 2007: The Supreme Court Holds—The Center Doesn't*, 117 YALE L.J. 1374, 1380–83 (2008) (suggesting that heterogeneous district strategies were in fact celebrated and encouraged by the Department of Justice).

[155] *See supra* notes 15–16 and accompanying text; *see also infra* notes 157–60.

[156] *See supra* note 3; *see also infra* notes 157–60 and accompanying text.

[157] Ulmer & Johnson, *supra* note 3.

[158] John H. Kramer & Jeffery T. Ulmer, *Sentencing Disparity and Departures from Guidelines*, 13 JUST. Q. 81 (1996); Jeffery T. Ulmer & John H. Kramer, *Court Communities Under Sentencing Guidelines: Dilemmas of Formal Rationality and Sentencing Disparity*, 34 CRIMINOLOGY 383 (1996).

[159] John H. Kramer & Jeffery T. Ulmer, *Downward Departures for Serious Violent Offenders: Local Court "Corrections" to Pennsylvania's Sentencing Guidelines*, 40 CRIMINOLOGY 897 (2002).

[160] Johnson, *supra* note 15, at 781–82.

[161] EISENSTEIN ET AL., *supra* note 3, at 271.

increases, the likelihood of downward departures for a violent crime increases and the likelihood of upward departures decreases.[162]

This Article suggests that such trends fit within a larger framework. Rural and suburban areas, and areas served by smaller courts, generally face lower crime levels than large urban areas.[163] Consequently, relative evaluations of severity can lead rural judges to judge the same cases more harshly than urban judges.

While there could be good arguments in favor of local courts serving their local communities,[164] including through adjusting sentencing levels, there are also reasons why awareness of such a phenomenon is important, and such a trend is not necessarily desirable. Justifying local norms that are based on substantive differences in attitudes, beliefs, and pressing needs is one thing. But if such norms are, at least to some extent, driven not just by social or ideological factors but are also mechanically nudged by caseload effects—that is a different rationale that perhaps deserves less weight.

C. Considerations for Case Assignments

Once the unintended consequences of limited docket composition have been acknowledged, the above analysis reveals several factors that can particularly exacerbate or attenuate the outcomes of relative judgments in the courts.

1. Scope of Severity in Caseloads.—The evaluation of relative severity will be particularly biased when a case is evaluated against a nonrepresentative sample of criminal severity. The distinction between a misdemeanor and a felony, or a violent and a nonviolent crime, is a fundamental one for a judge hearing all types of offenses. But for a judge who only hears misdemeanor cases or nonviolent cases, such features are irrelevant for differentiation—after all, all cases in that court are such.[165] The ranking of a particular case's severity will also be different when the caseload contains no representations of harsher cases.[166] The same misdemeanor will be ranked relatively low compared to all offenses (because all felony cases will be ranked higher), and it will be ranked higher compared against only misdemeanor cases. This can pose a problem

---

[162] Johnson, *supra* note 15, at 775, 785.

[163] *See*, *e.g.*, DETIS T. DUHART, BUREAU OF JUSTICE STATISTICS, U.S. DEP'T OF JUSTICE, URBAN, SUBURBAN, AND RURAL VICTIMIZATION, 1993–98 (2000); ELIZABETH KNEEBONE & STEVEN RAPHAEL, BROOKINGS INST., CITY AND SUBURBAN CRIME TRENDS IN METROPOLITAN AMERICA (2011).

[164] *See supra* notes 15–16 and accompanying text.

[165] *See supra* Section I.B.1.a.

[166] *See supra* Section I.B.1.b.

for a judge with a bounded scope of gravity in her caseload in evaluating the proper relation between the cases she observes and all other cases and shift the sentencing scales of judges evaluating only misdemeanors or only felonies.

The severity of cases is perhaps the most recognized and prevalent dimension of jurisdictional assignments in the judiciary, relevant to all the previously reviewed examples: this is the type of decision underlying the structure of unified or two-tiered divisions in state courts,[167] as well as the restriction of the jurisdiction of specialized courts to misdemeanor or low-class felonies.[168] Offense severity is also a criterion in deciding whether offenders are eligible for drug court treatment[169] and whether juveniles are charged at the juvenile or the adult court.[170] This is also the underlying problem in the exposure of different local courts to different levels of crime.[171] Exposure to cases of lower severity can explain all the phenomena reviewed in Section II.B above, and it is a serious concern for the current structure of court systems.

If the punishment curve used by the misdemeanor judge is to the right of the true distribution and the punishment curve used by the felony judge is to the left, then some misdemeanors could be sentenced more harshly than felonies. Such outcomes contradict retributive notions of punishment. They are also contradictory to maintaining marginal deterrence for crimes of increasing severity. They often also undermine the achievement of other social goals like getting tougher on domestic violence or emphasizing rehabilitation at the drug courts.

Such a paradoxical effect—where misdemeanors get sentenced more harshly than felonies involving similar conduct—is also in tension with the lesser procedural safeguards granted to misdemeanor defendants. Many procedural entitlements possessed by felony defendants—such as rights to a grand jury, to a preliminary hearing, and to increased discovery—are not awarded to defendants in misdemeanor cases.[172] The reason is that under existing law, the application of procedural safeguards rests almost entirely on the maximum term of incarceration authorized by the charged offense.[173]

---

[167] *See supra* Section II.B.1.

[168] *See supra* notes 19, 117–25 and accompanying text.

[169] NAT'L ASS'N OF CRIMINAL DEF. LAWYERS, *supra* note 19, at 22.

[170] *See supra* notes 19, 145–53 and accompanying text.

[171] *See supra* Section II.B.3.

[172] Paul T. Crane, *Charging on the Margin*, 57 WM. & MARY L. REV. 775, 802–06 (2016).

[173] This distinction between misdemeanor and felony defendants has already faced criticism. Some critiques highlight the value of fair process and avoiding wrongful convictions, regardless of how "petty" the potential punishment is. Alexandra Natapoff, *Misdemeanors*, 85 S. CAL. L. REV. 1313, 1352 (2012). Others emphasize the vast legal, economic, and psychological implications of misdemeanor

But when jurisdictions implement court structures that lead to systematically sentencing some misdemeanors more harshly than felonies, the focus on maximal rather than actual punishment is hard to justify.

Notably, courts of limited jurisdiction usually concentrate on categories of less serious offenses and are often restricted to hearing misdemeanors or low-class felonies.[174] As a result, the overall effect of dividing jurisdiction based on offense gravity is also a tendency towards increased punitiveness.

2. Correlation with Relevant Factors for Sentencing.—Even when caseloads are not assigned directly based on severity, they can affect judicial evaluation of relative severity by changing the relative weight of certain mitigating and aggravating factors. Most specialized courts are constructed around personal characteristics that are widely recognized as, and closely associated with, relevant considerations for sentencing. Yet once a mitigating or aggravating dimension becomes common to all persons before the court, the contrast is reduced between those characterized by that feature and those that are not.

Consequently, the unintended consequences of such specialization might be that an otherwise relevant consideration in sentencing loses its diagnostic value and becomes irrelevant.[175] Being a juvenile, for example, can be an important distinguishing factor in the criminal court but ironically a lesser one at the juvenile court. The seventeen-year-old offender would fair relatively well compared to a recidivist adult offender but much worse when evaluated against only younger juveniles. A particular veteran's military service record and mental and economic state would make her appear very sympathetic relative to the general population of offenders, but these qualities will not be distinguishing factors among defendants who are all veterans.

When the diagnostic value of such features is eliminated, they are less likely to be taken into account in the sentencing process. But although they do not possess much diagnostic value to distinguish among cases within the judges' caseloads, they are important factors in penal evaluations more

convictions, *id.* at 1325–27, including the potential for lengthier incarceration based on the aggregation of multiple misdemeanor charges, Colleen P. Murphy, *The Narrowing of the Entitlement to Criminal Jury Trial*, 1997 WIS. L. REV. 133, 157–69, and the increasing attachment of collateral consequences to misdemeanor convictions—such as disenfranchisement, deportation, sex-offender registrations, and firearm prohibitions—previously reserved only for convicted felons, Crane, *supra* note 172, at 780–82. All, however, remain focused on potential consequences—carceral or collateral—but have not extended the critique to a test of the actual practice of punishment in such offenses.

[174] *See supra* note 19; *supra* Sections II.B.1–2.

[175] *See supra* Section I.B.1.

generally. Forgoing them can undermine the relative scaling of such cases against other cases in the criminal justice system.

3. Uniqueness of Cases.—When a caseload is more homogenous than the overall distribution of offenses, it is also important to consider the expected magnitude of the punishment curve's effect. In particular, one should consider whether the caseload has the type of cases that already call for a separate sentencing scale or whether the cases should be evaluated proportionally to other types of cases.

For most criminal cases, proportionality is relevant across different levels and types of crimes, and therefore, nonrepresentative caseloads raise the serious concerns of biased sentencing curves discussed above. The division of misdemeanor and felony cases across courts, or the concentration of all crimes committed by juveniles under one court, are such examples.

Some crimes, however, are unique—they justify a separate sentencing scale or they are unlikely to be matched against crimes from other categories. For those crimes, the lack of a comparison to other cases might be less distortive. Traffic cases, for example, are mostly punishable by fines and often viewed as a separate category from the general criminal docket. Crimes that carry mostly monetary fines might be harder to scale compared to crimes that are punished mainly with imprisonment.

Another example is when policymakers are concerned that certain crimes have unique features whose gravity cannot be properly identified when they are immersed in a general criminal caseload, as is argued about domestic violence offenses. One might think that domestic violence cases are evaluated more accurately when judged separately because the focus becomes the abusive domestic environment rather than the physical harm itself. If domestic violence cases are unique in a way that justifies a separate sentencing curve, then separating them from other cases may be benevolent rather than distortive.

4. Types of Dispositions.—The proliferation of specialization in the courts has also brought with it an increased movement to establish therapeutic and treatment-oriented courts—focused on court-supervised treatment programs in lieu of criminal sanctions—and the consequences of relative judgments may be different based on the type of disposition utilized. In sentencing decisions, the concerns arising from relative judgments are most acute because they conflict with the desire for proportionality and consistency in sentencing. Rehabilitation, on the other hand (at least conceptually), is more tailored to the circumstances of an individual defendant. In therapeutic or treatment-oriented courts, the

problem of biasing the sentencing curve might be less acute due to those courts' focus on a judge-supervised treatment plan as an alternative to the adversarial criminal process—the judge "[a]ct[s] more like a probation officer than a jurist."[176] Because the underlying concept for these courts "begins with an understanding that routine criminal punishment will not address the participant's underlying problem,"[177] by virtue of being disinterested in punishments the courts escape the need for proportionality in sentencing—and, therefore, the concern of relative judgments as an obstacle to proportionality.

One should be cautious, though: as past experience with drug and juvenile courts teaches, it is not always easy to tell whether treatment-oriented courts are interested only in rehabilitation, and it is sometimes unclear how to evaluate case outcomes and their punitiveness. Scholars have long warned that "[t]he most serious vice of sanctioning systems that were associated with the rehabilitative ideal was their tendency to distort the definition of just proportion in punishment."[178] Increased awareness is required to avoid the pitfalls of overpunitiveness in the dispositions of such courts—like the concerns that guided states in legislating limits on the commitment of juveniles[179] or that guided the Supreme Court's extension of due process rights to juveniles in the juvenile courts of the mid-twentieth century.[180]

More generally, relative judgments might be less concerning only if such courts truly do not impose punishments. In modern juvenile courts, different mechanisms blur the line between juvenile and criminal court dispositions. All states implement some, and often several, mechanisms to charge juvenile offenders in adult courts[181] or to transfer juveniles between

---

[176] Hawkins, *supra* note 19, at 568.

[177] *Id.* at 570.

[178] Franklin E. Zimring, *Drug Treatment as a Criminal Sanction*, 64 U. COLO. L. REV. 809, 813 (1993); *see also*, *e.g.*, Josh Bowers, *Contraindicated Drug Courts*, 55 UCLA L. REV. 783, 788 (2008) ("[D]rug courts have not—as advertised—abandoned the 'traditional criminal justice paradigm' . . . . They have merely relocated the old paradigm to the background."); McLeod, *supra* note 131, at 1615–17, 1619 (warning that "cordoning off certain courts as purely involved in therapeutic interventions may both misstate what is actually occurring in those courts and undermine judicial self-consciousness about whether the punitive effects of particular decisions are proportional to the offending conduct and no greater than necessary to deter offending behavior").

[179] *See, e.g.*, 705 ILL. COMP. STAT. § 405/5-710(7) (2016); LA. CHILD. CODE ANN. art. 898(A) (2016); N.C. GEN. STAT. § 7B-2513(a)-(b) (2015); TENN. CODE ANN. § 37-1-137(a)(1)(B) (2015).

[180] *See supra* notes 143–44 and accompanying text.

[181] As of 2009, twenty-nine states offer statutory exclusions for certain offences, filed directly in criminal courts, and fifteen states leave the decision whether to charge the defendant in the criminal or juvenile court in the hands of the prosecutor. OFFICE OF JUVENILE JUSTICE & DELINQUENCY PREVENTION, *supra* note 19, at 3–6. In thirty-four states, a juvenile who was sentenced once by a

the juvenile and adult courts.[182] Many states also have blended sentencing laws, allowing both adult and juvenile courts to implement overlapping disposition and sentencing options.[183] Where the criminal and juvenile courts employ a parallel set of sanctions, the question of harmony—or proportionality—between cases at the different courts fully resurfaces.

