4-2018

Booker Disparity and Data-Driven Sentencing

Joshua M. Divine

Booker Disparity and Data-Driven Sentencing

JOSHUA M. DIVINE*

Sentencing disparity among similar offenders has increased at a disconcerting rate over the last decade. Some judges issue sentences twice as harsh as other judges on the same court, so a defendant's sentence often depends substantially on which judge is randomly assigned to the defendant's case. The old mandatory sentencing guidelines repressed disparity but only by causing unwarranted uniformity. The advisory guidelines swing the pendulum toward the opposite extreme, and this problem promises to grow worse as the lingering effects of the old regime diminish.

This Article proposes a system—data-driven appellate review—to curb sentencing disparity without re-introducing unwarranted uniformity. In a system where sentencing judges possess significant discretion, only meaningful appellate review can restrain the tendency for sentencing practices to deviate substantially between judges. Discretionary sentencing decisions are currently reviewed for abuse of discretion, which leads some judges to issue much harsher or much more lenient sentences than other judges. Congress can mitigate this problem of inter-judge disparity by changing the standard of review, creating a rebuttable presumption that outlier sentences among similar offenders are unreasonable. The data the U.S. Sentencing Commission collects on over 70,000 criminal cases annually can be used to define categories of similar offenders. Culling outlier sentences through data-driven appellate review would increase judicial awareness of sentences issued by peer judges, restricting inter-judge disparity without incurring unwarranted uniformity.

TABLE OF CONTENTS

INTRODUCTION

Federal sentencing policy leaves a substantial portion of each sentence subject to the identity of the sentencing judge. A card randomly drawn from a shuffled deck—a method some courts have used to assign judges to cases[1]—can alter the length of a sentence by several years. Between 2007 and 2011, one judge in the District of Nebraska sentenced drug offenders to a median 60 months while a colleague on the same court sentenced drug offenders to a median twice that amount.[2] This inter-judge disparity has swelled since the Supreme Court decision in *United States v. Booker*[3] rendered the sentencing guidelines advisory, and the disparity shows signs of only increasing.

Booker nullified an important part of the Sentencing Reform Act, which had created the Sentencing Commission and clothed the Commission with authority to create the mandatory Sentencing Guidelines. Reducing inter-judge disparity was a central aim—if not the central aim[4]—of the Act.[5] Congress sought to achieve this aim by pursuing an ideal of uniformity that required judges to sentence defendants within complicated and narrow guideline ranges that judges were to calculate based on facts they found.[6] The Act had no shortage of faults,[7] and it reduced inter-judge disparity at the cost of increasing unwarranted uniformity,[8] but it successfully reduced inter-judge disparity in measurable ways through the mandatory guidelines.[9]

---

1. *E.g.*, D. MINN. ORDER FOR ASSIGNMENT OF CASES 7 (Dec. 1, 2008).

2. Richard G. Kopf, *Judge-Specific Sentencing Data for the District of Nebraska*, 25 FED. SENT'G REP. 50, 51 tbl. 7 (2012). One of these judges was a senior judge and may not have had a random sampling of cases, but the average difference between active judges was still twenty-two percent in drug cases. *Id.*

3. 543 U.S. 220 (2005).

4. KATE STITH & JOSE A. CABRANES, FEAR OF JUDGING: SENTENCING GUIDELINES IN THE FEDERAL COURTS 104 (1998); Ryan W. Scott, Booker*'s Ironies*, 47 U. TOLEDO L. REV. 695, 714 (2016) ("Among Congress's primary objectives . . . was the reduction of inter-judge disparity . . . ."). Certainly "[e]liminating unwarranted sentencing disparity was the primary goal of the Sentencing Reform Act," even if eliminating inter-judge disparity, in particular, was not. U.S. SENTENCING COMM'N, FIFTEEN YEARS OF GUIDELINES SENTENCING 79 (2004).

5. Ryan W. Scott, *Inter-Judge Sentencing Disparity After* Booker: *A First Look*, 63 STAN. L. REV. 1, 3 (2010) (noting that the Sentencing Reform Act was intended to drive away sentencing disparity created by "the philosophy, politics, or biases of the sentencing judge").

6. 18 U.S.C. § 3553(b) (2012).

7. *See, e.g.*, STITH & CABRANES, *supra* note 4, at 51–60, 68, 82, 94, 98, 115.

8. Paul J. Hofer, *Data, Disparity, and Sentencing Debates: Lessons from the TRAC Report on Inter-Judge Disparity*, 25 FED. SENT'G REP. 37, 40 (2012) (noting that Guidelines departures are sometimes necessary to prevent disparity).

9. FIFTEEN YEARS OF GUIDELINES SENTENCING, *supra* note 4, at 97-99 ("The federal sentencing guidelines have made significant progress toward reducing disparity caused by judicial discretion."); James M. Anderson et al., *Measuring Interjudge Sentencing Disparity: Before and After the Federal Sentencing Guidelines*, 42 J.L. & ECON. 271, 294 (1999) (finding evidence that interjudge disparity fell thirty-six percent in a six-year span following the Sentencing Reform Act). *But see* U.S. SENTENCING COMM'N, *supra* note 4, at 98 (showing that judicial disparity increased for robbery and immigration

In a post-*Booker* world where pre-defined guideline ranges are advisory only, the Act has lost the element most responsible for reducing inter-judge disparity, and nothing remotely sufficient has replaced that element. Instead, "copious evidence" now shows that *Booker* decimated the central purpose of the Act.[10] To be sure, *Booker* is responsible for only a slight decrease in aggregate sentence averages (although average drug sentences decreased substantially).[11] But *Booker* instigated significant disparity between the individual sentences that make up those aggregate averages when *Booker* augmented judicial discretion and reduced meaningful appellate review.[12] *Booker* has at least doubled how important a specific judge is to the outcome of a sentence in certain districts, and the effect is even greater after controlling for mandatory minimum sentences because those sentences create an artificial uniformity that masks variation.[13] The example from the District of Nebraska also understates the extent of inter-judge disparity because judges in that district adhere to the recommendations of the advisory guidelines at higher rates than average, so offenders in that district experience below-average disparity.[14] This disparity is driven not only from differences in judicial philosophy, but also from the basic lack of awareness many judges have of sentences their peer judges—that is, other judges on the same court—impose, which in turn makes many judges unaware of their own harsh or lenient tendencies.[15]

This problem is likely to intensify. Although many judges continue to issue sentences within calculated ranges established by the advisory guidelines, that practice is attributable in part to the inertia of judges who grew used to sentencing when those ranges were mandatory and have not deviated much from their original practices. One should not expect newly appointed judges to exhibit the same tendency.[16]

What is to be done about a system that does not achieve its central aim of reducing arbitrary disparities between offenders? No shortage of

---

offenses); STITH & CABRANES, *supra* note 4, at 106 (arguing that inter-judge disparity was exaggerated before the Sentencing Reform Act).

10. Frank O. Bowman, III, *Dead Law Walking: The Surprising Tenacity of the Federal Sentencing Guidelines*, 51 HOUS. L. REV. 1227, 1266 (2014).

11. Jeffery Ulmer & Michael T. Light, *Beyond Disparity: Changes in Federal Sentencing After Booker and* Gall, 23 FED. SENT'G REP. 333, 335 (2011); Frank O. Bowman, III, *Prolegomenon on the Status of the Hopey, Changey Thing in American Criminal Justice*, 23 FED. SENT'G REP. 93, 94 (2010) (reporting that within-range sentences dropped from 70.9 % before *Booker* to 54.8 % in fiscal year 2010).

12. Bowman, III, *supra* note 10, at 1261, 1266–67.

13. Scott, *supra* note 5, at 5, 33 (reporting, among other things, that several judges in the post-*Booker* period depart three times as often as other judges).

14. Kopf, *supra* note 2, at 51.

15. Kopf, *supra* note 2, at 51.

16. *See* Nancy Gertner, *Supporting Advisory Guidelines*, 3 HARV. L. & POL'Y REV. 261, 270 (2009) (identifying adherence to the Guidelines post-*Booker* as the result of "habits ingrained during twenty years of mandatory Guideline sentencing").

scholars have suggested abolishing and replacing the Guidelines.[17] Yet the majority of judges favor the Guidelines in their advisory state.[18] Judges tend to appreciate the Guidelines and find them helpful;[19] presumably, they disfavor revolutionary departures from the current system. Moreover, the existing proposals seek other goals, not a reduction in inter-judge disparity. This Article contributes to the literature by providing the first reform tailored specifically toward attacking inter-judge disparity in the post-*Booker* world. It does so by supplementing the current system instead of replacing it and by delving deeply into the sentencing data the Commission collects to demonstrate how that data can be used to craft a remedy for inter-judge disparity.

In particular, this Article establishes that direct methods of attacking inter-judge disparity invariably produce unwarranted uniformity, which creates its own inequities. This was the approach taken by the mandatory guidelines. A more indirect approach that focuses on discrete, specific instances of inter-judge disparity should be tried instead.[20] Federal appeals courts currently review sentences under a highly deferential, abuse-of-discretion standard. Congress should alter this standard by creating a presumption for federal appellate courts that a sentence is unreasonable if it deviates too much from what similarly situated offenders have received. Creating that presumption would (1) induce sentencing judges to provide greater justifications for their sentences in an attempt to overcome a presumption of unreasonableness and (2) make sentencing judges more aware of the sentences imposed by other judges on their courts, an awareness that is currently lacking.[21] These two predominant effects would ameliorate system-wide inter-judge disparity.

Because this reform depends on measuring the distance a sentence deviates from those sentences typically issued to similarly situated offenders, this reform can work only if a feasible method exists for defining categories of similarly situated offenders. Fortunately, those categories can readily be defined by using the voluminous sentencing

---

17. *E.g.*, William K. Sessions III, *At the Crossroads of the Three Branches: The U.S. Sentencing Commission's Attempts to Achieve Sentencing Reform in the Midst of Inter-Branch Power Struggles*, 26 J.L. & POL. 305, 309 (2011) (arguing that Congress should reformulate and simplify the sentencing guidelines).

18. U.S. SENTENCING COMM'N, RESULTS OF SURVEY OF UNITED STATES DISTRICT JUDGES JANUARY 2010 THROUGH MARCH 2010 tbl.19 (2010), http://www.ussc.gov/research-and-publications/ research-projects-and-surveys (reporting that more than three-quarters of judges believe the advisory system "'best achieves' the sentencing goals outlined in the U.S. Code, 18 U.S.C. § 3553, and only three percent of judges believe the mandatory Guidelines were better.").

19. Alexander Bunin, *Reducing Sentencing Disparity by Increasing Judicial Discretion*, 22 FED. SENT'G REP. 81, 82 (2009).