Similarly, drug courts often carry the threat of criminal sentences as the stick alongside the carrot for compliance of avoiding conviction and sanctions.[184] But that stick does not necessarily need to be implemented by the same court that supervises the treatment program. Instead, defendants who fail to meet the conditions of their treatment or supervision could be transferred back to the general criminal system for sentencing.[185] To the extent treatment-oriented courts also impose criminal penalties, by performing such a function the court is no longer disinterested in proportionality. In fact, it might then be especially susceptible to the problems of different punishment curves.[186]

---

criminal court remains under adult jurisdiction for future cases regardless of the level or type of later offences. *Id.*

[182] As of 2009, forty-five states have discretionary transfer laws—generally triggered by a prosecutorial request but sometimes by the court—under which the juvenile court may transfer the case to the criminal court. *Id.* at 2–5, 7. Twenty-six states also have some form of presumptive or mandatory waivers for certain offenses or circumstances. *Id.* at 2–5, 7. And in twenty-four states, a reversed mechanism exists in which the criminal court can transfer cases of young offenders to be heard at the juvenile court. *Id.* at 2–3, 5, 7.

[183] Fourteen states provide juvenile courts with criminal-sentencing options as well (juvenile blended sentencing), and eighteen states allow criminal courts to impose juvenile dispositions (criminal blended sentencing). *Id.* at 7.

[184] NAT'L ASS'N OF CRIMINAL DEF. LAWYERS, *supra* note 19, at 18; Bowers, *supra* note 178, at 788.

[185] Other proposals by critics of the drug courts similarly aim to disentangle treatment and sentencing. Bowers envisions a voluntary opt-in process in which drug addicts can choose when to participate and successful treatment would lead to expungement of participants' past record. Bowers, *supra* note 178, at 832–33. McLeod proposes a shift of focus to emphasize a decarceration model for specialized courts leading to de facto decriminalization of certain conducts and enabling an alternative noncarceral social approach focused on social services and treatment institutions. McLeod, *supra* note 131. This Article does not advocate for any one particular form of reform. It supports the critique by highlighting the importance from a relative-judgment perspective, of overcoming the amalgamation of sentencing functions together with exposure to a bounded scope of severity or unrepresentative population of offenders.

[186] There is indeed some evidence of certain drug courts imposing "atypically long prison sentences for the very defendants that drug courts were supposed to keep out of prison and off of drugs," Bowers, *supra* note 178, at 786, "sentences [that] are significantly higher for those who seek drug treatment and fail than for those who simply avoid drug treatment and take a plea, at both the misdemeanor and felony level," NAT'L ASS'N OF CRIMINAL DEF. LAWYERS, *supra* note 19, at 29, and "unnecessary terms of incarceration for minor or technical violations," McLeod, *supra* note 131, at 1618.

The bias introduced by relative judgments, therefore, only can be alleviated if the line between treatment-oriented courts and criminal courts is carefully drawn. And the concern about its unintended consequences needs to be considered in drawing such jurisdictional lines and in deciding about the extent to which treatment and sentencing are interconnected. When courts operate under a blended model focusing on a population in need of treatment—such as those suffering from drug or alcohol abuse, the mentally ill, veterans, or juveniles—complete specialization together with sentencing functions raises a heightened risk for unintended consequences contrary to the court's mission.

* * *

Because the type of cases under the jurisdiction of courts affects the substance of the decisions reached, decisions about divisions in the court system are not merely administrative. They entail a decision about sentencing levels themselves by nudging the sentencing curve in a particular direction. In that sense, the increasing movement towards specialization in courts inherently undermines the quest for uniformity in sentencing.

And it undermines it in a particularly disturbing way since the sentencing curves that develop are not based on any penological considerations. Such biased sentencing clearly undermines notions of justice and equal treatment because similarly situated defendants are sentenced differently. These sentencing disparities cannot be justified as the price we pay for judicial discretion. It is not that sentencing differentials reflect differences in judicial ideology because even the same judge might order different sentences based on exposure to different cases *at the court*. This is also not about emphasizing individualization in sentencing because the decision in a particular case is affected by the characteristics of the *other* cases in a judge's caseload. Furthermore, because courts of limited jurisdiction often concentrate on relatively less serious offenses, specialization will often also be associated with increased punitiveness.

III. LESSONS FOR COURTS OF GENERAL, LIMITED, AND HYBRID JURISDICTION

Acknowledging the unintended consequences of limited jurisdiction raises the question of how one should approach specialization and jurisdictional divisions in the judiciary. If the problem of relative judgments is exacerbated when judges face homogenous caseloads—especially when those caseloads differ from the general distribution of crimes—a simple policy prescription is to ensure that judges' caseloads are

representative of the general distribution of crimes. This suggests a potential reason to prefer courts of general jurisdiction to courts of a limited one.

Of course, the bias introduced by relative judgments is one cost that should still be balanced against the potential benefits of specialization. But once it becomes clear that the impact of limited jurisdiction is not merely administrative (and outcome neutral) but rather that it has a significant substantive cost as well, there needs to be a significant benefit to outweigh it—and the need for a particular justification for specialization becomes acute. Where specialization is based on administrative or resource considerations—not on any benefits of specialization—generalist courts may be optimal. If specialization is required for particular benefits (such as special training or judicial expertise) or when complete unification is impractical (for example, because of geographic barriers), policymakers should consider the use of hybrid court models where possible.

A. Operational Reasons for Limited Jurisdiction

When the reason for separating the treatment of certain cases is merely operational, the benefits of using courts of limited jurisdiction are not uniquely tied to that particular organizational design and might, therefore, push towards preferring courts of general jurisdiction. The misdemeanor–felony distinction is one dimension where court structures are usually determined based on administrative considerations.[187] At the same time, the creation of two-tiered dockets can bias sentencing outcomes based on the dimension where proportionality is especially acute—the severity of the offense. This might, therefore, be an example for where courts of general jurisdiction would be better aligned with proportionality in sentencing.

Further, evidence implies that while random assignment of cases is superior to nonrandom dockets, even better would be to ensure that caseloads are not only random but also balanced with similar exposure to cases.[188] As judicial caseloads become more representative of the overall crime distribution, the punishment curve applied by the judges more closely tracks real-world proportionality and the decisions of different judges become more coherent.[189] Implementing case-assignment systems that take into account the current composition of judicial dockets in future

[187] *See supra* notes 106–15 and accompanying text.

[188] In previous scholarship, I conducted a "placebo" test and found that when judges' initial caseloads contained similar exposure to gravity, the judges' following sentencing outcomes were practically identical. Leibovitch, *supra* note 10, at 309, 314.

[189] *Id.* at 309.

assignment decisions could equalize judicial exposure and by that the background against which judicial decisions are made.

This is also an area that highlights the importance of devoting attention to such jurisdictional decisions at a centralized level. Currently, decisions on the unit of specialization are not limited to instituting courts of limited jurisdiction. They are also prevalent at the level of creating narrower divisions and specialized dockets within courts of general jurisdiction.[190] This empowers particular judges and administrators to make such decisions, often with little regard to the situation in neighboring courts or to the potential impact on the substantive outcomes reached by judges. It was such an oversight, for example, that led to the failure in Chicago of the aforementioned amendment to Illinois domestic violence laws (that enabled felony charges against recidivist offenders) because only the Chicago court separated between misdemeanor and felony cases.[191] When assigned to separate divisions, the felonies were sentenced more leniently than the misdemeanors.[192] Within a given jurisdiction concerned with equitable treatment across cases, greater emphasis should be granted to ensure that different courthouses within the jurisdiction implement similar case-assignment rules and to promote centralized decisionmaking at the county or state level rather than at the discretion of the presiding judges of particular courts.

The division of geographic jurisdiction is similarly driven not only by differences in local knowledge or approaches but to a large extent by practical constraints as well. A generalist court serving the entire nation is clearly infeasible, and so is one court per state. But at a lower administrative level, some narrower jurisdictional divisions perhaps should be reconsidered. In large counties, for example, local courts are usually divided based on municipal districts. Within the municipal districts, even narrower jurisdictions are often established with particular courthouses, or judges in a courthouse, focusing on cases of certain cities, towns, or villages.[193] Wide geographic distances across municipalities can perhaps

---

[190] *See supra* notes 109–12 and accompanying text.

[191] *See supra* notes 117–25 and accompanying text.

[192] *See supra* notes 121–25 and accompanying text.

[193] The largest county in the United States, Los Angeles County, California, is divided into nine county regions and operates twenty-four criminal courthouses. *Criminal Courthouses*, SUPERIOR COURT OF CAL., CTY. OF L.A., http://www.lacourt.org/courthouse/mode/division/criminal [https://perma.cc/HB2Y-VUYK]. The second largest county, Cook County, Illinois, has six different municipal judicial districts. As a practical matter, the criminal courthouses in suburban judicial districts often further divide municipalities across judges. *Organization of the Circuit Court*, STATE OF ILL., CIRCUIT COURT OF COOK CTY., http://www.cookcountycourt.org/ABOUTTHECOURT/OrganizationoftheCircuitCourt.aspx [https://perma.cc/E7NG-JWFA].

justify some divisions across courts. But when defendants from different cities arrive at the same courthouse, dividing municipalities across judges is likely less justified. Policymakers could also consider local crime levels to design jurisdictions with more similar crime exposure. With prison overpopulation already a salient issue in the United States, parole boards could take into account possible misalignments in the sentencing of offenders across district courthouses within the county or the state in early-release decisions.[194] More generally, any such decisions should also take into account the impact of caseload exposure on sentencing levels and balance them accordingly—something that is currently not done.

B. Substantive Reasons for Limited Jurisdiction

Specialization may also be desirable due to special professional knowledge, training, or resources that are required for the appropriate treatment of identified populations or subtypes of cases.

Importantly, what I term "substantive reasons" here are distinguishable from a mere desire to affect sentencing levels. Sentencing levels can be addressed in an array of ways—from separate sentencing schedules to acknowledging certain circumstances as mitigating and aggravating factors to legislated mandatory minimum or statutory maximum sentences.[195] Assigning cases to a separate court is not necessary to implement such penal goals.[196] And because of the bias introduced by

---

[194] This is especially relevant because the local districts developing different punishment levels are not the ones internalizing the costs associated with such sentences, and the cost of imprisonment is born at the state (for state prisons) or county (for county jails) level.

[195] Washington, for example, implements separate sentencing schedules for juveniles. WASH. REV. CODE § 13.40.0357 (2016). Illinois made a prior record an aggravating circumstance in domestic violence sentences—enabling it to charge recidivist offenders with felonies instead of misdemeanors. 720 ILL. COMP. STAT. 5/12-3.2 (2012).

[196] Indeed, this distinction lies at the heart of the famous debate between abolitionists and nonabolitionists of the juvenile courts. Scholars calling for the abolition of the juvenile courts emphasize that: "If shorter sentences for reduced culpability is the principal justification for juvenile courts, then providing youths with fractional reductions of adult sentences could just as readily meet that goal." Feld, *supra* note 142, at 461; *see also*, *e.g.*, Katherine Hunt Federle, *The Abolition of the Juvenile Court: A Proposal for the Preservation of Children's Legal Rights*, 16 J. CONTEMP. L. 23, 50 (1990); Barry C. Feld, *The Transformation of the Juvenile Court*, 75 MINN. L. REV. 691, 724 (1991). The opposition to abolition of the juvenile courts, on the other hand, focuses on the importance of the courts' procedures for social goals beyond sentencing: "[T]he juvenile courts do afford benefits that are unlikely to be replicated in the criminal courts, such as the institutionalized intake diversionary system, anonymity, diminished stigma, shorter sentences, and recognition of rehabilitation as a viable goal." Irene Merker Rosenberg, *Leaving Bad Enough Alone: A Response to the Juvenile Court Abolitionists*, 1993 WIS. L. REV. 163, 184–85 (footnotes omitted); *see also*, *e.g.*, Daniel M. Filler & Austin E. Smith, *The New Rehabilitation*, 91 IOWA L. REV. 951 (2006); Ira M. Schwartz et al., *Nine Lives and Then Some: Why the Juvenile Court Does Not Roll Over and Die*, 33 WAKE FOREST L. REV. 533, 550–51 (1998).

exposure to a bounded scope of cases, it can actually undermine them. In courts that are meant to exercise leniency in sentencing, concentrating on only low-level offenses or more sympathetic offenders can lessen the contrast between them and more serious criminals and lead to unintended punitiveness.[197] When divisions are instituted to get "tough on crime," a biased sentencing curve means either too lenient sentences for very serious offenses[198] or exceedingly harsh sentences even beyond what is socially desired.[199] Either way, penal goals are better addressed through legislative and policy measures not through courts of limited jurisdiction.

Many specialized courts, however, do not merely punish differently; they also treat and rehabilitate, protect victims, and promote other social goals. Juvenile, veterans, and mental-health courts are meant to emphasize the rehabilitation and social integration of unique populations of offenders—and they can implement less adversarial procedures in order to do so. Domestic violence or elderly courts focus on especially fragile victims who are often the subjects of ongoing, not just past, victimization and, therefore, in need of protection beyond the adjudicative process itself. Judges in such courts often receive special training to better understand and

---

[197] *See*, *e.g.*, *supra* Section II.B.2 (discussing juvenile courts).

[198] For instance, scholars have criticized international criminal tribunals addressing mass atrocities for issuing sentences which are lenient both generally and when specifically compared to sentences ordered for murder, torture, and rape in domestic courts. Such comparisons are not confined to the U.S. but also incorporate sentencing practices in Canada and various countries in South America, Europe, Asia, and Africa. *See, e.g.*, Kevin Jon Heller, *A Sentence-Based Theory of Complementarity*, 53 HARV. INT'L L.J. 85, 118 (2012) (comparing the sentences of the International Criminal Tribunal for the former Yugoslavia, the International Criminal Tribunal for Rwanda, and the Special Panels for Serious Crimes in East Timor to those of domestic courts in twenty-two countries and finding "*international* sentences . . . excessively lenient"); Jens David Ohlin, *Towards a Unique Theory of International Sentencing*, *in* INTERNATIONAL CRIMINAL PROCEDURE 373 (Göran Sluiter & Sergey Vasiliev eds., 2009) ("When compared against sentences handed down in the United States for regular crimes, the sentences of international criminal tribunals are typically far lower, even though the crimes at these tribunals are far greater in both moral depravity and legal significance." (footnote omitted)). While international criminal jurisdiction may well be different for various reasons, such a trend is in line with the theory of relative judgments. International courts focused exclusively on mass atrocities (arguably those crimes at the very top end of the distribution of severity) displace their sentencing scales downward and issue more lenient sentences than those issued by domestic courts (exposed to a wider spectrum of criminal behavior) for equivalent, and even milder, offenses. For a similar account of "narrow" versus "broad" proportionality, see Alexander K.A. Greenawalt, *International Criminal Law for Retributivists*, 35 U. PA. J. INT'L L. 969, 983 (2014).

[199] Measuring this bias is more complicated because it is in the same direction as the intended punitive goals of the court. It is, nevertheless, no less disturbing. Both liberal punishment theory and the substantial social cost of incarceration call for punishment no greater than necessary to achieve the desired penal goals. Even when aiming to get tough on crime any sentence beyond what is required to achieve retribution, deterrence, social protection, and even denunciation cannot be morally or practically justified.

address the needs of such populations.[200] They also develop working relations with agencies dealing with offenders and victims of such crimes—such as mental health and welfare agencies or community organizations—to develop and implement the body of professional knowledge on how best to tailor the court's proceedings to its underlying goals.[201] Court specialization is often tied to other specialized resources—from victim and witness specialists[202] to social workers and probation officers trained with the particular population of offenders.[203]

Often, substantive and operational needs are also strongly intertwined. Some domestic violence courthouses, for example, are physically structured to enhance the protection awarded to victims and to minimize disorder at the court: they maintain secured waiting rooms and childcare areas and separate victim-only elevators.[204] Criminal domestic violence courts are sometimes connected to civil courtrooms to facilitate the issuing of orders of protection in cases where the defendant is released on bail.[205] In integrated domestic violence courts, the court's focus extends beyond just criminal cases with one judge handling all criminal domestic violence cases and related family issues such as custody, visitation, civil protection orders, and matrimonial actions.[206] Such an overarching focus on different types of legal proceedings and remedies improves victim safety and offender

---

[200] LABRIOLA ET AL., *supra* note 131, at 71; LANDIS, *supra* note 118, at 152–53, 169–70; Judith S. Kaye, *Delivering Justice Today: A Problem-Solving Approach*, 22 YALE L. & POL'Y REV. 125, 141 (2004).

[201] Community courts, drug courts, and domestic violence courts in New York, for example, employed a resource coordinator as a link between the court and relevant social services for the defendant. Kaye, *supra* note 200, at 132–33, 136, 142. Drug courts helped link defendants to services like job training, health care, education, and housing, *id*. at 136, and mental-health courts linked mentally ill offenders with treatment programs, *id*. at 137. Community courts also engaged with the local community through advisory boards and neighborhood newsletters. *Id*. at 133. In domestic violence courts, judges have been working collaboratively with prosecutors, defense attorneys, criminal justice agencies, social services, victims' advocates, and community organizations in order to develop court models and training programs capable of addressing varying needs in combating domestic violence. *Id*. at 142.

[202] Domestic violence courts, for example, are assisted by witness specialists trained with aiding child witnesses or victims of sexual assault. LANDIS, *supra* note 118, at 129–30.

[203] *Id*. at 139–41.

[204] See, for example, in Chicago, *Domestic Violence Courthouse*, CIRCUIT COURT OF COOK CTY., http://www.cookcountycourt.org/ABOUTTHECOURT/CountyDepartment/DomesticViolence/DomesticViolenceCourthouse.aspx [https://perma.cc/7Q4A-P2BB]. For a review of 208 domestic violence courthouses in the United States, see also LABRIOLA ET AL., *supra* note 131, at 50–51.

[205] LABRIOLA ET AL., *supra* note 131, at 50–52; LANDIS, *supra* note 118, at 47.

[206] See, for example, in New York, *Integrated Domestic Violence Court*, CTR. FOR COURT INNOVATION http://www.courtinnovation.org/project/integrated-domestic-violence-court [https://perma.cc/C9VA-ZMKX].

accountability, which is the main goal of the court, but it conflicts with a narrower focus on proportionality within the criminal justice system.

In such instances, a better balance to enable the desired separate treatment, while avoiding undesired biased punishments, could be to implement hybrid models of partial specialization. Some judges could be appointed to specialize in the subset of cases (based on their prior experience or receipt of relevant training), instead of dividing them across all judges, but that subset of cases would constitute only part of their caseloads while they continue receiving cases of all other types in the remaining part of their caseloads.[207] The exposure to a general caseload, alongside the specialized one, could improve the perceptions of relative severity between cases from the specialized docket and all other cases.

Another possibility would be implementing rotations of judges between the generalist and specialized dockets. If judges were assigned to specialized dockets only after gaining exposure to the general caseload at the criminal court, they could use their previous wider perspective when evaluating the cases of the specialized subject matter. If judges were removed from the specialized docket back to the generalist court after a limited time period, this could prevent the cumulative effect of limited exposure at the specialized docket over a long period of time.

In domestic violence, for example, different courts have implemented different approaches. The first municipal district in Cook County, Illinois, implements full specialization of judges under the domestic violence division. The second municipal district uses a model of partial specialization where one judge hears domestic violence cases in addition to her general caseload. In New York, judges rotate between the criminal and the domestic violence divisions. Through mixed organizational design—like partial concentration and rotations—some of the benefits of specialization can be achieved while the exposure to other types of cases can mitigate the contextual influences that exist when handling only narrower categories of cases of lower gravity.

Hybrid solutions admittedly encompass some trade-offs as well. They might not always be practical and at times could still hinder some of the underlying unique reasons for specialization. The more unique and extensive the required training is, the less likely such hybrid solutions are to work efficiently. Rotations, for example, have been highly criticized by

[207] Baum draws an analytical distinction between "concentration of judges" and "concentration of cases" and acknowledges that "a court or an individual judge can be high on one dimension and low on the other." BAUM, *supra* note 1, at 6–10. Applying that terminology here, this means that there could be specialized dockets (i.e., concentration of cases in one court or before particular judges) but that judges would not focus solely on those cases (i.e., there would not be complete concentration of judges).

stakeholders at domestic violence courts for hindering judicial expertise and consistency of policy and, by that, undermining the unique policy goals sought.[208] But sometimes hybrid solutions can still offer an improvement over complete specialization, and their adequacy should be considered on a case-by-case basis.

IV. STANDARDIZING THE PUNISHMENT CURVE

The concern of nonrepresentative caseloads affecting determinations of proportionality and sentences highlights the importance of judicial exposure to cases and sentences across the entire spectrum in the relevant jurisdiction. As evident by now though, changing the relevant scale of cases evaluated through institutional design can only partially help in overcoming this bias.[209] Often, there are good reasons for specialized dockets that lead to a difficult balancing between the costs and benefits of such courts. Sometimes even careful initial attention can be later jeopardized by amendments and future changes in legislation or staffing of the courts. Some forms of limited jurisdiction—such as geographic divisions—are simply inevitable.

If the bounded scope of comparison used by judges leads to punishment curves that are shifted in a biased way, the solution is reshifting the punishment curve, and perhaps this can be achieved by providing judges with statistical information about sentences in cases beyond those they encounter firsthand. If a juvenile court judge compares a particular defendant only to other juveniles, exposing the judge to the full distribution of sentences imposed for the same offense by the adult courts can help her compare sentencing levels for adults and juveniles and rescale her sentencing decisions.

I call providing judges with that information "curving discretion." As I illustrate below, statistical curving will require policymakers or judges to first make straightforward and transparent choices about the relevant scale for comparison—similar offenses, the same charge, or even narrower categories defined by aggravating or mitigating circumstances—as well as choices about the relevant geographic unit. Then judges would determine the factual circumstances of the case that are relevant for sentencing similar to what they do today.[210] But instead of those circumstances only producing

---

[208] LABRIOLA ET AL., *supra* note 131, at 71–72, 77.

[209] *See supra* Part III.

[210] Such factual findings will remain, as they are today, largely dependent on the facts presented to judges by the parties. Facts that are not brought before the court obviously cannot be considered. This, however, does not limit the proposal in any way beyond the current practice. The facts that are presented—and that serve as the basis for calculating guidelines-recommended sentences—will also

a recommendation for the sentencing *range*, they would also produce a graph illustrating the sentences imposed for similar cases. Judges could also produce different graphs based on different considerations or for a slightly more or less serious charge. Armed with that information, sentencing decisions would reflect the relative severity of the particular case—in comparison to all other cases not just those in the judge's caseload.

Section IV.A develops the proposal for curving discretion. Section IV.A.1 explains how statistical curving can assist in standardizing sentences across trial courts and judges. Section IV.A.2 empirically illustrates and discusses the useful information that can be produced by curving discretion using examples from different geographic jurisdictions. Section IV.A.3 suggests an additional benefit of curving discretion: it can also enhance appellate review of sentences as an overarching mechanism across different courts.

Because statistical curving is a proposal for social design, it raises normative, constitutional, and institutional-competency questions. After the detailed explanations of how statistical curving could be implemented, Section IV.B addresses these policy questions.

A. A Proposal for Curving Discretion

1. Statistical Curving of Judicial Discretion.—Statistical curving can help to overcome sentencing disparities in two important ways. First, it can mitigate relative-judgments bias by changing the comparison group used to assess cases and by reintroducing forgotten dimensions into the sentencing decision. Second, it can more generally reduce disparate outcomes (regardless of their source) by providing judges better information on the sentences ordered in other cases. In that sense, curving discretion can further prove useful to solve part of the idiosyncrasy problem in judicial decisions.[211] Judges in several states, such as Pennsylvania and Minnesota, are already using computerized case-processing systems;[212] those systems

serve as the basis of the punishment curves. It offers an improvement over current practice in that such facts could be linked to more informative and granular sentencing recommendations.

[211] For an analysis of the general problem of lack of calibration in legal judgments across decisionmakers, see Frederick Schauer & Barbara A. Spellman, *Calibrating Legal Judgments*, J. LEGAL ANALYSIS, Nov. 8, 2016, at 1. For a discussion of how sentencing-information systems could be used to improve transparency and comparability in sentencing decisions, see Marc L. Miller, *A Map of Sentencing and a Compass for Judges: Sentencing Information Systems, Transparency, and the Next Generation of Reform*, 105 COLUM. L. REV. 1351 (2005).

[212] *See* Miller, *supra* note 211, at 1363–66.

could be extended, as was done in other countries,[213] to include statistical sentencing information as well.[214]

a. The relevant universe of cases.—Statistical curving would first require policymakers to make a straightforward choice about the relevant universe for comparison. The decision entails two main dimensions: a substantive element about the types of offenses or offenders to be compared and a local element about the relevant geographic unit.

The substantive element requires a deliberate decision about the relevant universe of offenses or offenders for equitable treatment, a decision that depends on the particular subject matter, the uniqueness of cases, and the policy goals that the courts aim to achieve. For juvenile and adult courts, for example, the substantive decision entails recognizing that the relevant comparison should be based on the particular offense across all offenders. Only in such a way can the offender's personal characteristics, including the dimension of being a juvenile, be properly taken into account. As a result, the decision about the sentencing curve would be that: (1) it should portray the sentences ordered in both adult and juvenile courts and (2) juvenile court judges should adjust their sentences in a way that will maintain harmony and proportionality between them and the sentences ordered for adults and juveniles at the criminal courts. There is no dispute that juveniles should be sentenced more leniently than adults should. The forced comparison between the sentence of juveniles and adults across courts could help ensure actual sentences indeed follow that desired policy.

When cases are determined to be unique, as discussed in Section II.C.3 above, the decision might go the opposite way. For domestic violence cases, for example, the societal stance is that what makes such cases severe is not the magnitude of physical harm caused but the creation of an abusive domestic environment. When domestic violence cases are heard by a general criminal court, the concern is that the comparison to assault and battery cases would make the dimension of physical harm too salient and lead to underestimating the uniqueness of the domestic aspect. With that policy goal in mind, when domestic violence offenses are heard by both specialized and generalist courts, the decision about the punishment curve would entail: (1) aggregating the sentences at both domestic violence and criminal courts and (2) ensuring that judges at the

[213] *Id*. at 1371–75 (discussing successful sentencing-information systems used in Scotland and New South Wales, Australia).

[214] *Id*. at 1380–81 ("[I]n most systems, much of the information that would serve as the basis for [a sentencing-information system] is already collected and analyzed; it is just not shared, or if shared, it is not shared in a form that judges and others . . . can use . . . .").

criminal courts adjust their sentences based on the sentencing levels developed by the specialized judges in domestic violence courtrooms.

The geographic element of the decision requires policymakers to decide what is the relevant local unit for the comparison of actual sentences—the district courthouse, the county, or the state. Different states might reach different conclusions on this question, but what is important is that they address it straightforwardly.[215]

The decision about the relevant universe of cases for comparison—and displaying information about sentences in all such cases through statistical curving—could help to mitigate the relative-judgments bias in judicial sentencing decisions. Making the necessity of such a comparison an explicit stage of the sentencing process can reintroduce neglected factors into the sentencing decision. If the dimension of being a juvenile might lose its diagnostic value in a comparison across juveniles, the forced comparison between juvenile and adult sentences makes it extremely relevant again. The regained salience of such relevant considerations for sentencing would ensure they are given adequate weight in the final decision of the court. The explicit comparison across cases of varying degrees of severity can also rescale the relative evaluations of severity onto a standardized scale. If the within-sample case evaluation were based on the relative severity of the juvenile case against the other cases in the juvenile court, mapping the case onto the joint distribution of sentences for all offenders would make salient the lesser severity of a juvenile compared to an adult.