20. For reasons explained below, this reform is best initiated through Congress, not courts.

21. Kopf, *supra* note 2, at 50.

data the U.S. Sentencing Commission already collects and publishes.[22] This data, combined with modern technology, permits this proposal to move forward in a way unfathomable when Congress passed the Sentencing Reform Act in 1984. Through computers and sentencing data, courts can readily construct categories of similarly situated offenders and calculate the ranges of sentences issued in each of those categories.

Other articles have correctly identified the main problem in the post-*Booker* world—a lack of meaningful appellate review—but none has constructed the contours of an appellate solution or established why an appellate solution is feasible. This proposal fills that gap, and this Article also defends the constitutionality of increasing appellate review. *Booker* struck down stronger appellate review when *Booker* rendered the Guidelines advisory,[23] and several scholars believe that increasing appellate review might unconstitutionally violate the remedial holding of *Booker*.[24] This Article explains why changing appellate review is feasible and constitutional. Finally, the literature has not yet addressed inter-judge disparity that occurs *within* guideline ranges. Within-range sentences occur overwhelmingly at the bottom of each range—96.6% of judges sentenced at the bottom of the range in the pre-*Booker* era—meaning that a judge who deviates from the norm operates as a significant source of inter-judge disparity.[25]

---

22. 28 U.S.C. § 994(w) (2012).

23. United States v. Booker, 543 U.S. 220, 245–46 (2005).

24. *E.g.*, Amy Baron-Evans & Kate Stith, Booker *Rules*, 160 U. PA. L. REV. 1631, 1721–29, 1731–36 (2012) (arguing that there are constitutional issues with Congress legislatively creating a presumption of reasonableness); Carissa Byrne Hessick, *A Critical View of the Sentencing Commission's Recent Recommendations to Strengthen the Guidelines System*, 51 HOUS. L. REV. 1335, 1359 (2014) (explaining *Booker* as striking a balance between district court discretion and fealty to the Guidelines, which creates "a real possibility that the Court might determine that [appellate review] proposals, if adopted, alter the *Booker* remedy to such an extent that it no longer fixes the Sixth Amendment problem").

25. Of 911 judges who sentenced at least 10 defendants between 1999 and 2001, sentencing at the bottom of the guideline range was the standard practice for the vast majority of judges, 880 of them. U.S. SENTENCING COMM'N, *supra* note 4, at 109.
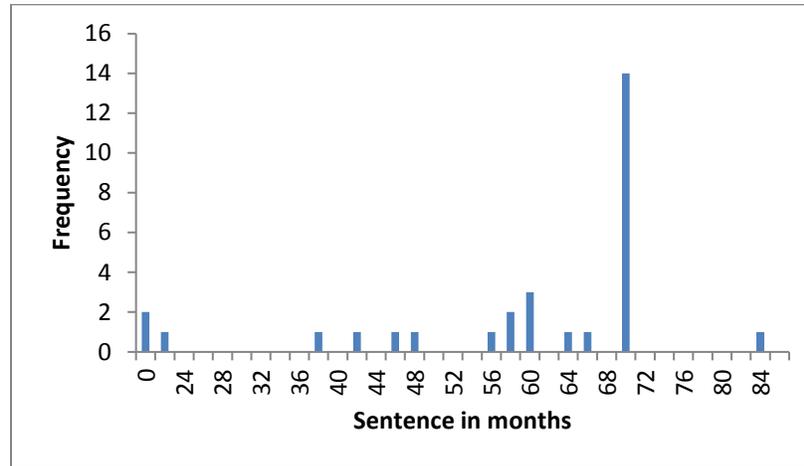
Figure 1: Sentences for Certain Drug Convictions, D. Neb.

Figure 1[26] shows the distribution of sentences issued to defendants in the District of Nebraska who bear the same major offender characteristics: statutory sentencing guideline, guideline range (seventy to eighty-seven months), criminal history, acceptance of responsibility, and whether a defendant was subject to and sentenced to a mandatory minimum.[27] This figure shows the dominant practice of sentencing offenders toward the bottom of a guideline range (seventy months in this example). The literature has failed to ask whether the offender who was sentenced to eighty-four months experienced inter-judge disparity despite being sentenced within the guideline range. This Article creates a system of appellate review that can readily address that question.

In Part I, this Article describes the genesis of the Guidelines. It briefly explains *Booker* and the Supreme Court's relevant subsequent decisions before exploring the studies that show a substantial increase in post-*Booker* inter-judge disparity. In Part II, this Article explains the proposal in more detail and suggests a statutory change that would lend effect to the proposal. In Part III, this Article discusses both the expected results of this reform on inter-judge disparity and the limits of the proposal. In doing so, it contrasts this proposal with others and shows that this reform is more capable of reducing inter-judge disparity.

---

26. To reduce the complexity of the horizontal axis, this Article classifies the data into fewer variables. The bin labeled "70," for example, represents all defendants who received sentences greater than sixty-eight months and less than or equal to seventy months. All defendants listed here received seventy-month sentences.

27. Because the Guidelines instruct judges to *disregard* mandatory minimums where an offender has substantially assisted the prosecution or is eligible for a safety valve, this Article analyzes all offenders together who were not issued a mandatory minimum.

Additionally, this Article undertakes the critical task of showing that the reform is constitutional (because the reform operates outside the structure of the Guidelines) and that other reforms may not be.

Part IV explains the methodology for how the data the Commission collects can be used (1) to create categories of similarly situated offenders and (2) to discern what sentences were issued to those offenders. This proposal can work only under a relatively simple formula for defining categories rather than the more complicated formula another author has proposed for a different but related purpose.[28] Part V establishes how this method of selecting similarly situated offenders works and provides several examples that display the distribution of sentences among different offenders.

I. CURRENT STATE OF SENTENCING

A proper understanding of current sentencing policy is incomplete without a general understanding of its historical developments.[29]

A. *BOOKER'S* HISTORY

1. Booker's *Genesis*

In early colonial history, juries ordinarily made sentencing decisions, in part because penitentiaries were not yet common. Without prisons, convictions were usually linked with one specific sentence—often death—so no rigorous sentencing mechanism was needed.[30] Once penitentiaries became more common, federal law frequently provided fixed statutory sentences[31] but also defined open ranges in which judges could sentence.[32] The pre-defined ranges were indeterminate—that is, judges had full discretion to sentence within each range—and federal law provided for virtually no appellate review, which was unusual among common law countries.[33] The lack of appellate review made developing federal common law on sentencing issues almost impossible.[34]

In the twentieth century, sentencing philosophy changed and altered with it the structure of sentencing policy. Early sentencing philosophy was largely retributive.[35] By the middle of the twentieth

---

28. *See* discussion *infra* Part III.A.3.

29. For an extended analysis of this history, see STITH & CABRANES, *supra* note 4, at 9–77; Nancy Gertner, *A Short History of American Sentencing: Too Little Law, Too Much Law, or Just Right*, 100 J. CRIM. L. & CRIMINOLOGY 691, 692–97 (2010).

30. Gertner, *supra* note 29, at 694.

31. Note, *The Admissibility of Character Evidence in Determining Sentence*, 9 U. CHI. L. REV. 715 (1942).

32. Apprendi v. New Jersey, 530 U.S. 466, 481 (2000).

33. Gertner, *supra* note 29, at 695–96.

34. STITH & CABRANES, *supra* note 4, at 23.

35. Gertner, *supra* note 29, at 694.

century, rehabilitative philosophies dominated. Necessary to those philosophies was broad discretion for judges.[36] But rehabilitation soon gave way to rationalism and scientific rigor.[37] The ideal of uniformity arrived,[38] and with it judicial discretion became less necessary.[39] A system based on rehabilitation required the careful judgment of judicial actors. In contrast, a scientifically rigorous algorithmic system was naturally more retributive and made Congress and the public the experts.[40]

One example that exhibits the newer paradigm of scientific rigor is the American Law Institute's attempt to "rationaliz[e] reforms in the area of crime definition" by instituting the Model Penal Code.[41] At the time, the Supreme Court was busy fundamentally reshaping criminal procedure but was doing so only at the adjudication stage, leaving sentencing largely untouched.[42] The Model Penal Code made sentencing appear anti-scientific and non-rigorous in comparison[43] and invited criticisms against the sentencing system, which many viewed as dependent on the whims of judges and parole authorities.[44]

During the push for the Sentencing Reform Act, scholars argued that sentence variation was partly explained by the temperaments of individual judges and sensitive defendant characteristics, such as race, education, and class.[45] This culminated into a virtual crisis when Marvin Frankel, then a judge on the Southern District of New York and the "father of sentencing reform,"[46] created a questionnaire of hypothetical cases. The questionnaire was mailed to federal judges in the Second Circuit. The results evinced a "glaring disparity" in sentencing practices on identical cases.[47] In eighty percent of the hypothetical cases, judges

---

36. United States v. Mueffelman, 327 F. Supp. 2d 79, 83 (D. Mass. 2004) (Gertner, J.) (describing historical sentencing practices and declaring the judge's role in rehabilitative sentencing as "almost like a doctor or social worker exercising clinical judgment").

37. STITH & CABRANES, *supra* note 4, at 11–13, 23.

38. Michael M. O'Hear, *The Original Intent of Uniformity in Federal Sentencing*, 74 U. CIN. L. REV. 749, 751 (2006).

39. STITH & CABRANES, *supra* note 4, at 28–30.

40. Gertner, *supra* note 29, at 691.

41. STITH & CABRANES, *supra* note 4, at 29–30.

42. Williams v. New York, 337 U.S. 241, 252 (1949) (rejecting the argument that a judge's consideration of a pre-sentence report, consisting of information provided by witnesses not available for cross-examination, violated the defendant's due process rights).

43. STITH & CABRANES, *supra* note 4, at 29–30.

44. STITH & CABRANES, *supra* note 4, at 21; *see also* MARVIN E. FRANKEL, CRIMINAL SENTENCES: LAW WITHOUT ORDER 5 (1973); Anderson, *supra* note 9, at 274 (regarding disparity as "that variation caused by the identity" of specific judges). *See generally* KENNETH CULP DAVIS, DISCRETIONARY JUSTICE: A PRELIMINARY INQUIRY (1976) (arguing that the greatest and most frequent injustice regarding sentencing occurs when judges are afforded broad discretion).