The exercise of mapping particular cases against the universe of relevant conduct will also serve to inform judges. It will help judges become aware not only of the existence of disparities but also of their sources and of the fact that they are truly unwarranted. Currently, even where judges are aware of the existence of sentencing disparities, disparities prevail because judges often either rationalize them or accept them as based on differences of opinion. They think either those different courts have justified reasons to sentence differently or that other judges are wrong.[216]

---

[215] Policymakers, for example, could be sensitive to the varied demographic or socioeconomic conditions across jurisdictions that are correlated with the development of different sentencing patterns or to which central prison defendants are being sent. Even when disparate sentencing curves across courts persist, the geographic element can account for the differences in the geographic domains of trial courts and jail or prison facilities, and parole boards could use the different curves to adjust sentences and to enable better comparisons across prisoners for early-release decisions.

[216] During my research for this Article, I talked to many state and federal judges. Overwhelmingly common to all judges was how quickly the issue of geographic disparities surfaced. Practically all judges recognized that different district courthouses sentence differently. And all judges referred to

Educating judges about the bias underlying the problem can help change behavior by questioning the current underlying justifications. An explicit policy decision that all cases within a certain universe should be treated equitably—and that the universe includes cases from all courts—can induce judges to move from explaining the differences to changing them. Providing the curve according to which sentences should be standardized will offer guidance on exactly how to do so.

b. Addressing disparate outcomes.—Statistical guidelines can address the bias introduced by relative judgment in an additional way: targeting its disparate outcomes directly. If relative judgments lead to sentencing disparities, we can address them similarly to how we would address any other kind of disparities regardless of their sources. Like standardization in other fields, statistical curving can curb sentencing disparities by standardizing the curve on which cases are evaluated.

Statistical curving offers a metric to evaluate trial judges' sentencing behavior and thus can help the judiciary develop and implement norms of uniformity from within. Court actors are usually immersed in the local caseload, but statistical curving will make the decisions by other judges and courts more accessible and will be more informative about the trends they represent.[217] Judges are also likely averse to being outliers—if they are concerned about not being retained (for appointed judges) or reelected (for elected judges), they might not want to be viewed as excessively harsh or lenient. Appellate review may have a greater bite once appellate courts are better informed about how other judges exercise their discretion.[218]

Further, because the punishment curve is generated for cases with similar circumstances, statistical curving can promote better reasoning. The reasons provided by judges as relevant considerations in sentencing will

such differences along at least one of these two lines of reasoning. The first is that the different sentencing levels developed at different courts are the result of different needs or goals. Different district courthouses, the explanation goes, cater to different constituents, or deal with different local pressing needs and, therefore, reach different outcomes. When such substantive justifications do not exist, judges often think the judges at the other court are simply wrong. Judges at suburban districts facing relatively low crime levels acknowledged that their sentencing levels are much higher than those of a neighboring urban district—one judge even facetiously wondered why any offender would choose to commit a crime only a few blocks into their jurisdiction. Their explanation was that while the urban court is overwhelmed with cases, they send the proper message that crime does not pay. Judges at larger urban districts observed the same trends but offered an opposite explanation: the suburban judges don't see "real crime," they said, and therefore do not realize how petty the things they punish so harshly are. They, at the urban court, see all crimes, and they know what is truly important.

[217] For a discussion of the relationship between sentencing guidelines and statistical curving, see *infra* Section IV.B.1.

[218] For the desirability of appellate review as an overarching mechanism in the post-*Booker* era, see *infra* Section IV.B.2.

affect the curve that is produced, and judges therefore will benefit from providing better and more detailed reasons for their decisions. When a judge decides to impose a sentence different than the sentence generally ordered for similar cases by other judges, the need to justify the particular deviation from the majority view of the courts would also increase. Because such reasons can only include permissible considerations in sentencing, curving discretion can also mitigate the weight granted to impermissible considerations and, by that, help more generally with overcoming other biases in sentencing decisions.

The focus of this Article is on the impact of caseload composition on judicial sentencing decisions, and it focuses on that feature because the caseload effect has not gained scholarly (or practically any) attention to date. The existence of other biases in sentencing, however, and of interjudge sentencing disparities, is widely recognized. Curving discretion can more effectively address some of those issues and at least partially solve part of the idiosyncrasy problem in judicial decisions.

2. Illustrating Statistical Curving.—To illustrate what curving discretion might look like, this Section offers an empirical example, comparing the punishment curves across courts of different geographic jurisdiction. While no division of jurisdiction is completely arbitrary, geographic jurisdiction is perhaps the closest. This is also an area where at least the rhetoric around sentencing guidelines is that sentencing disparities are unwarranted. And both penal codes and sentencing guidelines are promulgated at the state level to be applied by local courts. For these reasons, it is a useful example to illustrate how statistical curving can be used, if we want to harmonize sentencing levels across courts.

Section IV.A.2.a reviews the data. Section IV.A.2.b discusses the intra- and interdistrict differences that are revealed when comparing actual sentences imposed to the sentences recommended by the guidelines. Section IV.A.2.c explains how to construct the sentencing curve that relates to the sentencing guidelines range and how it is affected by the decision about the relevant geographic unit for comparison. Section IV.A.2.d extends the uses of sentencing curves to comparisons across offense-severity levels and guidelines categories.

This Section's aim is illustrative, not analytical. The data available for researchers is obviously less granular than the information before the judge hearing the case. Therefore, the analysis is not meant to suggest what sentence should have been imposed in any particular case nor to identify what sentencing levels should be for any offense. I also do not make any causal claim regarding the source of the observed disparities. The analysis is only meant to illustrate how such a model would work. In practice,

statistical curving could be based on any information available to the judge, including anything currently used for calculating the guidelines range as well as any other relevant circumstances of the offense and the offender.[219]

a. Data.—I use sentencing data from the Pennsylvania Courts of Common Pleas for the years 2001–2012.[220] Pennsylvania is particularly suited for the purpose of this analysis due to its sentencing guidelines regime. Because Pennsylvania uses a sentencing guidelines matrix, it offers the opportunity to compare the type of guidance for judicial discretion that is offered under the current guidelines with the one that could be achieved using statistical curving. Pennsylvania's guidelines have been advisory throughout the period of study[221]—as is now the regime in the federal courts and in all states that use sentencing guidelines, following the Supreme Court's decision in *United States v. Booker.*[222]

Pennsylvania has an indeterminate sentencing system[223] with sentencing guidelines setting a range for the minimum sentence, stated in months of incarceration. The guidelines' instructions for the minimum sentence are similar to those under the federal system. The recommended sentencing range is the result of two scores assigned by the guidelines for each offense[224]: (1) an "Offense Gravity Score" (OGS), which takes into account the gravity of the offense of conviction and ranges from one to fourteen,[225] and (2) a "Prior Record Score" (PRS), which weighs the seriousness and extent of the offender's prior criminal record and is divided

---

[219] Miller, *supra* note 211 (explaining how sentencing-information systems can include a wide array of sentencing considerations beyond those covered by the guidelines, including offense- and offender-specific circumstances and even punishment goals).

[220] The sentencing recommendations for the examined offenses remained the same under the different guidelines editions over the years. *Sentencing Guidelines and Implementation Manuals,* PA. COMM'N ON SENTENCING, http://www.pcs.la.psu.edu/guidelines/sentencing/sentencing-guidelines-and-implementation-manuals [https://perma.cc/JUB9-GFV4].

[221] Under advisory guidelines, judges retain discretion to impose a sentence outside the guidelines range subject to mandatory minimum and statutory maximum sentences when such apply. If a judge deviates from the guidelines range, she must specifically state the reasons for departure in a written statement. Sentences inside the guidelines also pose a heightened burden on the defendant on appeal compared to sentences outside the guidelines range. 42 PA. CONS. STAT. § 9781 (2016).

[222] 543 U.S. 220 (2005).

[223] When imposing a sentence of partial or total confinement, Pennsylvania requires the court to impose both a minimum and a maximum sentence. 42 PA. CONS. STAT. §§ 9755–56 (2016). The sentencing guidelines and most mandatory sentencing provisions address only the minimum sentence, and that is my focus here as well. The range between the minimum and maximum sentence is a part of the sentence to be decided later by the parole board.

[224] 204 PA. CODE § 303.2 (2015).

[225] *Id.* § 303.15. Except for murder in the first and second degree, which is prescribed either a mandatory life sentence or capital punishment. 18 PA. CONS. STAT. § 1102 (2016).

into eight categories.[226] Based on a matrix of OGS and PRS combinations, the sentencing guidelines categorize offenses into five levels. Within each level, the guidelines prescribe three ranges: (1) a standard range, for use under normal circumstances, (2) an aggravated range, for use when the judge determines that there are aggravating circumstances, and (3) a mitigated range, for use when the judge determines that there are mitigating circumstances.[227]

For the purposes of this illustration, the analysis focuses on one group of cases with circumstances that are as similar as possible. The category of cases that has both a large number of cases for such a comparison and sufficient granularity about particular circumstances is drug cases. Among drug cases, the largest sample for cases that routinely result in substantial periods of incarceration is for those involving possession of cocaine with intent to distribute. To guard against the impact of other sentencing provisions that may override the sentencing guidelines' recommendations, I exclude cases that carry a youth, school, or gun statutory enhancement or a mandatory minimum sentence, as well as cases where the defendant was found to need treatment for drug or alcohol dependence.[228]

The sentencing recommendation for a case is determined based on the intersection of the OGS and the offender's PRS. For the main analysis, I focus on one guidelines category at the intersection of (1) the offense of possession of cocaine with intent to distribute at least 2.5 to less than 10 grams, a felony with an OGS of 7,[229] and (2) defendants with a PRS of level 2.[230] An amount of 2.5 to 10 grams is, indeed, a relatively wide range. However, that is the quantity range currently included under the same guidelines category. I use the same quantity range in the main analysis in

---

[226] The categories, by order of increasing severity, are: No prior record (0), five categories based on the number and severity of prior record (1–5), Repeat Felony Offender (RFEL), and Repeat Violent Offender (REVOC). 204 PA. CODE § 303.4 (2015).

[227] *Id*. § 303.16(a).

[228] For the same reason, I also drop a few cases where the recorded guidelines sentencing range differed from the recommended range under the guidelines.

[229] The three mildest quantity categories for the offense under the guidelines have the highest caseloads by a large margin (with between 3,775 and 12,856 cases in each category compared to 98 to 874 cases in the other three categories). From those three, the main analysis focuses on the middle category (2.5–<10 grams) in order to allow for cross-category comparison with the less and more serious categories in Section IV.A.2.d.

[230] The three largest prior record categories for this offense are no prior record (n = 2,432), a prior record score of 2 (n = 899), and a prior record score of 5 (n = 1,015). For first-time offenders, the likelihood of incarceration is lower as the recommended sentencing range includes zero incarceration, and the focus of the graphic presentation is on sentences of incarceration. For "career criminals," sentences are substantially affected by the defendant's prior conduct rather than mainly driven by the offenses in question. I therefore focus on the category of OGS = 2. Results are similar for different categories.

order to illustrate the difference between the current guidance given to judges—i.e., that guidelines range—and the guidance that statistical curving could offer. But I also show how sentences are divided based on finer gradations of quantity. The same patterns persist when looking at different offense categories and at offenders with varying degrees of prior record.

b. The difference between recommended and actual sentences.—The recommended sentencing range for an OGS of 7 and a PRS of 2 is twelve to eighteen months.[231] If the judge finds that mitigating or aggravating circumstances exist, she may reduce or increase the sentence by up to six months, resulting in an overall sentencing range of six to twenty-four months.[232]

Philadelphia is the most populous county with the largest caseloads. Neighboring Philadelphia is the midsized Delaware County. A little further, still within the eastern district of Pennsylvania, is the smaller Lehigh County.[233] During the period of the study, 56 judges in the three counties sentenced 380 such cases—with individual dockets ranging from 1 to 70 cases.[234] Figure 1 displays the distribution of actual sentences ordered in those cases.

As is evident from Figure 1, the range of sentences actually imposed is confined within the guidelines' recommended sentencing range only in Lehigh. In both Philadelphia and Delaware, the range of actual sentences is much wider than that recommended by the guidelines. While the guidelines' recommended range for sentencing (even the extended range) is six to twenty-four months, actual sentences range from zero to thirty-six months in Delaware and from zero to sixty months in Philadelphia. Granted, differences across cases can result from differences in unobservable case characteristics, not just from divergence in judicial propensities to punish. Still, the distributions shown in Figure 1 exhibit greater variance than what is implied by the guidelines regime. Actual sentences are spread over thirty-six- to sixty-month intervals—200% to 330% wider even in comparison to the recommended *extended* range of eighteen months (between six to twenty-four), which incorporates aggravating and mitigating circumstances. This, remember, is for cases that, at least from observable data, do not include what legislators

[231] PA. CODE § 303.16.