45. STITH & CABRANES, *supra* note 4, at 31.

46. STITH & CABRANES, *supra* note 4, at 35.

47. STITH & CABRANES, *supra* note 4, at 31.

disagreed on whether incarceration was even appropriate; in one hypothetical extortion case, the most lenient judge would have applied only a three-year sentence compared to the twenty-year sentence a harsher judge would have applied.[48]

The desire to rationalize sentencing and make it scientifically rigorous finally led to the Sentencing Reform Act.[49] Only a single voter in the Senate rejected the bill.[50] But by that time, the Act had morphed into a tool bent on creating utopia. Senator Ted Kennedy, one of the principle architects, had first tried to create advisory Guidelines,[51] but the Act as passed instead sought to reduce inter-judge disparity through brute force by (1) limiting judges to sentences within the calculated guideline range for each specific offender and (2) granting appellate courts broad appellate review.[52] A sentence outside the calculated guideline range was to be reversed except under a few situations where departure was authorized.[53] Congress further ramped up enforcement of the Guidelines in 2003—in apparent reaction to higher-than-acceptable departure rates—by creating *de novo* appellate review for certain circumstances.[54] Originally intended to apply to all unenumerated downward departures,[55] the 2003 statute ultimately made departing harder only in crimes involving pornography, sexual abuse, sexual activity with minors, and child kidnapping and trafficking.[56] The Act turned a process that had been intended to be more acceptable to offenders into one designed to discourage crime and encourage cooperation between defendants and prosecutors in the name of uniformity.[57] This model, centered on an ideal of uniformity, was largely rejected by the states.[58]

---

48. ANTHONY PARTRIDGE & WILLIAM B. ELDRIDGE, FED. JUD. CTR., THE SECOND CIRCUIT SENTENCING STUDY: A REPORT TO THE JUDGES OF THE SECOND CIRCUIT 6–7 (1974).

49. *See* O'Hear, *supra* note 38, at 751 ("[C]ritics argued in the 1970s that rehabilitation was an uncertain concept that might be misused as cover for irrational and inhumane practices.").

50. STITH & CABRANES, *supra* note 4, at 43.

51. STITH & CABRANES, *supra* note 4, at 41 n.21.

52. 18 U.S.C. § 3742(a) (2012).

53. 18 U.S.C. § 3742(f)(2) (2012); 18 U.S.C. § 3553(b) (permitting departure if the court finds that there exists an aggravating or mitigating circumstance "of a kind, or to a degree, not adequately taken into consideration by the Sentencing Commission in formulating the guidelines that should result in a sentence different from that described").

54. PROTECT Act, Pub. L. No. 108-21 § 401(d)(2), 117 Stat. 650, 667–68 (2003) (amending 18 U.S.C. § 3742(e) to grant appellate courts *de novo* review over "the district court's application of the guidelines to the facts").

55. H. Amend. 19 to H.R. 1104, 108th Cong. (2003) (Rep. Feeney).

56. PROTECT Act § 401(a), (b) (permitting only those departures listed in United States Sentencing Guidelines Manual § 5H, not those departures recognized solely in United States Sentencing Guidelines Manual § 5K).

57. O'Hear, *supra* note 38, at 816.

58. O'Hear, *supra* note 38, at 816.

The Act created the Sentencing Commission[59] and required it to draft a comprehensive, binding[60] sentencing code that controlled essentially all significant sentencing decisions.[61] Congress directed the Commission to create a rigorous and comprehensive system that would "reflect every important factor relevant to sentencing for each category of offense and each category of offender, give appropriate weight to each factor, and deal with various combinations of factors."[62] In addition to the comprehensive system of sentencing, the Commission included "commentaries" and "policy statements" that delineated the purpose of the Guidelines. The Court later interpreted these statements to be as authoritative as the Guidelines themselves.[63]

The desire to introduce scientific rigor into sentencing certainly found its imprint on the Act. As written, the Act requires judges to follow a lengthy process of discerning (by a preponderance of the evidence) whether any number of voluminous adjustments applies to an offender.[64] A judge first determines which Guidelines category covers the offender before applying numerous "specific offense characteristics" to calculate a "base offense level" score.[65] This process "reflects the Commission's determination to minimize the need for sentencing judges to exercise their judgment."[66] The task is frequently difficult. Categories often cross-reference each other,[67] and a judge may have to compute multiple calculations. After the base offense level is determined, the judge adjusts the level based on enhancements that exist for an offender's role in the crime, the vulnerability of the victim, etc. The judge then calculates the offender's criminal history.

---

59. 28 U.S.C. § 991(a) (2012). Initially, only one member of the Commission had ever been involved in sentencing. STITH & CABRANES, *supra* note 4, at 49.

60. Judges were prohibited from departing from the calculated guideline ranges except in rare, exceptional circumstances. 18 U.S.C. § 3553(b)(1) (2012) (excised in United States v. Booker, 543 U.S. 220, 220 (2005)); United States Sentencing Guidelines Manual § 5K1.1; *id.* at § 5K2.0.

61. 28 U.S.C. § 994(a) (2012). The Act also abolished parole. Kate Stith & Steve Y. Koh, *The Politics of Sentencing Reform: The Legislative History of the Federal Sentencing Guidelines*, 28 WAKE FOREST L. REV. 223, 226–42 (1993) (examining the abolition of parole and the many proposed, but never enacted, statutory provisions that would have kept some indeterminate parole). Parole had been in place at least since 1910. Act of June 25, 1910, ch. 387, 36 Stat. 819, 819.

62. S. REP. NO. 225, 98-225, at 166 (1983).

63. Stinson v. United States, 508 U.S. 36, 41–47 (1993) (holding commentary statements to be authoritative); Williams v. United States, 503 U.S. 193, 201 (1992) (holding policy statements to be authoritative).

64. Among many others, these include whether a gun was possessed or fired, the amount of fiscal loss, the quantity of drugs, whether a burgled building was a residence, and whether any bodily injury occurred. U.S. SENTENCING GUIDELINES MANUAL ch. 2 (U.S. SENTENCING COMM'N 2015).

65. *Id.* at ch. 2, introductory cmt.

66. STITH & CABRANES, *supra* note 4, at 68.

67. *See* U.S. SENTENCING GUIDELINE MANUAL § 1B1.5 (U.S. SENTENCING COMM'N 2015); *e.g., id.* § 2K2.1 (felon-in-possession of firearms guideline, instructing the judge to apply section 2X1.1—which covers attempts, solicitations, and conspiracies—if the possession was in connection with another offense and the resulting sentence is greater).

The judge then locates the intersection of the criminal history and offense level scores on the Commission's matrix chart. Each intersection yields a narrow range, and the top of the range generally cannot exceed the bottom by more than twenty-five percent.[68] Still not yet finished, the judge then takes into account several dozen Guidelines policy prescriptions that dictate whether the judge can depart from the calculated guideline range in narrow circumstances.[69]

2. Booker *and Its Progeny*

This rigorous and binding system survived for eighteen years until *United States v. Booker* made the Guidelines effectively advisory. Decided at the apotheosis of the efforts by the Rehnquist Court to establish itself as "one of the most vigorous protectors of the jury guarantee in Supreme Court history,"[70] the Court reasoned that a Sixth Amendment violation occurs when a maximum sentence is increased based on facts proved only by preponderance to a judge instead of beyond a reasonable doubt to a jury.[71] The Court also held later that elements that increase the *minimum* sentence a court can impose must also be proved before a jury.[72]

In fashioning a remedy, the Court could have simply required prosecutors to present those elements to a jury.[73] Instead, the Court, in an attempt to divine Congressional intent, excised two provisions from the Act.[74] The first had required *de novo* review. The Court found that the Act, in the light of the Court's constitutional holding, implied a substitute standard of "reasonableness" review.[75] The second provision had made the Guidelines mandatory. Excising that provision the Court "ma[de] the Guidelines effectively advisory,"[76] requiring judges to both consider whether a sentence was reasonable in the light of the statutory factors Congress had identified as relevant to sentencing *and* to calculate and

---

68. 28 U.S.C. § 994(b)(2) (2012); U.S. Sᴇɴᴛᴇɴᴄɪɴɢ Gᴜɪᴅᴇʟɪɴᴇs Mᴀɴᴜᴀʟ ch. 1, pt. A, 2 (U.S. Sᴇɴᴛᴇɴᴄɪɴɢ Cᴏᴍᴍ'ɴ 2015).

69. U.S. Sᴇɴᴛᴇɴᴄɪɴɢ Gᴜɪᴅᴇʟɪɴᴇs Mᴀɴᴜᴀʟ ch. 1, pt. A, 6–7 (U.S. Sᴇɴᴛᴇɴᴄɪɴɢ Cᴏᴍᴍ'ɴ 2015).

70. Rachel E. Barkow, *Originalists, Politics, and Criminal Law on the Rehnquist Court*, 74 Gᴇᴏ. Wᴀsʜ. L. Rᴇᴠ. 1043, 1044 (2006).

71. United States v. Booker, 543 U.S. 220, 232 (2005). Because the Guidelines were mandatory before *Booker*, the district court was statutorily limited to imposing a sentence within the 210-262 month calculated range based on what the jury had found (minus certain statutory exceptions not applicable). *Id.* at 227. But the district court exercised judicial fact-finding to increase the offender's sentencing range to 360 months to life. *Id.*

72. Alleyne v. United States, 133 S. Ct. 2151, 2162–63 (2013).

73. *See, e.g.*, *Booker*, 543 U.S. at 325 (Thomas, J., dissenting in part).

74. *Id.* at 227, 246 ("We seek to determine what 'Congress would have intended' in light of the Court's constitutional holding").

75. *Id.* at 261.

76. *Id.* at 245.

consider the Guidelines before sentencing a defendant.[77] Calculating the Guidelines is important because "in the ordinary case, the Commission's recommendation of a sentencing range will 'reflect a rough approximation of sentences that might achieve § 3553(a)'s objectives.'"[78]

Booker quickly became a subject of both scholarly and judicial interest. One of the most cited Supreme Court cases,[79] *Booker* also spawned a lengthy doctrine.[80] Most importantly, the Court expounded on its promulgation of reasonableness review shortly after *Booker*.[81] The Court highlighted the close relationship between the Guidelines and the § 3553(a) statutory sentencing factors. Section 3553(a) outlines a list of

---

77. *Id.* (invoking 18 U.S.C. § 3553(a)).

78. Kimbrough v. United States, 552 U.S. 85, 109 (2007) (quoting Rita v. United States, 551 U.S. 338, 350 (2007)).

79. Courts have cited *Booker* more than 30,000 times compared to the just over 15,000 citations to *Chevron, U.S.A., Inc. v. Nat. Res. Def. Council, Inc.*, 467 U.S. 837 (1984) according to citations recognized in the Westlaw database as of November 2017 (last visited Mar. 3, 2018).

80. As of 2017, the Court has heard a number of cases expanding on or related to *Booker*. Beckles v. United States, 137 S. Ct. 886, 894 (2017) (holding that the Guidelines are not subject to challenges under the void-for-vagueness doctrine because "they merely guide the exercise of a court's discretion"); Molina-Martinez v. United States, 136 S. Ct. 1338, 1348–49 (2016) (holding that a defendant need not establish that unpreserved error affected the defendant's substantial rights); Alleyne v. United States, 133 S. Ct. 2151, 2162–63 (2013) (holding that any fact increasing the mandatory minimum is an "element" required to be submitted to a jury); Peugh v. United States, 133 S. Ct. 2072, 2081, 2088 (2013) (holding that the *Ex Post Facto* Clause prohibits sentencing offenders under Guidelines promulgated after an offender has committed a crime when the newer Guidelines provides a higher sentencing range, despite the Guidelines being advisory); S. Union Co. v. United States, 132 S. Ct. 2344, 2350, 2357 (2012) (applying *Apprendi v. New Jersey*, 530 U.S. 466 (2000), the *Apprendi* rule, to courts issuing criminal fines); Pepper v. United States, 562 U.S. 476, 504–05 (2011) (permitting district courts to consider post-sentencing rehabilitation during re-sentencing after an offender's sentence has been vacated); United States v. O'Brien, 560 U.S. 218, 224, 235 (2010) (requiring the status of a firearm as a machinegun to be proved beyond a reasonable doubt to a jury, not to be included as a sentencing factor to be proved to a judge); Dillon v. United States, 560 U.S. 817, 819, 828–30 (2010) (holding that the Guidelines are *mandatory* under an 18 U.S.C. § 3582(c)(2) resentencing proceeding when the Guidelines have been altered in favor of the offender); Spears v. United States, 555 U.S. 261, 262, 265–66 (2009) ("clarify[ing] that district courts are entitled to reject and vary categorically from the crack cocaine Guidelines based on a policy disagreement" alone after *Kimbrough*); Oregon v. Ice, 555 U.S. 160, 168 (2009) (permitting states to assign to judges finding of facts regarding whether a consecutive sentence for multiple offenses should be imposed); Moore v. United States, 555 U.S. 1, 3–4 (2008) (vacating a sentence where the judge believed the judge had no discretion to reject the crack-powder ratio); Greenlaw v. United States, 554 U.S. 237, 243–44 (2008) (prohibiting courts of appeals from increasing an offender's sentence absent a federal government appeal or cross-appeal); Irizarry v. United States, 553 U.S. 708, 712–13, 716 (2008) (holding that FED R. CRIM. P. 32(h), requiring notice when a court is considering a non-Guidelines sentence, does not apply for a regular sentence variance from the recommended guideline range); *Kimbrough*, 552 U.S. at 91 (permitting courts to depart from the Guidelines' 100-to-1 crack-powder ratio disparity); Gall v. United States, 552 U.S. 38, 41 (2007) (requiring courts of appeals to review all federal criminal sentences under a deferential abuse-of-discretion standard); Rita v. United States, 551 U.S. 338, 341 (2007) (permitting courts of appeals to apply a presumption of reasonableness to a district court sentence within the recommended guideline range); Cunningham v. California, 549 U.S. 270, 275 (2007) (striking down California's determinate sentencing law as violating *Booker*).

81. *See, e.g.*, *Kimbrough*, 552 U.S. at 91; *Gall*, 552 U.S. at 41; *Rita*, 551 U.S. at 341.

"Factors to be Considered [by a District Court Judge] in Imposing a Sentence" and includes, for example, considering what sentence is needed "to afford adequate deterrence" and "to avoid unwarranted sentence disparities among defendants with similar records who have been found guilty of similar conduct."[82] The Court held that appellate courts may adopt a rebuttable presumption that within-Guidelines sentences are reasonable.[83] The whole purpose of the Guidelines was to "seek to embody the § 3553(a) considerations."[84] As such, the guideline ranges "reflect a rough approximation of sentences that might achieve § 3553(a)'s objectives."[85] A presumption of reasonableness recognizes that both the Commission and the district court have arrived at the same conclusion.[86]

Although *district* courts cannot presume sentences to be reasonable, the Court said that within-range sentences ordinarily need no explanation other than a statement that the court is following the Guidelines rationale but that greater explanations are often necessary the further a sentence deviates from the calculated range.[87] (This statement ignored the potential problem of inter-judge disparity *within* a guideline range.) Yet the Court later prohibited any type of "rigid mathematical formula" from determining the strength of the justification that should be required for departures.[88] Using a mathematical formula would "come too close to creating an impermissible presumption of unreasonableness" for non-Guidelines sentences.[89] The worry appears to be that a presumption of unreasonableness could reintroduce the Sixth Amendment problem by making the Guidelines effectively mandatory.[90] The Court explicitly declared that a heightened appellate standard "is inconsistent with" *Booker*'s abuse of discretion standard,[91] although courts can "consider the extent of a deviation from the Guidelines" when determining whether a sentence is reasonable.[92]

Not yet done with disarming appellate review after *Booker*, the Court showed just how broad district court discretion is when it decided *Kimbrough v. United States*. There, the Court permitted district court

---

82. 18 U.S.C. § 3553(a), (a)(2)(B), (a)(6) (2012).

83. *Rita*, 551 U.S. at 347.

84. *Id.* at 350.

85. *Id.*

86. *Id.* at 347.

87. *Id.* at 356–57.

88. Gall v. United States, 551 U.S. 38, 47 (2007).

89. *Id.*

90. Brief for United States at 34–35, Rita v. United States, 551 U.S. 338 (2007) (No. 06-5754) (implying that a presumption of unreasonableness might be problematic because it might "transform an 'effectively advisory' system . . . into an effectively mandatory one" (quoting United States v. Moreland, 437 F.3d 424 (4th Cir. 2006)).)

91. *Gall*, 552 U.S. at 49.

92. *Id.* at 50.

judges to vary downward based solely on a policy disagreement the district court judge had with the 100-to-1 crack-powder ratio outlined in the Guidelines.[93] The Court extended this holding in *Spears v. United States*, when it affirmed a district court's decision not only to disagree with the Guidelines ratio but to create its own.[94]

B. *BOOKER*'S PROBLEMS

The return of judicial discretion established by *Booker* has generally been welcomed[95] (although more than a few find the case objectionable[96]). The mandatory system was too constraining in its utopian pursuit of an algorithm that could be used to sentence unique offenders. It had reacted to the tendency the indeterminate sentencing framework had of focusing "too much on individualization and not enough on avoiding unjust disparities."[97] But it weighed too heavily drug quantities and how much financial loss a crime involved and left little room for judges to provide feedback to the Commission through their sentencing.[98]

But in bringing back wide judicial discretion, the Court also destroyed any sense of meaningful appellate review. Not only is the reasonableness review of *Booker* a highly unusual standard for reviewing district court decisions,[99] but the Court promoted judicial discretion by removing the central functions of ordinary appellate review. For instance, the presumption of reasonableness in *Rita v. United States* diverges from ordinary standards of review because it is an *optional* presumption that appellate courts may make, and *Kimbrough*, by requiring reasonableness review for policy disagreements with the Guidelines, departs from the ordinary norm that legal determinations of district courts are reviewed *de novo*.[100] The Court has essentially created a safe harbor for within-range sentences, *Rita*; has prohibited courts from rigorously defining proportional justification requirements for departures, *Gall v. United States*; and has permitted virtually limitless discretion based solely on judicial philosophy, *Kimbrough/Spears*.

---

93. Kimbrough v. United States, 552 U.S. 85, 109–10 (2007).

94. Spears v. United States, 555 U.S. 261, 267 (2009).

95. *E.g.*, Baron-Evans & Stith, *supra* note 24, at 1632–33.

96. *E.g.*, Frank O. Bowman, III, *Nothing Is Not Enough: Fix the Absurd Post-*Booker *Federal Sentencing System*, 24 FED. SENT'G REP. 356, 356 (2012) (criticizing the post-*Booker* system as "retain[ing] most of the flaws of the system it replaced, while adding new ones").

97. Rachel E. Barkow, *Sentencing Guidelines at the Crossroads of Politics and Expertise*, 160 U. PENN. L. REV. 1599, 1619 (2012).

98. Baron-Evans & Stith, *supra* note 24, at 1675, 1706–07.

99. Carissa Byrne Hessick & F. Andrew Hessick, *Appellate Review of Sentencing Decisions*, 60 ALA. L. REV. 1, 14 (2008).

100. *Id.* at 19, 25.

This last point is especially important. Although the *Kimbrough* Court placed heavy emphasis on facts that were particular to *Kimbrough*—that the crack-powder ratio existed because of congressional inertia and despite the Commission's public proclamation that the ratio should be changed[101]—the Court later emphasized that sentencing judges are afforded "wide discretion" and may disregard Guidelines policies in other areas, such as when a judge considers post-sentencing rehabilitative conduct during resentencing.[102] The Court has even permitted sentencing courts to not just *disregard* policy statements, but to *substitute* into the Guidelines their own policy statements—at least with regard to crack-powder ratios[103]—and courts have read this holding as a grant of authority of vast judicial discretion to disregard Guidelines policy statements.[104] Finally, although *Kimbrough* hinted

---

101. Kimbrough v. United States, 552 U.S. 85, 99, 109 (2007); *see also* U.S. SENTENCING COMM'N, REPORT TO THE CONGRESS: COCAINE AND FEDERAL SENTENCING POLICY viii (2002) ("[T]he Commission . . . unanimously and firmly concludes that the various congressional objectives can be achieved more effectively by decreasing substantially the 100-to-1 drug quantity ratio."). The Commission had passed an amendment to equalize the crack/powder penalties, but President Clinton signed legislation nullifying the amendment. *Id.* at v.

102. Pepper v. United States, 562 U.S. 476, 480, 500–01 (2011) (quoting Williams v. New York, 337 U.S. 241, 247 (1949)). The Guidelines explicitly prohibited consideration of post-sentencing rehabilitation during resentencing. U.S. SENTENCING GUIDELINES MANUAL § 5K2.19 (U.S. SENTENCING COMM'N 2010). But the Court declared that sentencing courts must only "give 'respectful consideration'" to policy statements, a sharp post-*Booker* departure from its original holding that such policy statements were authoritative. *Pepper*, 562 U.S. at 501 (quoting *Kimbrough*, 552 U.S. at 101); *cf.* Williams v. United States, 503 U.S. 193, 201 (1992) (holding policy statements to be authoritative).