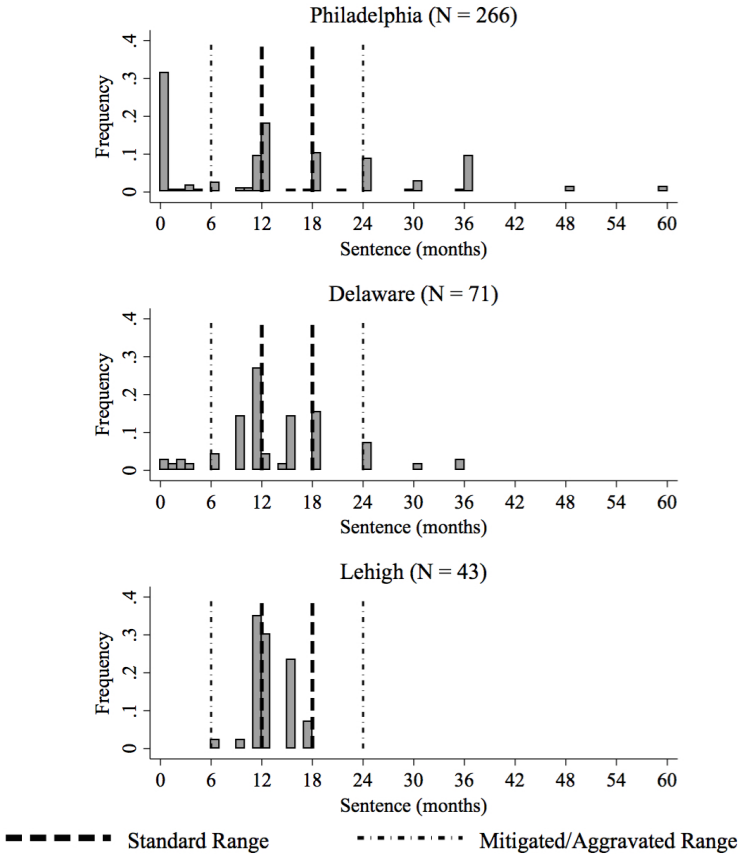
[232] *Id.*

[233] These are the counties with the largest caseloads for the offense category that are within the same district.

[234] Thirty-three judges in Philadelphia, fifteen judges in Delaware, and eight judges in Lehigh.

recognize as extreme circumstances that justify different sentencing schedules. These are cases that do not involve circumstances that trigger statutory enhancements or mandatory minimum sentences, and these are defendants who were not referred to treatment for drug or alcohol abuse.

FIGURE 1: DISTRIBUTION OF SENTENCES, BY COUNTY



Note: Figure 1 presents the distribution of sentences for the offense of possession of cocaine with intent to distribute 2.5–<10 grams (OGS = 7) for defendants with a prior record score of 2. Each bar represents the proportion of sentences within the one-month interval.

The cross-county comparison also reveals the great local discrepancies in sentencing levels. Philadelphia County has the most cases in that category and high case volume in general, and the distribution of sentences is highly skewed with 30.5% of cases not sentenced to incarceration at

all.[235] In the smaller counties, however, completely avoiding incarceration is rare: that is the disposition only in 1.4% of cases in the midsized Delaware County[236] and not even in a single case in Lehigh County. The sentencing distributions extend over a narrower range in the smaller counties—twelve months in Lehigh and thirty-six months in Delaware—compared to the sixty-month range in Philadelphia. They are also much more bell shaped—with the mode of the distribution (the most frequent sentences) lying within the recommended guidelines range—potentially suggesting that judges in courts with smaller caseloads implement less variance (and fewer within-category distinctions) in their sentencing decisions.

c. Constructing the sentencing curve.—If different geographic jurisdictions develop different sentencing levels, the decision about the relevant unit for jurisdiction can change the sentencing curves used by judges and provide guidance to judges about desired sentencing levels overall.

As a first step, suppose each county produces its own sentencing curve—similarly to how sentencing levels are locally developed today. From the distribution of actual sentences presented in Figure 1, it is possible to calculate the density curve of sentences and construct the sentencing curve used by the judges in each county, as displayed in Figure 2.[237] Those sentencing curves can serve as the county's statistical-curving recommendation.

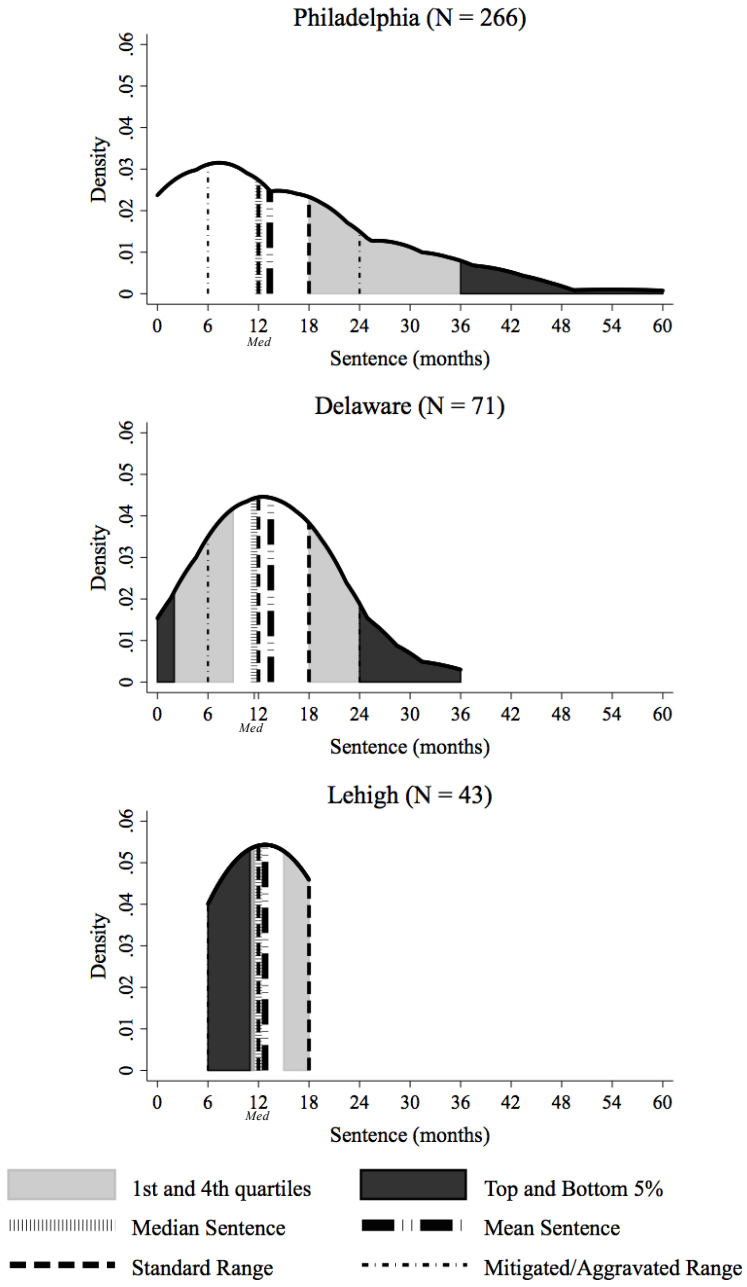
The cross-county comparison reveals one way in which the current guidelines may serve as focal points for sentencing: the median sentences in the counties tend to cluster around the lower end of the guidelines range and equal about 11.5 to 12 months in all counties. The average sentence is in the lower half of the standard sentencing range and equals between 12.8 and 13.5 months. But the comparison also reveals the insufficiency of such focal points to harmonize sentencing levels.

[235] 81 out of 266 cases.

[236] In only 1 case out of 71 there was no incarceration sanction.

[237] I use kernel density curves with a width of six months. Because, as is evident from Figure 1, sentences tend to cluster around six-month figures (at months six, twelve, eighteen, etc.), the choice of width enables displays of smoother curves and eases the visual comparison. The choice is esthetic, not substantive, and using narrower widths (or even the histograms themselves) would not change the comparisons drawn.

FIGURE 2: SENTENCING CURVES, BY COUNTY



Note: Figure 2 presents the kernel density curves that correspond to the histograms in Figure 1.

The standard guidelines' sentencing range is twelve to eighteen months. Yet in both larger counties, only one-third of cases are actually sentenced within it,[238] compared to 60% of the cases in Lehigh County. In most cases, different counties end up with very different sentencing outcomes and depart from the guidelines ranges at very different rates. Downward departures are infrequent in Delaware, where they occur in 8.5% of cases, and nonexistent in Lehigh. But they occur in 33.8% of cases in Philadelphia. Departures above the guidelines in Philadelphia are four times more likely than in Delaware and did not occur in even one case in Lehigh.[239]

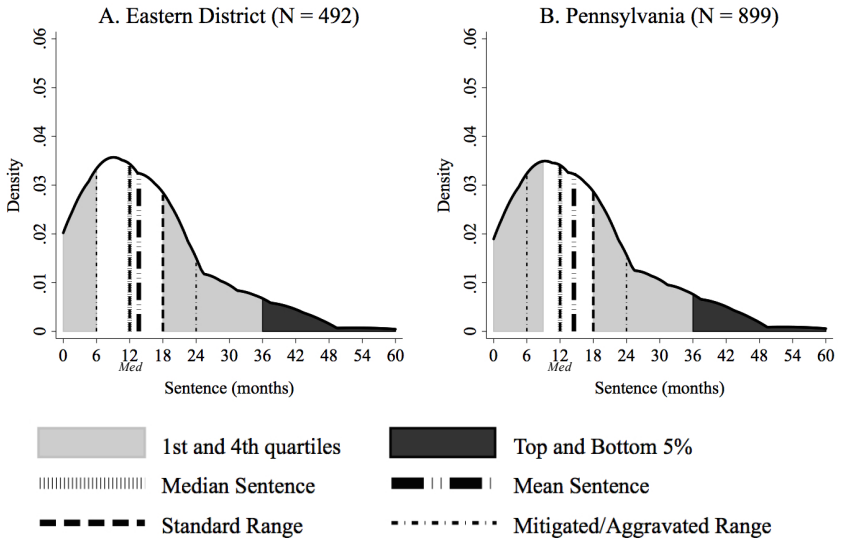
The inter-county differences are manifested not only in the treatment of extreme cases at the ends of the distribution but also in the full range of sentences applied by the courts. The width of the interquartile range varies by county by 600%—from three to eighteen months. Sentencing levels relative to the guidelines' recommendations also vary widely. In Lehigh, the interquartile range accommodates only the lower half of the standard sentencing range. In Delaware, the interquartile range is wider and includes both the standard and part of the mitigated range. In Philadelphia, the interquartile range includes the standard range, the mitigated range, and below it.

If such disparate sentencing practices across counties are unwarranted, policymakers might wish to consolidate the sentencing curve across the different counties. The relevant jurisdiction could include, for example, all cases in the eastern district of Pennsylvania (Figure 3.A) or be wider to spread across the state of Pennsylvania as a whole (Figure 3.B). As the chosen area of jurisdiction widens, the curve averages the differences across extreme counties and follows more closely the sentencing guidelines' recommendations. As can be seen from Figure 3, the interquartile range based on the Pennsylvania sentencing curve resembles more closely the guidelines' recommendations: it extends from nine to eighteen months—slightly wider than the standard recommended sentencing range of twelve to eighteen months but still within the mitigated range above six months.

[238] In Philadelphia, 29% of cases (77 out of 266 cases) and in Delaware, 35% (25 out of 72).

[239] Sentences higher than twenty-four months are ordered in 16% of cases in Philadelphia (43 cases) in comparison to 4.2% of cases (3 cases) in Delaware.

FIGURE 3: THE JOINT SENTENCING CURVE



Note: Figure 3 presents the kernel density curves for the eastern district (Panel A) and for the State of Pennsylvania (Panel B) for the offense of possession of cocaine with intent to distribute 2.5–<10 grams (OGS = 7) for defendants with a prior record score of 2.

Notice that because the statistical curving is based on the actual sentences imposed, it does not necessarily impose rigidity or narrower categories. In the current example, the interquartile range is three months wider than the standard sentencing range recommended by the guidelines. According to the Pennsylvania sentencing curve, 90% of the cases will be sentenced between zero and thirty-six months—allowing for a range of discretion twice as wide as that proposed by the guidelines' extended range (six to twenty-four months). But that discretion could be applied more consistently across the districts—emphasizing variation that is rooted in the circumstances of particular cases, not in local sentencing levels.

The fact that the curve is not identical to the sentencing range recommended by the guidelines is not unexpected. Any biases that currently exist are incorporated into it. Theoretically, the sentencing curve does not necessarily need to be completely aligned with the sentencing guidelines' recommendations. If, for example, many courts display outright policy disagreement with the guidelines recommendations (as they may under *Kimbrough v. United States*[240] and *Spears v. United States*[241]), the sentencing curve will reflect the same disagreement.

[240] 552 U.S. 85 (2007).

The conceptual desirability of such curving does not necessarily depend on either the width or the location of the range. The point is that statistical curving based on how judicial discretion was applied in actual cases gives judges flexibility to develop sentencing levels while at the same time facilitating coherence across judges.[242] To the extent that existing variation is rooted in too little guidance from the guidelines, this variation will gradually decrease as judges move their sentencing choices closer to each other. To the extent variance does not decrease, judges should then evaluate their sentencing choices based on the real "going rates" because those are clearly only loosely tied to the recommended guidelines range.[243] Importantly, the decision on the relevant level of jurisdiction as it is currently implemented already entails a decision about the sentencing levels that will be developed in the courts. Statistical curving does not change that in any way; it simply makes it transparent.

d. Levels of comparison.—Curving discretion can also enable judges to take into account the impact of different considerations in sentencing. This exercise can facilitate coherence both across and within guidelines categories. Across guidelines categories, it will assist judges in evaluating the relationships between the actual sentences associated with increasing offense gravity scores or prior record scores. Within categories, this will better enable judges to standardize a particular sentence against the distribution of overall sentences in the category.