103. Spears v. United States, 555 U.S. 261, 265–66 (2009).

104. *See, e.g.*, United States v. Henderson, 649 F.3d 955, 960 (9th Cir. 2011) (permitting a downward variance because of a policy disagreement with the child pornography Guidelines); United States v. Grober, 624 F.3d 592, 608–09 (3d Cir. 2010) (same); United States v. Dorvee, 616 F.3d 174, 188 (2d Cir. 2010) (same); United States v. Herrera-Zuniga, 571 F.3d 568, 584–85 (6th Cir. 2009) (interpreting *Kimbrough* as declaring "the broad authority of sentencing judges" to "categorically reject" a calculated sentencing range); *see also* Clare Freeman, Spears v. U.S.—*Getting the Kimbrough Point Across*, SIXTH CIRCUIT BLOG (Jan. 22, 2009, 9:02 AM), http://circuit6.blogspot.com/2009/01/spears-v-us-getting-kimbrough-point.html. *But see* United States v. Pugh, 515 F.3d 1179, 1201 n.15 (11th Cir. 2008) (rejecting a downward variance for disagreements with the child pornography Guidelines).

This reading of *Kimbrough* is internally coherent with *Booker*'s creation of an advisory system and is the better reading after *Spears* and *Pepper*. Yet one can plausibly read *Kimbrough* more narrowly in the light of its emphasis on the history of the crack-powder ratio and the legislative inertia that had, at the time, prevented the Commission from crafting a new issuing. *See* Kimbrough v. United States, 552 U.S. 85, 94–100 (2007). In *Pepper*, too, the Court permitted disregarding the Guidelines policy statements in part because the Court found the statement to be "wholly unconvincing." *Pepper*, 562 U.S. at 501.

Judge Hardiman, who sits on the U.S. Court of Appeals for the Third Circuit, argues that the *Kimbrough* line should not be interpreted to allow judges to vary from the Guidelines based on policy disagreements alone. Hon. Thomas M. Hardiman & Richard L. Heppner Jr., *Policy Disagreements with the United States Sentencing Guidelines: A Welcome Expansion of Judicial Discretion or the Beginning of the End of the Sentencing Guidelines?*, 50 DUQ. L. REV. 5, 32–33 (2012). Hardiman instead counsels that, before varying, district court judges should consider whether the Guidelines reflect the Commission acting in its archetypal institutional role and whether the advisory sentencing

that "closer review" *might* be appropriate for certain policy disagreements,[105] the Court left open whether closer review *is* ever appropriate.[106] The weight of the doctrine leaves little to the imagination; sentencing judges enjoy broad discretion.

Furthermore, reasonableness review nominally permits courts to vacate sentences as substantively unreasonable, but the extremely deferential standard of review the Court has promulgated has ensured that sentences are almost never vacated for this reason.[107] Many of the sentences that are vacated as substantively unreasonable might also be mislabeled, overstating the actual number. Errors some circuits have held were substantive errors are considered by other circuits to be procedural errors.[108]

What's more, procedural error review hardly constrains a judge's substantive decisions because a judge is largely free to issue to an offender the same sentence on remand. Procedural reasonableness has been interpreted to require only that a district judge show that he or she has correctly calculated the applicable guideline range, has not treated the Guidelines as mandatory, has considered the § 3553(a) factors, and has not relied on clearly erroneous facts.[109] Yet after calculating a guideline range, a judge is fully permitted to disregard the Guidelines.

These rules establish a system of wide judicial discretion with limited appellate oversight, which was the major problem with

---

range is reasonable in light of the § 3553(a) factors. *Id.*

105. *Kimbrough*, 552 U.S. at 109.

106. Peugh v. United States, 133 S. Ct. 2072, 2080 n.2 (2013).

107. Courts of appeals ordinarily reverse barely one percent of sentences that are challenged as substantively unreasonable. U.S. SENTENCING COMM'N, 2011 SOURCEBOOK OF FEDERAL SENTENCING STATISTICS tbl.59 (2012), http://www.ussc.gov/sites/default/files/pdf/research-and-publications/annual-reports-and-sourcebooks/2011/Table59_0.pdf [hereinafter 2011 SOURCEBOOK]; U.S. SENTENCING COMM'N, 2012 SOURCEBOOK OF FEDERAL SENTENCING STATISTICS tbl.59 (2013), http://www.ussc.gov/sites/default/files/pdf/research-and-publications/annual-reports-and-source books/2012/Table59_0.pdf [hereinafter 2012 SOURCEBOOK]; U.S. SENTENCING COMM'N, 2013 SOURCEBOOK OF FEDERAL SENTENCING STATISTICS tbl.59 (2014), http://www.ussc.gov/sites/default/files/pdf/research-and-publications/annual-reports-and-sourcebooks/2013/Table59.pdf. [hereinafter 2013 SOURCEBOOK]; U.S. SENTENCING COMM'N, 2014 SOURCEBOOK OF FEDERAL SENTENCING STATISTICS tbl.59 (2015), http://www.ussc.gov/sites/default/files/pdf/research-and-publications/annual-reports-and-sourcebooks/2014/Table59.pdf [hereinafter 2014 SOURCEBOOK].

108. For instance, the Sixth Circuit considers it to be *substantively* unreasonable when a court considers impermissible factors. *See* United States v. Ward, 506 F.3d 468, 478 (6th Cir. 2007). While *Ward* has been favorably cited elsewhere, *United States v. Pugh*, 515 F.3d 1179, 1192 (11th Cir. 2008), impermissible-factors cases are largely relegated to the Sixth Circuit. That is probably because *Gall* suggests that considering impermissible factors is procedural error, not substantive error. *Gall* held that procedural unreasonableness includes failing to consider each of the § 3553(a) sentencing factors. *Gall*, 552 U.S. 38, 51 (2007). This holding implies a negative component that the § 3553(a) factors are exhaustive and that courts should not consider other factors.

   The Sixth Circuit also categorizes as substantively unreasonable a court's failure to adequately justify a sentence. *See* United States v. Borho, 485 F.3d 904, 911 (6th Cir. 2007). But *Gall* explicitly considers this failure under the umbrella of procedural unreasonableness. *Gall*, 552 U.S. at 51.

109. *Gall*, 552 U.S. at 51.

indeterminate sentencing before the Sentencing Reform Act.[110] Current doctrine now resembles a bifurcated system. The Supreme Court, concerned with the arbitrary imposition of death sentences, has used the appellate process to create greater safeguards in capital cases, but concern at the appellate stage with arbitrary results induced by inter-judge disparity in non-capital cases is lacking.[111]

This reduction in meaningful appellate review appears to have been the primary driver of increases in inter-judge disparity. *Booker*'s return of judicial discretion certainly had an effect, but the brunt of the increase did not occur until after *Rita*, *Gall*, and *Kimbrough*.[112] The needed destruction of uniformity, coupled with the decimation of meaningful appellate review, has created a system in which virtually no parts are oriented toward reducing inter-judge disparity. The only aspects that do still limit inter-judge disparity are the anchoring effects of the Guidelines and judicial inertia.[113] After *Booker*, some feared that judges would issue radically disparate sentences.[114] After all, sentencing lengths tremendously increased (and non-incarceration rates significantly decreased) after the Sentencing Reform Act and the Guidelines.[115] But the decrease since *Booker*, while measurable, has been far from severe. The mean and median sentences decreased by three to six months and may be on the rise again.[116]

---

110. STITH & CABRANES, *supra* note 4, at 82 (identifying the lack of appellate review as one of the greatest deficiencies of the indeterminate era).

111. *See* Rachel E. Barkow, *The Court of Life and Death: The Two Tracks of Constitutional Sentencing Law and the Case for Uniformity*, 107 MICH. L. REV. 1145, 1148, 1204 (2009) (stating that non-capital defendants receive "almost no Court oversight of their sentences," which "languish in a backwater devoid of any procedural protections").

112. Crystal S. Yang, *Have Interjudge Sentencing Disparities Increased in an Advisory Guidelines Regime? Evidence from* Booker, 89 N.Y.U. L. REV. 1268, 1333 (2014).

113. The Court has implicitly recognized this effect. *E.g.*, Peugh v. United States, 133 S. Ct. 2072, 2081, 2088 (2013) (holding that the *Ex Post Facto* Clause prohibits sentencing offenders under advisory Guidelines promulgated after an offender has committed a crime when the newer Guidelines provides a higher sentencing range).

The effect is also well-documented in scholarship. Birte Englich, et al., *Playing Dice with Criminal Sentences: The Influence of Irrelevant Anchors on Experts' Judicial Decision Making*, 32 PERSONALITY & SOC. PSYCHOL. BULL. 188, 194 (2006) (suggesting that giving judges starting values, even ones known to be arbitrary, has a discernible anchoring effect on judicial sentencing); Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 SCI. 1124, 1128–30 (1974) (demonstrating this psychological bias in non-judicial realms); *see also*, *e.g.*, Nancy Gertner, *Thoughts on Reasonableness*, 19 FED. SENT'G REP. 165, 167 (2007) (recognizing the effect); Stephanos Bibas et al., *Policing Politics at Sentencing*, 103 NW. U. L. REV. 1371, 1387 (2009) (same); Michael M. O'Hear, *Appellate Review of Sentence Explanations*: *Learning from the Wisconsin and Federal Experiences*, 93 MARQ. L. REV. 751, 758 (2009) (same).

114. Carl Hulse & Adam Liptak, *New Fight over Controlling Punishments Is Widely Seen*, N.Y. TIMES, Jan. 13, 2005, at A29 (identifying some commentators as calling *Booker* an "egregious overreach" and saying it will lead to "wildly inconsistent" outcomes while others praise the decision for allowing greater leniency).

115. STITH & CABRANES, *supra* note 4, at 5–6.

116. Bowman, *supra* note 10, at 1236.

But even inertia and anchoring effects are unlikely to stem the flow of inter-judge disparity. The relative stability of sentences is true only in the aggregate, not across the board. The means and medians are level, but only because firearm and economic sentences increased significantly to counter the decrease in drug and immigration sentences.[117] And the aggregate means and medians mask the variation between judges. So, too, mandatory minimums cloud the data by making median sentences appear more uniform, disguising the true effect of *Booker* by mixing into the same pot sentences where judges are statutorily constrained and sentences where judges are free to disregard the Guidelines.[118] Most notably, even with a high adherence to the Guidelines, geographic disparity and inter-judge disparity have increased measurably.[119] Quite apart from the hard data, the anecdotal evidence has been tremendous.[120] Professor Frank Bowman concludes that "uncontrolled disparity will again become a rallying cry" because inter-judge disparity "should matter to anyone who believes that one important component of a just system of criminal punishment is that similarly situated offenders are treated substantially similarly."[121]

Inertia and anchoring effects are also unlikely to stem the tide of inter-judge disparity because those effects are weaker on new judges. Immediately after *Booker*, nearly all judges were used to the mandatory Guidelines regime and—quite apart from the anchoring effect of the Guidelines—may have remained relatively inert in their sentencing practices.[122] But by the end of his term, President Obama had appointed more than one-third of the district court bench.[123] To the extent the inertia hypothesis is correct, inter-judge disparity should increase as

---

117. Bowman, *supra* note 10, at 1241.

118. Bowman, *supra* note 10, at 1257.

119. Bowman, *supra* note 10, at 1261.

120. Scott, *supra* note 5, at 3–4.

121. Bowman, *supra* note 10, at 1267–68.

122. Scott, *supra* note 5, at 5 (finding two judges in the District of Massachusetts to be relatively inert in their sentencing practices). However, Scott found that no greater inertia effect occurred for judges who were appointed after 1987 than those appointed before; both sentenced below the Guidelines more than thirty percent of the time in the post-*Booker* era. Scott, *supra* note 5, at 43–44.