Figure 4 depicts how sentencing curves can be used to compare the overall impact of offense gravity on sentencing outcomes across the guidelines categories.[244] Figure 5 illustrates the use of curving to create more granular distinctions within offenses by displaying the sentences for possessing different amounts of cocaine in 2.5 gram intervals. Panel A (less than 2.5 grams) corresponds to an OGS score of 6. Panel E (at least 10 and

---

[241] 555 U.S. 261, 264–66 (2009) (per curiam).

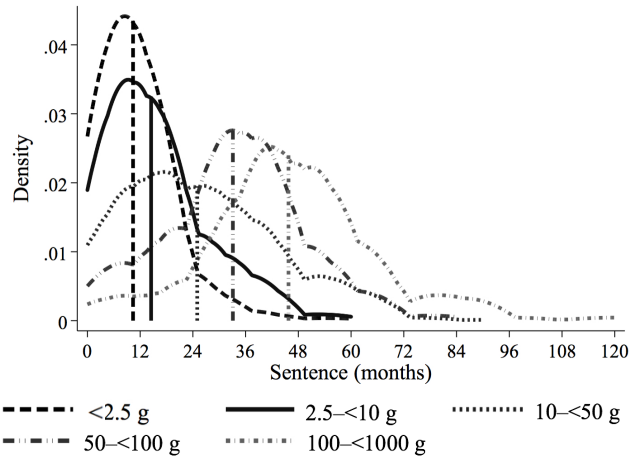
[242] I discuss the practical and normative aspects in the division of powers to determine sentencing levels across judges and legislators further below. *See infra* Section IV.B.3.

[243] Similarly, punishment curves can be useful whether the sources of observed disparities are plea bargaining by prosecutors in the shadow of the judge or, perhaps, judicial decisions in the shadow of prosecutors. If the source of variation is judge-based disparities, then statistical curving could better fulfill the current role aspired to by the sentencing guidelines. If the variance is driven by plea bargains, then exposing the real "going rates" supports providing judges with similar power to that of prosecutors.

[244] Different guidelines categories recommend different sentencing ranges, and to simplify the graphic presentation, Figure 4 displays only the sentencing curves and average sentences for each category. Each weight category is associated with a higher offense gravity score of 6, 7, 8, 10, or 11. The relevant associated guidelines categories for a PRS of 2 are in order of increasing severity: 9–16 (±6) [n = 1909], 12–18 (±6) [n = 899], 15–21 (±9) [n = 488], 36–48 (±12) [n = 122], and 48–66 (±12) [n = 126].

less than 12.5 grams) is carved from cases charged for possession of 10–<50 grams, an offense with a gravity score of 8. Panels B, C, and D divide the possession of at least 2.5 and less than 10 grams analyzed in Sections IV.A.2.b–c above (OGS = 7) to three equal-weight categories.

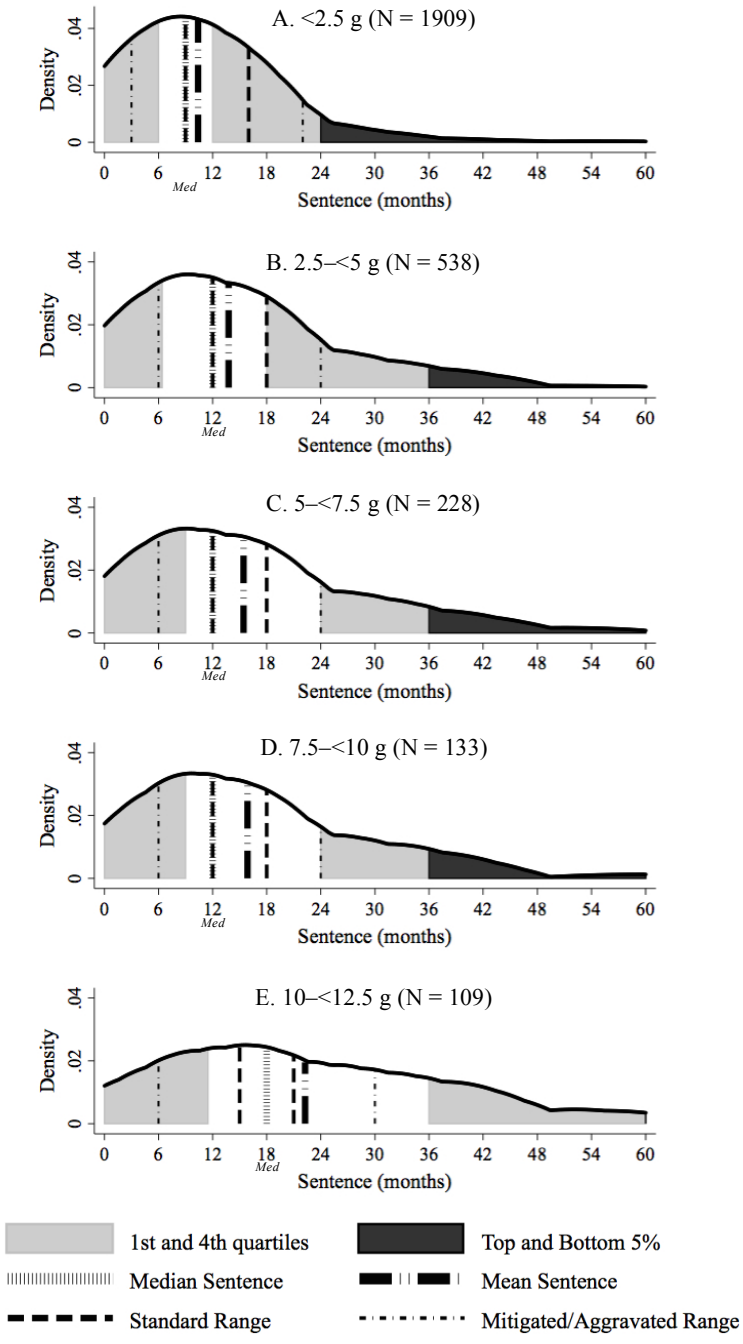
FIGURE 4: SENTENCING CURVES FOR PENNSYLVANIA, BY OGS CATEGORIES OF DRUG QUANTITY, PRS=2



Note: Each curve represents the kernel density curve for the distribution of sentences for each OGS–PRS combination for all cases in Pennsylvania. The solid curve corresponds to the curve in Figure 3.B. Vertical lines mark the mean sentence ordered in each guidelines category.

The comparison demonstrates the two problems statistical curving can answer. First, even across distinctive categories, great sentencing variance exists, and actual sentencing ranges largely overlap (Figure 4). Second, the weight given to a particular circumstance—drug weight—crucially depends on the amount of guidance given by the sentencing guidelines (Figure 5). Where the guidelines draw a clear line across drug quantities, average sentences increase with drug quantity. But within the same guidelines category, higher drug quantity barely affects the average sentence.

FIGURE 5: SENTENCING CURVES, BY DRUG QUANTITY, PRS = 2



From Figure 5, for possession of less than 2.5 grams, the median and mean sentences are 9 and 10.4 months, respectively. In comparison, for possession of between 2.5 and 5 grams, the median and mean sentences are 3 months longer (12 and 13.8 months, respectively). Similarly, for possession of between 10 and 12.5 grams, the median and average sentences are 18 and 22.7 months, respectively, in comparison to 12 and 15.9 months, respectively, for possession of between 7.5 and 10 grams—a difference of 6 to 7 months.

While intervals of 2.5 grams play a large role in sentencing across guidelines ranges, within the recommended range for possession of at least 2.5 and less than 10 grams, changes in drug quantity have at most a modest effect on sentencing. The average sentence is 13.8 months for possession of at least 2.5 and less than 5 grams, 15.4 months for possession of at least 5 and less than 7.5 grams, and 15.9 months for possession of at least 7.5 and less than 10 grams.

Even the source of that variation is mainly driven by extreme decisions, and across all three categories, the median sentence equals twelve months. On the mildest end of the sentences, possessing more than 7.5 grams decreases the probability of avoiding incarceration completely by one-fifth: 16.4% and 15.4% of cases involving possession of 2.5 to 5 and 5 to 7.5 grams, respectively, are not sentenced to incarceration, compared to 13.5% of the cases involving possession of 7.5 to 10 grams. On the upper end of the sentences, the probability of imposing a sentence of thirty-six months or higher increases by one-fifth—from 10% to 12% to 15%—as drug quantity increases. After excluding cases with sentences equal to 0 or to 36 months and higher, even the small gap in sentences practically disappears, and average sentences are 13.5, 14.5, and 13.9 months, respectively. Similar patterns persist when looking at other offense categories.

Taken together, the effect illustrated by the guidelines fits with a theory of relative judgments: across guidelines categories, where different sentencing recommendations exist, a difference of 2.5 grams in the quantity of cocaine possessed leads to substantial differences across all cases, as reflected in the average and median sentences. Within a guidelines category, absent more granular guidance, a difference of 2.5 grams in drug quantity leads to a different outcome for extreme cases at the ends of the distributions but barely creates any differentiation across the bulk of cases within the range.

This result should be normatively perplexing. Guidelines cutoff points are largely arbitrary—there is no substantively prevalent reason to draw a line exactly at 2.5 or 10 grams instead of at 5 grams. From a normative

perspective, "uniformity does not consist simply of similar sentences for those convicted of violations of the same statute . . . . It consists, more importantly, of similar relationships between sentences and real conduct."[245] Yet, actual sentences reflect a differentiation that is not based on real conduct but rather on the statutory veil of that conduct. A potential advantage of statistical curving is, therefore, that it can create continuous differentiation within the ranges given by the penal code or guidelines matrix. That way it can better address the desire for proportionality across cases of increasing severity and overall coherence in sentencing.

3. The Overarching Role of Appellate Review.—The specialized–generalist spectrum can also contribute to assessing the role of appellate review in harmonizing sentencing levels. Appellate courts, after all, are meant to serve as an overarching mechanism—overseeing the decisions of all judges and trial courts under their jurisdiction, vacating extreme decisions, and providing guidance to trial judges on how to exercise their discretion.[246] Notably, appellate courts are also generalists. And as trial courts' jurisdiction becomes increasingly narrower, the generalist nature of appellate courts gains greater importance.

If the institutional structure of trial courts leads to biased sentencing curves, this provides an additional justification for increased appellate review of sentences. Statistical curving can facilitate such appellate review. In Pennsylvania, sentences within the guidelines range are appealable only if they "involve[] circumstances where the application of the guidelines would be clearly unreasonable,"[247] and sentences outside the guidelines can be vacated on appeal if "the sentence is unreasonable."[248] So departures are appealable, but because the guidelines are only advisory, what would determine the unreasonableness of a sentence?

---

[245] United States v. Booker, 543 U.S. 220, 253–54 (2005).

[246] For the role of appellate review as a harmonizing mechanism, see generally, Jose A. Cabranes, *Reforming the Federal Sentencing Guidelines: Appellate Review of Discretionary Sentencing Decisions*, 1 HARV. LATINO L. REV. 177 (1994); Kevin R. Reitz, *Sentencing Guideline Systems and Sentence Appeals: A Comparison of Federal and State Experiences*, 91 NW. U. L. REV. 1441 (1997). This was also the view of early reformers. *See* MODEL SENTENCING AND CORRECTIONS ACT § 3-208 cmt. (NAT'L CONFERENCE OF COMM'RS ON UNIF. STATE LAWS 1979) ("Appellate review of sentencing is a fundamental aspect of the act's presumptive sentencing system. . . . Appellate review of each sentencing court's decisions in these matters will insure that departures from the guidelines are consistent with the standards established in the act and also that new situations will be uniformly handled throughout a state. Appellate review will facilitate the development of a 'common law of sentencing' to buttress and supplement the guidelines of the commission.").

[247] 42 PA. CONS. STAT. § 9781(c)(2) (2016).

[248] *Id.* § 9781(c)(3).

Instead of evaluating the reasonableness of a sentence based on cases cherry-picked by the parties or the nonrepresentative sample of cases that reach the appellate courts, statistical curving provides that information in a much more extensive format. Statistical curving can assist courts in evaluating not just the existence of certain sentences for certain offenses but also their prevalence and frequency of use. The distribution arising from statistical curving, based on a large number of actual sentences, might suggest, for example, a closer appellate review for sentences that are in the top 5% of the sentencing distribution—in this example, sentences higher than thirty-six months. Or the criterion might be sentences that are further than two standard deviations from the average—in this example, with an average of 14.5 months and a standard deviation of 11.7 months, sentences higher than 37.9 months. Appellate courts could still be sensitive to local norms and other factors in deciding whether or not to vacate a sentence in a particular case, but the curve would provide a useful starting point for the evaluation of reasonableness.

As judges change their sentencing practices, what is reasonable and unreasonable would change accordingly, facilitating both uniformity and flexibility over time. As a theoretical matter, if the sentencing range offered by the sentencing curves narrows, it will be because there is greater judicial consensus around the sentencing of an offense. Curbing judicial departures in such cases through appellate review is, therefore, less troubling, even to opponents of narrow limits on judicial discretion, because it is not subordinating judicial discretion to mandatory ranges. Rather, it is truncating judicial outliers according to the majority view of the courts.[249]

Such curbing would also not prevent a dynamic change in the sentencing curve caused by a change in the judiciary's or legislature's approach to certain crimes. A particular crime epidemic might justify enhancing sentencing levels. A societal shift that views certain crimes as less repellant could justify more mitigated sentences. When legislators change the recommended sentence for certain offenses, judges should take account of the change. In such cases, judges could depart from current sentencing practices because the goal of statistical curving is to offer guidance rather than strict rules. If such cases are appealed, appellate courts could, similarly, weigh the reasons for and desirability of changing sentencing practices. The fact that initial outliers might reach the appellate courts should be viewed as a benefit not a shortcoming. This would enable

[249] For a discussion of the desirability of appellate review and the standard of review on appeal pre- and post-*Booker*, see *infra* Section IV.B.2.

appellate courts to review the matter sooner and offer guidance to the trial courts.

B. Curving Discretion and Sentencing Policy

Statistical curving is admittedly a proposal for social design, and therefore, the call for its implementation should be supported both practically and normatively. This Section offers such support in three aspects. Section IV.B.1 explains why statistical curving suits the spirit of the guidelines enterprise and of penal justifications more generally. Section IV.B.2 demonstrates why, under existing sentencing jurisprudence, statistical curving is both an adequate and a constitutional way to address the desire for coherence and proportionality in sentencing. Section IV.B.3 explains how statistical curving also addresses some often-raised concerns about the institutional roles of legislators, judges, and prosecutors in shaping sentencing levels, as well as the political economy of sentencing.

It is worth emphasizing one important point: If the institutional design of courts and the mechanisms for case assignment affect case outcomes, not addressing these issues is its own form of social design. Embracing the status quo of varied court models and wide discretion by local administrators is itself a choice to maintain a particular political economy of franchised sentencing. As long as one recognizes two basic truths—that some forms of limited jurisdiction are inevitable and that consistency in sentencing is desirable—the only potential remedy can come through the design of sentencing recommendations for the courts and their enforcement through appellate review.

1. Empiricism and "A Common Law of Sentencing."—Existing sentencing guidelines aim to achieve a similar goal. Seeking to address the increasing sentencing disparities across judges, sentencing commissions have offered a menu of sentences to be used for different offenses, curbing judicial discretion to impose extreme sentences or consider impermissible factors.[250] Historically, sentencing guidelines achieved this function by restricting departures from the guidelines range; after *United States v. Booker*, advisory guidelines serve as an informative baseline for calculating sentences and as focal points.[251]

[250] For an overview of the desire to balance uniformity and proportionality in the sentencing guidelines, see, for example, Susan R. Klein & Jordan M. Steiker, *The Search for Equality in Criminal Sentencing*, 2002 SUP. CT. REV. 223.

[251] *See generally* John F. Pfaff, *The Continued Vitality of Structured Sentencing Following* Blakely*: The Effectiveness of Voluntary Guidelines*, 54 UCLA L. REV. 235 (2006).

But that task was not as easy to achieve, and "[t]he Guidelines have made few friends in their two decades of existence, garnering criticism from judges and scholars alike."[252] Ranking all offenses proved to be immensely difficult and even more difficult to do ex ante: "A legislature cannot know *ex ante* all the situations that will fall within the law's ambit, and in the criminal context, it tends to err on the side of sweeping too much rather than too little within a law's terms."[253] The desire to curb sentencing disparities on the one hand, while leaving enough room for judicial discretion in particular cases on the other, resulted in wide ranges for each offense—at least several months and often several years[254]—and caused a vigorous debate about the role and desirability of guidelines departures.[255]

Guidelines matrices, however, give no guidance on how to implement (and evaluate) judicial discretion, even within the range. Judges have little guidance on how to rank particular cases or how to translate ordinal rankings into cardinal rankings. If different courts develop different sentencing curves, the sentences for similar offenses will vary based on their relative rankings on each of the different sentencing curves. But even where such differences seem unwarranted to judges themselves, no mechanisms currently exist to facilitate convergence around similar sentences. Court actors are usually immersed in the local caseload. Appellate courts could potentially perform an overarching role in standardizing sentencing levels, but the deferential standard of appellate review makes vacated sentences a rare phenomenon. Meanwhile, the increased use of computer programs by judges for calculating guidelines ranges offers a potential avenue for constructing sentencing recommendations that can help overcome some of the informational hurdles.

---

[252] Rachel E. Barkow, *The Devil You Know: Federal Sentencing After* Blakely, 16 FED. SENT'G REP. 312, 312 (2004); *see also, e.g.*, KATE STITH & JOSE A. CABRANES, FEAR OF JUDGING: SENTENCING GUIDELINES IN THE FEDERAL COURTS 5–6 (1998); Bowman, *Dead Law Walking*, *supra* note 17, at 1230; Freed, *supra* note 18, at 1685–87; Marc Miller, *Rehabilitating the Federal Sentencing Guidelines*, 78 JUDICATURE 180, 180–81 (1994).

[253] Barkow, *supra* note 252, at 314.

[254] Under the federal Guidelines, the minimum interval for a range is 6 months and the widest interval is 81 months (not including ranges that extend from 360 months to a life sentence). U.S. SENTENCING GUIDELINES MANUAL § 5A (U.S. SENTENCING COMM'N 2016). In Pennsylvania, sentencing intervals for the standard sentencing range can be as narrow as 1 month or as wide as 18 months, and up to 42 months including the mitigated and aggravated ranges. For ranges that include a life sentence, the widest range extends from 72 months to life. 204 PA. CODE § 303.16(a) (2015).

[255] *See generally*, Douglas A. Berman, *Balanced and Purposeful Departures: Fixing a Jurisprudence that Undermines the Federal Sentencing Guidelines*, 76 NOTRE DAME L. REV. 21 (2000).

Conceptually, ex post construction of sentencing recommendations is aligned with common law practices for the development of law. Although, doctrinally, a sentence in one case does not pose a precedent for future cases, it nevertheless carries informational value that should not be disregarded completely. And while the rationale for not using *individual* sentences as precedent is that they constitute particular case outcomes, not general rules, valuable knowledge can be gained from looking at the patterns arising from sentencing decisions in the *aggregate.*

Early reformers of the Sentencing Commission saw importance in learning from the exercise of judicial discretion in particular cases and in utilizing appellate review to "facilitate the development of a 'common law of sentencing' to buttress and supplement the guidelines."[256] The federal Sentencing Guidelines were developed based on an extensive statistical analysis of sentencing patterns and the variables affecting them.[257] The "characteristic institutional role" of the Sentencing Commission is to "take account of 'empirical data and national experience,'"[258] and that is the fundamental reason why sentencing guidelines have continued to serve as a baseline in sentencing even under an advisory regime. When sentencing guidelines are not based on empirical data but rather on political pressure from legislatures (as in the case of the federal crack cocaine Guidelines), judges are granted greater leeway to depart based on policy disagreement, and "a categorical disagreement with and variance from the Guidelines is not suspect."[259]

Statistical curving, therefore, is conceptually what sentencing guidelines are meant to be based on. Historically, before the heavy reliance on computerization, we used the closest approximation possible—sentencing commissions were charged with collecting massive amounts of data on sentencing and translating it into sentencing ranges based on different offense and offender characteristics. Today, statistical curving can provide judges with an additional—and a more effective—tool to support decisionmaking in particular cases.

The sentencing curves offered by statistical curving could fulfill the desire of commissioners, judges, and scholars for a mechanism to enhance

---

[256] MODEL SENTENCING AND CORRECTIONS ACT § 3-208 cmt. (NAT'L CONFERENCE OF COMM'RS ON UNIF. STATE LAWS 1979).

[257] U.S. SENTENCING COMM'N, SUPPLEMENTARY REPORT ON THE INITIAL SENTENCING GUIDELINES AND POLICY STATEMENTS (1987); *see also* Stephen Breyer, *Federal Sentencing Guidelines Revisited*, 11 FED. SENT'G REP. 180, 181 (1999) (explaining how the Commission based Guidelines punishments on "typical past practice").

[258] Kimbrough v. United States, 552 U.S. 85, 109 (2007) (quoting United States v. Pruitt, 502 F.3d. 1154, 1171 (10th Cir. 2007)).

[259] Spears v. United States, 555 U.S. 261, 264 (2009) (per curiam).

uniformity across the sentences actually ordered. The continually updating nature of such curving, based on dynamic sentencing trends, would also allow for flexibility and updating of the sentencing curve based on judicial practices. And connecting the information on such trends in easily displayed curves across judges and courts can guard against increased disparities.

2. Uniformity of Sentencing in the Post-Booker *Era.*—Curving discretion does not just fit logically within the guidelines enterprise in an improved and enhanced way. In the post-*Booker* era, it will also be more effective than the guidelines matrices in addressing sentencing disparities without raising the same constitutional concerns.

The Supreme Court decision in *United States v. Booker* held unconstitutional the mandatory nature of the Sentencing Guidelines.[260] The Court also explicitly stated its position that advisory guidelines or wide judicial discretion do not pose the same difficulty. As the Court emphasized:

> If the Guidelines as currently written could be read as merely advisory provisions that recommended, rather than required, the selection of particular sentences in response to differing sets of facts, their use would not implicate the Sixth Amendment. We have never doubted the authority of a judge to exercise broad discretion in imposing a sentence within a statutory range . . . . For when a trial judge exercises his discretion to select a specific sentence within a defined range, the defendant has no right to a jury determination of the facts that the judge deems relevant.[261]

The Court also acknowledged the importance of uniformity in its decision not to strike down the Guidelines completely but rather to "make the Guidelines system advisory while maintaining a strong connection between the sentence imposed and the offender's real conduct—a connection important to the increased uniformity of sentencing that Congress intended its Guidelines system to achieve."[262] Advisory guidelines can assist in enhancing uniformity to the extent judges follow the guidelines' recommendations. But in later cases, the Court emphasized that "[the district judge] may not presume that the Guidelines range is reasonable . . . . He must make an individualized assessment based on the

---

[260] 543 U.S. 220 (2005).

[261] *Id.* at 233; *see also* Rita v. United States, 551 U.S. 338, 352 (2007) ("This Court's Sixth Amendment cases do not automatically forbid a sentencing court to take account of factual matters not determined by a jury and to increase the sentence in consequence . . . . The Sixth Amendment question, the Court has said, is whether the law *forbids* a judge to increase a defendant's sentence *unless* the judge finds facts that the jury did not find (and the offender did not concede).").

[262] *Booker*, 543 U.S. at 246.

facts presented."[263] The question is: How will uniformity be achieved where judges do not follow advisory guidelines?

Appellate judges could perform the role of overseeing sentencing decisions. But once guidelines are only advisory, departure from the guidelines ranges no longer merits de novo review on appeal.[264] Instead, the *Booker* decision directed courts of appeals to "review sentencing decisions for unreasonableness."[265] Consequently, "courts of appeals must review all sentences—whether inside, just outside, or significantly outside the Guidelines range—under a deferential abuse-of-discretion standard"[266]—which makes vacatur of sentencing decisions rare. Indeed, the implications for appellate review were emphasized by the dissent in *Booker*, and the majority conceded, "We cannot and do not claim that use of a 'reasonableness' standard will provide the uniformity that Congress originally sought to secure."[267]

After *Booker*, therefore, if one cares about achieving uniformity in sentencing, the goal should be to find a mechanism that facilitates appellate review of departures that is constitutional. Sentencing curves can be such a mechanism, and they pass the *Booker* test: They do not impose rigid, legislature-created ranges that are tied to factual findings; rather, they provide information to judges on how to apply their discretion based on the similar sentences ordered in similar cases. Appellate courts could vacate extreme sentences under such a model because heightened scrutiny on appeal is not based on the mandatory nature of the guidelines; it is based on the distance of the sentence in the particular case from the majority view of the courts.

3. The Political Economy of Sentencing.—Numerous discussions have focused around the political economy of sentencing, and those are largely outside the scope of this Article. What this Section emphasizes, however, is several key points as to how the implementation of statistical curving can fit within that discourse.

First, regarding sentencing *levels*. If one subscribes to the view that legislatures are the ones that should determine sentencing levels, statistical curving does not undermine such a view. Legislated sentencing limits could be achieved through mandatory minimum and statutory maximum sentences, as others before me have suggested in the wake of *Blakely v.*

---

[263] Gall v. United States, 552 U.S. 38, 49–50 (2007); *see also Rita*, 551 U.S. at 350–51.

[264] *Booker*, 543 U.S. at 259.

[265] *Id.* at 264.

[266] *Gall*, 552 U.S. at 41.

[267] *Booker,* 543 U.S. at 263.

Washington.[268] In such a world where binding lower and upper limits on sentencing are statutorily prescribed, the proposed statistical curving would be bounded by such limits on both ends, but it would still provide judges a useful tool for applying sentences *within* the range.[269] The addition of statistical curving on top of any statutory limits would help standardize sentencing across different judges and courts over one joint curve.

Current penal–constitutional jurisprudence, however, has taken a different direction, implying judges should have broad discretion over sentencing.[270] The Supreme Court repeatedly emphasized that wide sentencing discretion by judges does not violate the Sixth Amendment but that narrower mandatory sentencing ranges do.[271] Justice Antonin Scalia's dissent in *Booker* highlights that: "Since the Guidelines are not binding . . . the sentencing judge need only state that 'this court does not believe that the punishment set forth in the Guidelines is appropriate for this sort of offense.'"[272] Using the ranges prescribed by the Guidelines as a baseline in sentencing has been upheld because the Commission based the ranges on the actual sentences imposed in cases.[273] And where sentencing ranges are not the product of an empirical investigation but rather of political influence by Congress, the Supreme Court's decisions in *Kimbrough v. United States*[274] and *Spears v. United States*[275] allow judges to depart based on policy disagreement with the Guidelines. In a world that subscribes to the view that judges should be the ones determining sentencing levels but also wishes to maintain broad uniformity in sentencing, statistical curving that is based on the aggregation of actual judicial decisions can provide a mechanism to increase uniformity while placing the responsibility over sentencing levels with the judiciary.

---

[268] 542 U.S. 296 (2004); *see, e.g.*, Barkow, *supra* note 252, at 312–13; Stephanos Bibas, Blakely*'s Federal Aftermath*, 16 FED. SENT'G REP. 333, 338–39 (2004).