These findings do not necessarily counter the conjecture that newly appointed judges will abide by the Guidelines less. Scott's hypothesis is that judges appointed before 1987 "might cast off the yoke of the Guidelines more readily." Scott, *supra* note 5, at 43. But even if a judge had enjoyed greater sentencing freedom before 1987, by the time *Booker* was decided, the judge would have been sentencing under a mandatory regime for nearly two decades. The difference between a pre-1987 judge and a post-1987 (but pre-*Booker*) judge is not terribly meaningful. But the difference between a pre-*Booker* judge and a post-*Booker* judge is tremendous. One should not expect a pre-1987 judge to sentence much differently from a judge appointed in, say, 1999. But one should expect that both will sentence very differently from a judge who never sentenced under a mandatory regime.

123. *Judgeship Appointments by President*, U.S. Cts. http://www.uscourts.gov/judges-judgeships/authorized-judgeships/judgeship-appointments-president (last visited Mar. 3, 2018).

pre-*Booker* judges retire and are replaced by post-*Booker* judges.[124] If one cares about inter-judge disparity, the time for action is yesterday. A more meaningful form of appellate review is necessary.

C. RECENT STUDIES DEMONSTRATING INTER-JUDGE DISPARITY

Measuring inter-judge disparity has traditionally been notoriously difficult.[125] The Commission gathers a wealth of data, compiling extensive information for almost all of the 70,000-plus defendants each year. This data includes the guidelines calculations, statutory offense category, and as many as thousands of other data points (depending on how the judge sentences the offender). The problem is the Commission removes all identifying information before publicly releasing the data, which makes measuring the differences in sentences by judges more difficult.[126] Even where identifying information is available, studies have had difficulty controlling for relevant sentencing characteristics.[127] Several recent studies have nevertheless overcome these boundaries to show the increasing significance of inter-judge disparity.

Professor Ryan Scott surveyed the data released by judges in the Boston division of the District of Massachusetts. The only district at the time to publicly release its data with judge identifiers,[128] the district opened up a critical insight into a world before unknown. The results are

---

124. *See* Gertner, *supra* note 16, at 270 (identifying adherence to the Guidelines post-*Booker* as the result of "habits ingrained during twenty years of mandatory Guideline sentencing"); Yang, *supra* note 112, at 1319 (noting that as newer judges take the bench, the Guidelines are likely to have less of an anchoring effect).

125. Scott, *supra* note 5, at 21 (noting that "changes are almost impossible to detect"); Caleb Mason & David Bjerk, *Inter-Judge Sentencing Disparity on the Federal Bench: An Examination of Drug Smuggling Cases in the Southern District of California*, 25 FED. SENT'G REP. 190, 190 (2013) (identifying as "principal shortcomings" of these measurements the difficulty in either identifying judges or controlling for "relevant offense-conduct facts").

126. U.S. SENTENCING COMM'N, PUBLIC ACCESS TO SENTENCING COMMISSION DOCUMENTS AND DATA, 54 Fed. Reg. 51279 (Dec. 13, 1989) (delineating an agreement between the Commission and the Judicial Conference to remove all judge and defendant identifiers from publicly released datafiles); U.S. SENTENCING COMM'N, GUIDE TO PUBLICATIONS & RESOURCES 2010/2011 28 (2010), ("Pursuant to the policy on public access to Sentencing Commission documents and data, all case and defendant identifiers have been removed from the data." (internal citation omitted)).

127. *E.g.*, Anderson et al., *supra* note 9, at 274 (employing a methodology that considered "solely that variation caused by the identity of the decision maker," as opposed to variation in punishment among different classes of offenders); Paul J. Hofer et al., *The Effect of the Federal Sentencing Guidelines on Inter-Judge Sentencing Disparity*, 90 J. CRIM. L. & CRIMINOLOGY 239, 282 (1999) ("We recognize that this measures only the 'tip of the iceberg' of disparity created by judges, because it fails to account for interaction effects between judges and particular offense and offender characteristics."); Joel Waldfogel, *Aggregate Inter-Judge Disparity in Federal Sentencing: Evidence from Three Districts* (D. Ct., S.D.N.Y., N.D. Cal.), 4 FED. SENT'G REP. 151, 151–52 (1991). *But see* Mason & Bjerk, *supra* note 125, at 191–92 (employing a methodology that allowed the researchers to limit the cases to offenders who had dealt in drug quantities triggering mandatory minimums but were eligible for the sentencing safety valve and received reduced sentences).

128. Scott, *supra* note 5, at 4.

striking. Limiting his analysis to 10 active-status judges who had sentenced at least 50 defendants,[129] Scott found that the 3 most lenient judges impose sentences that average 25.5 months or less; the other 2 judges impose sentences that average more than double (51.4 months) the more lenient judges. Not unsurprisingly, four judges now sentence below the guideline ranges at more than triple their pre-*Booker* rates.[130]

Following the Boston Study, the District of Nebraska voted to release its sentencing data to the public.[131] In Nebraska, sentences are more likely to be uniform because judges impose within-range sentences at rates much higher than the national average, yet one judge sentenced drug offenders at twice the median rate of a second. Although this effect might be overstated because the stricter judge had sentenced fewer drug offenders and had greater control over the docket because the judge had taken senior status, the difference in median rate between the two active judges with a comparable number of sentences was still twenty-two percent.[132]

Researchers in a third study manually collected data by accessing complaints and dockets in the Southern District of California, allowing the researchers to identify the judges.[133] But rather than look at average or median sentences, the researchers employed a new methodology. They recognized that offenders with somewhat disparate drug quantities can receive the same guideline range calculations. For instance, the drug quantity table assigns the same offense level to individuals who deal fifteen kilograms of cocaine as to those who deal forty-nine kilograms of cocaine,[134] yet one would expect judges to sentence the latter somewhat more heavily. The researchers isolated their data so they were looking at offenders who differed *only* in type and quantity of drugs.[135] This practice allowed them to use regression analysis to create an "expected" sentence for each drug quantity and to measure how far from the expected sentence each judge deviated.[136]

The researchers found "substantial disparity."[137] Ten of thirteen judges had low variance from the expected sentence, but two judges were noticeably more lenient (sentencing around seventeen to eighteen

---

129. Scott, *supra* note 5, at 25.

130. Scott, *supra* note 5, at 4–5.

131. Kopf, *supra* note 2, at 50.

132. Kopf, *supra* note 2, at at 51 app. tbl. 7 (Drug Trafficking (2007–2011)).

133. Mason & Bjerk, *supra* note 125, at 191.

134. U.S. Sentencing Guidelines Manual § 2D1.1(c)(4) (U.S. Sentencing Comm'n 2016).

135. The researchers ensured they were measuring similarly situated offenders by limiting their analysis to individuals who were given sentences below mandatory minimums despite carrying drug quantities that would trigger a mandatory minimum but for the offenders being eligible for a safety valve that prevented the mandatory minimum, a benefit available only to offenders with criminal history scores of one. Mason & Bjerk, *supra* note 125, at 191.

136. Mason & Bjerk, *supra* note 125, at 191–92.

137. Mason & Bjerk, *supra* note 125, at 196.

percent below the "expected" sentence) and one was noticeably harsher (sentencing around thirty percent above the "expected sentence").[138] These differences "reflect individual judges' preferences about the appropriate sentence for the same crime."[139]

Another study launched by the Transactional Records Access Clearinghouse ("TRAC") program at Syracuse University[140] acquired most of the sentencing data covering a five-year period: roughly 370,000 cases nationwide.[141] Because the Sentencing Commission does not voluntarily release data that identifies judges, TRAC gathered this data from other sources, such as the Office of Personnel Management[142] and, evidently, PACER.[143] Like Scott's study, TRAC limited its data to judges who had sentenced at least fifty defendants, and TRAC subdivided the data into categories of judges who had served continuously during the five-year post-*Booker* span and for judges who had continuously served on active status during that time.[144]

Although some of the districts maintained low variance, the results were largely consistent with all other studies. The median sentences of judges in Dallas, Texas, varied between 60 months and 121.5 months; in Fort Worth between 102.5 months and 160 months; and in the District of Columbia between 27 and 77 months.[145]

The most crucial and developed study comes from Professor Crystal S. Yang, who built upon the work of the TRAC Report and Scott's study. In merging the TRAC Report data with sentencing data the Commission makes available, Yang identified the sentencing judge for more than 400,000 sentences, eliminating some of the most serious hindrances to studying inter-judge disparity.[146] The most comprehensive study to date, Yang's study found that the effect of inter-judge disparity doubled in the post-*Booker* period.[147]

D. Limits of These Studies

These studies are not without their limits, and each included caveats about interpreting the findings too broadly.[148] But each also shows

---

138. Mason & Bjerk, *supra* note 125, at 193, 195.

139. Mason & Bjerk, *supra* note 125, at 192.

140. *Surprising Judge-to-Judge Variations Documented in Federal Sentencing*, TRAC REP. (Mar. 5, 2012), http://trac.syr.edu/tracreports/judge/274.

141. Susan B. Long & David Burnham, *TRAC Report: Examining Current Federal Sentencing Practices: A National Study of Differences Among Judges*, 25 FED. SENT'G REP. 6, 6 (2012).

142. *Id.* at 7.

143. Hofer, *supra* note 8, at 37.

144. Long & Burnham, *supra* note 141, at 7.

145. Long & Burnham, *supra* note 141, at 9–10.

146. Yang, *supra* note 112, at 1275, 1296.

147. Yang, *supra* note 112, at 1275.

148. *E.g.*, Yang, *supra* note 112, at 1292 (noting that "a single courthouse is likely unrepresentative of other courthouses across the United States"); Long & Burnham, *supra* note 141, at 15 (stating that

increases in inter-judge disparity, no study has shown an absence of an increase, and the sources of anecdotal evidence are tremendous.[149]

The limits do not overcome these findings. The first major limit is that each of the studies assumes that cases are randomly assigned. Recent history indicates potentially significant problems (for the purposes of data collection) in the way cases are assigned.[150] Additionally, prosecutors are not necessarily assigned randomly, and they may tailor their conduct to the judges they draw.