[269] It is similarly irrelevant whether statutory ranges are based on a matrix of offense gravity and the offender's prior record (as in the current guidelines regime) or on offense categories alone (as in guidelines-less states). Whatever the parameters that sentencing ranges are based on, statistical curving can guide decisions within such ranges.

[270] *See, e.g.*, Richman, *supra* note 154, at 1376 ("In the Supreme Court, the trio of *Rita v. United States*, *Gall v. United States*, and *Kimbrough v. United States* enshrined the reasonable district court as the ineffable place where federal criminal policy, sentencing philosophy, and individualized judgment merge." (footnotes omitted)).

[271] Rita v. United States, 551 U.S. 338, 352 (2007); *Booker*, 543 U.S. at 233; *Blakely*, 542 U.S. at 308–09.

[272] *Booker*, 543 U.S. at 305 (Scalia, J., dissenting in part).

[273] *Rita*, 551 U.S. at 350–51; Gall v. United States, 552 U.S. 38, 46 (2007).

[274] 552 U.S. 85 (2007).

[275] 555 U.S. 261, 264–66 (2009) (per curiam).

Statistical curving can also enrich the current discourse regarding the political economy of sentencing guidelines by highlighting three key points. First, statistical curving does not raise the same separation of powers concerns that the promulgation of fact-specific sentencing ranges by sentencing commissions raises and can help to distinguish between practical necessity and theoretical desirability of that delegation.[276] Constitutionally, statistical curving does not remove from the judicial branch the power to set sentencing ranges in categories any narrower than those offered by the penal code, as it is under a guidelines regime.[277] That is because the range offered by statistical curving is determined by the actual sentences imposed by judges, not by the prescribed range set by legislatures. Following, they also do not raise the problem of entrusting the Sentencing Commission with "legislative power outright and not as an incident to some other executive or judicial function."[278] Today, where statistical curving based on judicial decisions is practically feasible, this can dissipate the rationale for the historical constitutional bargain that placed such power in the hands of the Commission.

Once uniformity can be achieved organically, the discourse on involvement in sentencing by legislatures and sentencing commissions should be transparent about its focus on questions of harshness and leniency—questions of *magnitude*, not of range. To the extent desired, those could be achieved through mandatory minimum and statutory maximum sentences, as mentioned above. But that discourse, as well, would benefit from directly addressing the desire for punitiveness or leniency unbundled from the captivating rhetoric about fighting disparities.

Second, forming guidelines recommendations in such a dynamic manner could have an additional benefit. It is not so much about the identity of the body setting the ranges; it is about the point in time when the ranges are set. The specific time when any decision is made—including decisions about sentencing matrices—is a time that creates opportunities for focused political influences that then continue to be entrenched through the matrices as they are adopted and used. Dynamic curves can avoid such

---

[276] In *Mistretta v. United States*, 488 U.S. 361 (1989), the Court rejected the challenge to the constitutionality of the Commission and its Guidelines and concluded that Congress had not violated the separation of powers principle or the nondelegation doctrine (over the dissenting opinion of Justice Scalia). Yet, scholars have continued to raise such concerns in the following years. *See*, *e.g.*, Rachel E. Barkow, *Separation of Powers and the Criminal Law*, 58 STAN. L. REV. 989 (2006); Bibas, *supra* note 268, at 336, 340. Since 1989, the Court has also changed its guidelines jurisprudence on some issues, including *Booker*'s removal of the mandatory force of the sentencing guidelines. 543 U.S. at 233 (majority opinion).

[277] *See* Barkow, *supra* note 252; Barkow, *supra* note 276, at 1028, 1041–42.

[278] *Barkow, supra* note 276, at 1041–42.

influence points because they are updated simply by sentencing practices themselves.

In criminal law in particular, the outcomes of such political decision points tend not to reflect a legislative compromise or current social views but rather tend to converge on more extreme options.[279] One reason for this tendency is the cumulative nature of criminal and sentencing laws, where each political decision point often is affected by the social battles and high-profile, heinous cases of the era.[280] Another is the one-sidedness of interest groups' pressure on such issues.[281] From the procedural perspective, a main problem is that criminal law and sentencing law do not define rules in whose shadow parties bargain; they define a menu of options for prosecutors to choose from. High penalties may be desirable to legislators even if they exceed their real sentencing preferences and even when they do not shift actual sentencing practices. Because the public understands that prosecutors have discretion, harsher penalties mainly move the locus of public criticism from the policy choice by the legislature to the implementation of that policy by the prosecutor in the particular case.[282] In a reality where the overwhelming majority of cases are resolved by plea bargains, basing sentencing recommendations on actual sentences will increase the likelihood that sentences reflect social views.

This brings me to my last point. Regardless of one's position on the separation of powers between legislators and judges, in practice many believe the substantial sentencing power lies neither with judges nor legislators but rather with the prosecutor's office. Much criticism has been heard that current guidelines do not necessarily minimize real disparities but rather shift discretion from judges to prosecutors and shift the decision point from the in-court adjudication to the plea bargain.[283] This is

---

[279] *Id.* at 1029–30; William J. Stuntz, *The Pathological Politics of Criminal Law*, 100 MICH. L. REV. 505, 549 (2001) [hereinafter Stuntz, *Pathological Politics of Criminal Law*]; William J. Stuntz, *Reply: Criminal Law's Pathology*, 101 MICH. L. REV. 828 (2002) [hereinafter Stuntz, *Criminal Law's Pathology*].

[280] William J. Stuntz, *Plea Bargaining and Criminal Law's Disappearing Shadow*, 117 HARV. L. REV. 2548, 2558 (2004).

[281] Stuntz, *Criminal Law's Pathology*, *supra* note 279, at 836.

[282] Stuntz, *supra* note 280, at 2557 ("[W]ho bears the blame if something goes wrong—if sympathetic defendants are punished or if unsympathetic defendants are punished more harshly than the public thinks just? In the United States, the answer is almost always the overzealous prosecutor, not the overcriminalizing or oversentencing legislator."); *see also* Stuntz, *Pathological Politics of Criminal Law, supra* note 279, at 547–57; Stuntz, *Criminal Law's Pathology*, *supra* note 279, at 836.

[283] *See*, *e.g.*, Alschuler, *supra* note 18, at 926–28 (explaining that "[t]he sentencing reform movement has not restricted sentencing discretion so much as it has transferred discretion from judges to prosecutors"); Alschuler, *supra* note 154, at 102 (reviewing the empirical findings and concluding that "[t]he best judgment . . . is that some reduction in judge-produced disparity was more than offset by

concerning not only because of the difference in the public roles of judges and prosecutors but also because, to the extent disparities are masked by charging and factual agreements, they are harder to detect and overcome.[284] Further, the more restricted judges are relative to prosecutors, the greater the pressure on a defendant to reach a plea bargain and the less reliable are such pleas of guilt and the outcomes of such proceedings.[285] Scholars, practitioners, and judges have warned of the following implications that are not limited to sentencing—there is also a drift away from the adversarial model, reduced transparency, and reduced possibility to challenge the government conduct or to challenge the government's version of the events.[286] Enhancing the ability and legitimacy of judges to sentence based

---

an increase in prosecutor-produced disparity"); Rachel E. Barkow, *Institutional Design and the Policing of Prosecutors: Lessons from Administrative Law*, 61 STAN. L. REV. 869, 877 (2009) ("The United States Sentencing Guidelines . . . have also increased prosecutorial leverage by curbing judicial sentencing discretion. They have prompted more pleas and fewer trials.") [hereinafter Barkow, *Institutional Design and the Policing of Prosecutors*]; Rachel E. Barkow, *Sentencing Guidelines at the Crossroads of Politics and Expertise*, 160 U. PA L. REV. 1599, 1622–25 (2012) ("As long as guidelines apply only to judges, they will never resolve the disparity in the system and in fact may end up exacerbating the disparity by failing to provide a valuable check on prosecutors."); Bowman, *Failure of the Federal Sentencing Guidelines*, *supra* note 17, at 1336–39 (discussing the "increasing array of tools for controlling sentencing outcomes" given to prosecutors); Freed, *supra* note 18, 1697–98 (describing the Sentencing Guidelines as "administrative handcuffs that are applied to judges" but not to prosecutors); Richman, *supra* note 154, at 1386 ("With substantial control over the flow of offense-related facts to the judge . . . prosecutors were left with unprecedented sway over sentencing."); Daniel C. Richman et al., *Panel Discussion: The Expanding Prosecutorial Role from Trial Counsel to Investigator and Administrator*, 26 FORDHAM URB. L.J. 679, 682–87 (1999) (Judge John S. Martin's critique of the "secret nature" of the Guidelines, and "the totally non-public, non-reviewable discretion that is in fact given to some people who are as old as twenty-nine"); Stephen J. Schulhofer & Ilene H. Nagel, *Plea Negotiations Under the Federal Sentencing Guidelines: Guideline Circumvention and Its Dynamics in the Post*-Mistretta *Period*, 91 NW. U. L. REV. 1284 (1997) (concluding that prosecutorial sentencing discretion "if unchecked, has the potential to recreate the very disparities that the Sentencing Reform Act was intended to alleviate"); Jeffrey Standen, *Plea Bargaining in the Shadow of the Guidelines*, 81 CALIF. L. REV. 1471, 1475 (1993) (noting that the Guidelines have "curtailed judges' ability to constrain prosecutors"); STITH & CABRANES, *supra* note 252, at 130 (explaining that "prosecutorial discretion is now greater *relative* to judicial discretion in criminal sentencing"); Kate Stith, *The Arc of the Pendulum: Judges, Prosecutors, and the Exercise of Discretion*, 117 YALE L.J. 1420, 1425 (2008) (noting that the Guidelines "provided prosecutors with indecent power relative to both defendants and judges").

[284] Barkow, *supra* note 276, at 1027–28; Richman et al., *supra* note 283, at 687.

[285] For some vivid examples, see David E. Patton, *Federal Public Defense in an Age of Inquisition*, 122 YALE L.J. 2578 (2013); Richman et al., *supra* note 283, at 686 (comments of Judge John S. Martin). For a theoretical analysis explaining why, when statutory punishments exceed prosecutorial sentencing preferences, the settlement curve reflects the same punishments even for cases of lower evidentiary strength, see Stuntz, *supra* note 280, at 2552–53, and Richman et al., *supra* note 283, at 689–90 (comments of Judge Gerard E. Lynch).

[286] Barkow, *supra* note 276, at 1034, 1047–48; Barkow, *Institutional Design and the Policing of Prosecutors*, *supra* note 283, at 878–79; Gerard E. Lynch, *Our Administrative System of Criminal Justice*, 66 FORDHAM L. REV. 2117 (1988); Patton, *supra* note 285, at 2598; Richman et al., *supra* note 283, at 686; Stuntz, *Criminal Law's Pathology*, *supra* note 279, at 829, 832, 838.

on the sentencing curves—themselves based on de facto "going rates"—will give judges the same power of discretion already held by prosecutors.

CONCLUSION

Defining courts' scope of jurisdiction and assigning judges to dockets are commonly viewed as administrative decisions. Further, viewing judicial professionalism as a virtue has cultivated a movement towards increased specialization in the courts—resulting in varying types of case concentrations often decided on at the locality level. This Article identifies an additional effect of case assignments: they affect the substance of judicial decisions as well.

Because legal judgments are relative, the decision in a particular case depends on the contours of the other cases in judicial caseloads. Relative judgments can inject concerning arbitrariness into judicial decisions in particular cases based on caseload compositions. Worse, relative judgments can systematically bias sentencing practices across courts of different subject matter or geographic jurisdiction. This Article demonstrates how such relative judgments are evident across many courts of limited jurisdiction—from courts under two-tiered models focusing on only misdemeanors or felonies to specialized courts such as juvenile courts and domestic violence divisions. It also illustrates how relative judgments are associated with the development of geographic disparities in sentencing.

The normative implications are extensive. If the types of cases under the jurisdiction of the courts affect the substance of the decisions reached, this means that decisions about the division of jurisdiction are not merely administrative. They entail a decision about sentencing levels themselves by nudging the sentencing curve in a particular direction. In light of the variance in court models, the increased movement toward specialization in the courts and the great discretion given to local court administrators in case assignments, such consequences are especially concerning. From the policymaker's perspective, as specialization becomes increasingly prevalent in the legal system, connecting divergence in outcomes to institutional capacity—and learning the reasons for any such influences—can serve as important tools in implementing future policies. It can also shed new light on the role of other institutions in the chain of sentencing, including parole boards and courts of appeals.

The effect of relative judgments can also help explain why guidelines have been less successful in certain areas than others. In an important sense, the increasing movement towards specialization in the courts inherently undermines the quest for uniformity in sentencing. And it undermines it in a particularly disturbing way because the sentencing

curves that develop are not based on any penologically relevant considerations. Such biased sentencing clearly undermines notions of justice and equitable treatment because similarly situated defendants are sentenced differently. At the same time, such sentencing disparities cannot be justified as the price we pay for judicial discretion. It is not that sentencing differentials reflect differences in judicial ideology because even the same judge might impose different sentences based on her exposure to different cases. This is also not about emphasizing individualization in sentencing because the decision in a particular case is affected by the characteristics of the other cases in the group of cases before the judge. Furthermore, because courts of limited jurisdiction often concentrate on relatively less serious offenses, specialization will also often be associated with increased punitiveness.

In light of the wide effect of relative judgments in the criminal justice system, this Article also suggests complementing existing sentencing guidelines with statistical curving. Exposing judges to the overall punishment curve, across all courts and jurisdictions, would enable better standardization of particular sentences in a way that would promote consistency across courts and judges. It would also enable appellate courts to perform an overarching role in reviewing and standardizing sentencing levels.