Second, each study limited itself to judges who had sentenced more than fifty defendants, but only the TRAC study further qualified its data by sentence *type* (although it considered forty sentences instead of fifty).[151] On the surface, fifty seems like a reasonable number—introductory statistics books often paint thirty as the magic number[152]—but fifty may be too small if judges received skewed dockets. For instance, one judge may receive fifty randomly allotted sentencings that proportionately represent the standard docket of cases while another judge may receive a thirty-defendant drug trafficking case where many defendants are subject to mandatory minimums. Yet these sentencing patterns are most likely rare and would largely affect judges who had sentenced relatively few offenders. Many of the judges in the studies sentenced far more than fifty offenders. The Scott study also ameliorated this difficulty by looking at the difference between the sentence imposed and the sentence suggested by the calculated guidelines.[153]

---

the large differences in median sentences "is not sufficient to establish that such differences are indeed *unwarranted sentencing disparities*").

149. Scott, *supra* note 5, at 3–4.

150. *See, e.g.*, Benjamin Weiser & Joseph Goldstein, *Federal Court Alters Rules on Judge Assignments*, N.Y. TIMES, (Dec. 23, 2013), www.nytimes.com/2013/12/24/nyregion/federal-court-alters-rules-on-judge-assignments.html (asserting that the litigation over the stop-and-frisk policy in New York City has exposed problems in supposed random assignment); Katherine A. Macfarlane, *The Danger of Nonrandom Case Assignment: How the Southern District of New York's "Related Cases" Rule Shaped Stop-and-Frisk Rulings*, 19 MICH. J. RACE & L. 199, 203 (2014) (stating that one judge was randomly assigned a single case and then received all "related" cases); Joe Palazzolo, *The Problem with Not-So-Random Case Assignment*, WALL ST. J. (NOV. 4, 2013, 4:44 PM), http://blogs.wsj.com/law/2013/11/04/the-problem-with-not-so-random-case-assignment (discussing a congressionally authorized pilot program for judges to take related cases from other judges in the name of judicial efficiency).

151. Long & Burnham, *supra* note 141, at 9.

152. *See, e.g.*, GREGORY W. CORDER & DALE I. FOREMAN, NONPARAMETRIC STATISTICS FOR NON-STATISTICIANS: A STEP-BY-STEP APPROACH 2 (2009) (stating that most researchers choose sample sizers at least as large as thirty); Alan Agresti & Yongyi Min, *On Sample Size Guidelines for Teaching Inference About the Binomial Parameter in Introductory Statistics* 1 (Aug. 15, 2002) (unpublished manuscript), http://www.stat.ufl.edu/~aa/articles/agresti_min_binomial.pdf (identifying the similarity a population distribution obtains with a normal distribution when sample sizes are thirty or greater as the reason for suggesting a sample size of thirty in the first place).

153. Scott, *supra* note 5, at 25.

Finally, the most noteworthy objection is that the data on which the studies relied is unreliable. A random spot check conducted by Federal Defenders of the TRAC Reports revealed several inaccuracies, and TRAC appears to have taken some of its data from documents that are compiled for the purpose of budgetary requests, not sentencing research.[154] To the extent budgetary documents carry their own biases, the data may be less reliable. What's more, the Commission draws its data from paperwork completed by courts, but judges do not always fill out those reports; courtroom deputies or probation officers sometimes do.[155] And there are often multiple ways for a person filling out the paperwork to represent the same sentencing factor. Despite these difficulties, nobody has yet introduced a better proposal for measuring inter-judge disparity. And the Commission and Congress consider the data good enough for the Commission to satisfy its duty to research and report on sentencing patterns.[156] Moreover, as explained in greater detail below, the proposal in this Article does not depend on the precision of the data. Rather, the proposal uses the data to shift the focus of judges' onto the sentencing patterns of other judges on the same court.[157] Even if the data suffers from reliability issues, it is still sufficient to create this paradigm-shifting function because the benefit of this Article's proposal is intended to be felt more in the aggregate than on the individual level.

Although good reasons exist to be skeptical of these studies, the studies' consistent findings coupled with the knowledge that *Booker* took away the greatest weapon under the Sentencing Reform Act against inter-judge disparity make it easy to conclude with high confidence that inter-judge disparity has become more problematic in the post-*Booker* world. It is not an understatement to say that the weight of scholarly authority strongly favors the notion that *Booker* has caused an increase in inter-judge disparity.

II. A Proposal for Presumed Unreasonableness

The weight of the evidence above shows that inter-judge disparity has increased and will likely grow worse as newer judges who never sentenced under the mandatory Guidelines take the bench. But a return to a mandatory regime[158] would pursue a reduction in inter-judge

---

154. Sentencing Res. Counsel Project, Fed. Pub. Defs., *TRAC's Report Claiming "Surprising Judge-to-Judge Variation" Fails to Compare Similar Cases, Relies on Poor Quality Data, Uses an Unreliable Method of Identifying Case Type, Uses Incorrect Methods of Reporting Sentence Length, and Contains Numerous Errors*, 25 Fed. Sent'g Rep. 20, 25–26 (2012).

155. Nancy Gertner, *Judge Identifiers, TRAC, and a Perfect World*, 25 Fed. Sent'g Rep. 46, 47 (2012).

156. 28 U.S.C. §§ 995(a)(12), (14)-(16) (2012).

157. *See* Part II; Part III.A.3.

158. Congress could constitutionally return to a mandatory regime by adopting, through statute, the remedy proposed by the main *Booker* dissent: providing for mandatory Guidelines only when the facts necessary to a sentence are either admitted to by the defendant or found beyond a reasonable

disparity at the expense of creating unwarranted uniformity. After the mandatory Guidelines era, there is little reason to suggest that sentencing can be pressed into a rigorous, scientific system that places a quantifiable weight on every single relevant sentencing factor.

Instead, reducing inter-judge disparity must be pursued through *indirect* means that have ancillary effects on sentencing behavior. The problem that caused an increase in inter-judge disparity was not *Booker* alone; it was the subsequent reduction of any meaningful system of appellate review. Increasing appellate review is the logical reform, but the details are elusive. Notably, increasing appellate review raises some constitutional issues. A system of appellate review that is (1) tied to the Guidelines and (2) so constraining as to make the Guidelines system de facto mandatory likely violates the Sixth Amendment as construed by the Court since *Booker*.[159]

It follows that an effective form of stricter appellate review should be independent of the Guidelines. But to adequately reduce inter-judge disparity, the appellate review system must also induce judges to provide greater justifications for their sentences and make judges more aware of their own sentencing tendencies in relation to the sentences their peers impose. Simply knowing how a judge's peers sentence should have an anchoring effect on most judges,[160] and creating a rebuttable presumption of unreasonableness should induce judges to engage more critically with the distinctions between each specific sentence, an effect that would itself reduce inter-judge disparity.

This reform could be achieved under existing law, which requires judges to consider "the need to avoid unwarranted disparities among" similarly situated offenders.[161] But because the Supreme Court has layered onto appellate review an abuse-of-discretion standard,[162] rendering appellate review effectively toothless, and because the record in a given case rarely includes evidence of what sentences similarly situated offenders received, the best chance for achieving this reform lies in a statutory amendment. Congress should require appellate courts to presume that a sentence is unreasonable if the sentence falls too far from the median sentence similarly situated offenders receive. (Part IV explores the methodology the Commission should adopt in defining categories of similarly situated offenders.) This statute could be added to 18 U.S.C. § 3742 to establish the following:

---

doubt by a jury.

159. *See* Gall v. United States, 552 U.S. 38, 47 (2007) (prohibiting a presumption of unreasonableness for sentences outside the Guidelines).

160. *See, e.g.*, Tversky & Kahneman, *supra* note 113, at 1128 (establishing that study participants' estimates of the percentage of African countries in the United Nations were strongly influenced by what number arbitrarily appeared in a game of chance before they made their estimates).

161. 18 U.S.C. § 3553(a)(6) (2016).

162. *Gall*, 552 U.S. at 41.

> **Presumption of Unreasonableness**: A court of appeals shall presume a sentence is unreasonable when it departs too far from the median sentence for like offenders. This presumption can be rebutted if the court of appeals finds that the sentencing judge's explanation is sufficiently compelling.
>
> The Sentencing Commission shall have the authority to define categories of like offenders and determine how far a sentence may vary before being presumed unreasonable.[163]

Congress could itself determine exactly how to define categories of like offenders, but it is likely the specific details may be crime-dependent and that the Sentencing Commission, as the institution with greater expertise in this matter, is in a better position to tailor specific formulas to specific crimes.

The benefits of the proposal are explored in detail in Part III. Briefly, this reform would accomplish two important things: it would avoid constitutional problems by creating an appellate system wholly outside the Guidelines (but which uses data created by the Guidelines system), and it would create a system that has indirect ameliorating effects on inter-judge disparity. By allowing judges to overcome the presumption of unreasonableness, this reform would create a system where judges would be required to include more detailed and comprehensive explanations for their sentences. Although the § 3553(a) factors already require judges to avoid unwarranted disparities among similarly situated offenders,[164] this proposal would require even greater justification when a judge decides to sentence an offender too far from the median.

This Article suggests that courts presume that a sentence is unreasonable when the sentence falls within the bottom or top fifteen percent of sentences that similarly situated offenders receive. (An exception might be made when the difference between the bottom and top fifteen percent is, in absolute numbers, small. Inter-judge disparity carries comparatively less harm in those situations.) This number is exclusive: A sentence length would not be presumed unreasonable for falling in the ninetieth percentile if the same length also fell within the eighty-fourth percentile. The fifteen-percent figure is chosen for illustrative purposes. The figure might be too blunt to rectify inter-judge disparity across the board, or a more robust statistical metric, such as

---

163. Congress would also need to amend 28 U.S.C. § 994, which concerns the powers of the Commission, to read as follows:

> The Sentencing Commission shall promulgate and distribute methods for defining categories of offenders who are similarly situated to the instant defendant. In doing so, the Commission shall detail which of the Commission's individual offender database variables shall be used, how many years of sentencing data should be included in the sample, what the minimum sample size of similarly situated defendants should be, how to treat Guidelines amendments, and any other factors that would be appropriate to implement the appellate system of review spelled out in 18 U.S.C. § 3742.

164. 18 U.S.C. § 3553(a)(6) (2016).

median absolute deviation, might be more appropriate.[165] This Article uses the fifteen-percent figure merely for illustrative purposes.

A few safeguards may also be established to avoid placing too much focus on this reform at the expense of the guidelines and to avoid drowning the already voluminous dockets of many courts. This proposal affords no official place for sentencing data at the initial sentencing decision, so this proposal should not increase litigation costs at that stage. And the proposal can avoid excessive appeals—federal appellate courts already receive more than 3000 procedural reasonableness challenges each year[166]—by making appeal permissive only. Defendants will not be permitted to appeal unless they are granted something similar to a certificate of appealability in the habeas context by a neutral federal authority whose job it is to review each sentence and notify those defendants who are eligible to appeal. Part IV shows that calculating whether a sentence is presumptively unreasonable should take very little time. Permissive appeal is appropriate because the system of appellate review proposed in this Article is designed to affect inter-judge disparity at the aggregate level. Appeals are the vehicle the proposal uses to obtain a shift in the sentencing conversation, encourage judges to be more thorough in their sentencing habits, and indirectly inform judges about the sentencing habits of their peers in ways that will mitigate the effects of inter-judge disparity on all offenders.

III. THE BENEFITS OF THE PROPOSAL

A. DISTINCTIONS FROM EXISTING PROPOSALS

Before outlining in detail the merits of this proposal and defending its constitutionality, it is worth exploring how this proposal differs from other suggested proposals and avoids some of the weaknesses that affect those proposals.

1.   *Systematic Reforms*

Some reforms suggested shortly after the Supreme Court decided *Booker* are now foreclosed by precedent,[167] but others remain. The

---

165. Using the median absolute deviation—which measures the statistical dispersion centered around the median instead of the mean—instead of the standard deviation, would likely be more appropriate because this more robust statistical measure better protects against the influence of outlier sentences. The sentencing data does not fall into a normal distribution, where the mean and median would be the same, because of a number of factors: The Guidelines exercise a large anchoring effect on judges, and there is a strong inclination among judges to sentence at the bottom of a calculated range when judges do not depart. *Supra* notes 25-26 and accompanying text. Sentencing data does not come close to resembling a normal distribution, so the median is a more appropriate tool of analysis because it diminishes the significance of outlier sentences such as life sentences.

166. 2013 SOURCEBOOK, *supra* note 107, at tbl.59; 2014 SOURCEBOOK, *supra* note 107, at tbl.59.

167. Several individuals, including Professor Frank Bowman and then-Attorney General Alberto

Constitution Project Sentencing Initiative, which included then-Judge Alito, proposed that the Guidelines effectively be re-written. The sentencing ranges would be wider; the adjustments would be found by a jury (or admitted to by the offender); and judges would enjoy more permissive departures than under the Guidelines.[168] Judge William Sessions, who served on the Commission from 1999–2010, proposed a similar reform.[169] Sessions believed advisory Guidelines "are by definition unenforceable and thus allow for the emergence of sentencing disparities that motivated many American sentencing reforms in the first instance."[170] He pushed for a binding approach with "sub-ranges" within the wider ranges; the wider ranges would be mandatory, and judges would use the presence or absence of aggravating factors to sentence within sub-ranges.[171] Departures "would need to be based on truly extraordinary mitigating circumstances" and would be subject to "relatively strict" appellate review; appellate courts would review the sufficiency of the evidence for whether mandatory aggravating factors were proved to a jury.[172]

Judge William Pryor, who currently serves as Acting Chair of the U.S. Sentencing Commission, recently proposed a reform similar to the reform Sessions proposed. Like Sessions's reform, Pryor's would be much simpler than the current Guidelines and would include broad, effectively binding ranges, which he calls "presumptive."[173] "[O]nly the most important and common aggravating and mitigating factors" would remain for calculating a guidelines range, and a jury would have to find those factors.[174] The more complicated and *sui generis* aggravators and mitigators would be moved to the commentary section of the Guidelines

---

Gonzalez, proposed "topless guidelines" where judges would be bound by the minimum calculated guideline ranges but would have advisory maximum ranges. Douglas A. Berman, *Tweaking* Booker*: Advisory Guidelines in the Federal System*, 43 Hous. L. Rev. 341, 356–64 (2006). This form of guidelines is, under current doctrine, undoubtedly unconstitutional. *See* Alleyne v. United States, 133 S. Ct. 2151, 2162–63 (2013) (requiring any factor that increases the mandatory minimum of a crime to be proved before a jury).

168. CONSTITUTION PROJECT SENTENCING INITIATIVE, *Recommendations for Federal Criminal Sentencing in a Post-*Booker *World*, 18 FED. SENT'G REP. 310, 314–17 (2006); *see also* Bowman, *supra* note 96, at 364 (summarizing the plan).

169. Sessions III, *supra* note 17, at 337–53. Bowman endorsed both Sessions' and the Constitution Project Sentencing Initiative's proposals. Bowman, *supra* note 96, at 364.

170. Sessions III, *supra* note 17, at 339 (internal quotation marks omitted) (citing Kevin R. Reitz, *The New Sentencing Conundrum: Policy and Constitutional Law at Cross-Purposes*, 105 COLUM. L. REV. 1082, 1114 (2005)).

171. Sessions III, *supra* note 17, at 347.

172. Sessions III, *supra* note 17, at 351, 354.

173. William H. Pryor, Jr., *Returning to Marvin Frankel's First Principles in Federal Sentencing*, 29 FED. SENT'G REP. 95, 98 (2017).

174. *Id.*

and would affect only where *within* a "presumptive" guideline range a judge decided to sentence a defendant.[175]

The Commission has proposed several reforms. First, it asked for a statutory change that would require judges to abide by a "three-step" process: judges would first calculate the Guidelines, consider the Commission's commentary and policy statements—which are strongly oriented toward preventing departures—and finally consider the § 3553(a) sentencing factors.[176] Second, the Commission asked Congress to codify its policy statements that currently conflict with congressional statutory policy. Congress requires the Commission to "assure the guidelines . . . reflect the general inappropriateness of considering the education, vocational skills, employment record, family ties and responsibilities, and community ties of the defendant."[177] But Congress also requires judges to consider the "history and characteristics of the defendant,"[178] which has created a "tension" between statutory law and the provisions in the Guidelines that instruct judges to ignore certain offender characteristics.[179] Finally, and most importantly, the Commission asked Congress to create a presumption of reasonableness for within-Guideline sentences, require proportionally "greater justification[s]" when sentences are further outside the guideline range (abrogating *Gall*), and require appellate courts to apply a heightened standard of review when district courts disagree with the Commission's policy statements.[180]

The American Bar Association's proposal adopts the position taken by the main *Booker* dissent, which argued that the correct remedy in *Booker* was to require the government to prove aggravating elements to juries.[181] But because presenting to a jury every single relevant element of an offense is procedurally unmanageable, the ABA proposes that the Guidelines be greatly simplified to reduce the number of adjustments,

---

175. *Id.*

176. *Uncertain Justice: The Status of Federal Sentencing and the U.S. Sentencing Commission Six Years After* U.S. v. Booker*: Hearing Before the H. Subcomm. On Crime, Terrorism, and Homeland Sec. of the H. Comm. on the Judiciary*, 112th Cong. 55–60 (2011) (statement of Judge Patti B. Saris, Chair, U.S. Sent'g Comm'n), http://www.ussc.gov/sites/default/files/pdf/news/congressional-testimony-and-reports/testimony/20111012_Saris_Testimony.pdf.

177. 28 U.S.C. § 994(e) (2012). The Commission has adopted this policy. U.S. SENTENCING GUIDELINES MANUAL § 5H, introductory cmt. (U.S. SENTENCING COMM'N 2016).

178. 18 U.S.C. § 3553(a)(1) (2016).

179. *Uncertain Justice*, *supra* note 176, at 57; *see also Improving the Advisory Guideline System Public Hearing Before the U.S. Sentencing Comm'n* app. question 4 (Feb. 16, 2012) (statement of Henry Bemporad, Fed. Pub. Def., W. Dist. of Tex.), https://www.fd.org/sites/default/files/criminal_defense_topics/essential_topics/sentencing_resources/defender_recommendations/bemporad_statement_2_16_12.pdf (requesting that Congress codify the Commission's policy statements that education, family ties, etc., are "not ordinarily relevant").

180. *Uncertain Justice*, *supra* note 176, at 55–56.

181. Berman, *supra* note 167, at 365.

offense levels, and other complex factors.[182] Under this proposal, a prosecutor who sought to obtain a higher sentence for a robbery would have to allege in an indictment, "in addition to the elements of robbery, whether the defendant possessed a firearm, whether he brandished or discharged it, [or] whether he threatened death," among other things.[183]

A separate proposal calls not for a new set of guidelines but instead a reformulation of what it means for a judge or jury to make fact findings. Douglas Berman and Stephanos Bibas argue that a distinction exists between facts about the offender and facts about the offense.[184] According to Berman and Bibas, juries have historically discerned offense facts and should continue to do so, but courts are better equipped to discern offender facts. The implication of the distinction is that a legislature might establish guidelines that bind judges based on facts judges discern that relate to the offender, not to the offense.

Whatever merits or demerits these proposals have on other grounds, they are poorly suited to attacking inter-judge disparity. The proposals offered by Sessions, and the Commission re-create the Guidelines in much of their complex rigor. Amy Baron-Evans and Professor Kate Stith characterize the Commission's proposal as one meant to "undo the holdings of the Supreme Court in *Booker* and its progeny and to reestablish the Commission's guidelines and policy statements as the 'law' of sentencing."[185] Pryor's and the ABA's are much simpler, but they too lack the ability to attack inter-judge disparity head-on. To the extent any of these proposals will increase uniformity, they will have an incidental and marginal effect on inter-judge disparity as well, but the wide ranges that would be established under these proposals suggest that inter-judge disparity would remain high. At the price of avoiding unwarranted uniformity—a noteworthy goal—these proposals are unable to fully address inter-judge disparity.

Some of these proposals also operate on constitutionally tenuous grounds under the current doctrine of the Supreme Court. The Court has exercised tremendous influence in the area of reasonableness review under the Guidelines, and it is not necessarily clear that a stricter system would not draw out the Sixth Amendment issue *Booker* tried to put away. Several scholars argue that *Booker* and its progeny create a delicate balance between judicial discretion and uniformity, a balance that can be easily upset.[186] Baron-Evans and Stith have directly declared Sessions'

182. Am. Bar Ass'n, Crim. J. Sec., *Report on* Booker *and Recommendation*, 17 FED. SENT'G REP. 335, 339 (2005).

183. United States v. Booker, 543 U.S. 220, 254 (2005).

184. Douglas A. Berman & Stephanos Bibas, *Making Sentencing Sensible*, 4 OHIO ST. J CRIM. L. 37, 38 (2006).

185. Baron-Evans & Stith, *supra* note 24, at 1731.

186. Hessick, *supra* note 24, at 1336, 1348 ("[T]here is little doubt that this recommendation would shift the delicate balance the Court has struck after *Booker* away from district court discretion towards

and the Commission's proposals unconstitutional.[187] And the Commission's proposal to require appellate courts to presume that within-Guidelines sentences are reasonable directly contradicts *Rita*'s holding that such presumptions be *permitted* out of a "desire to avoid creating a bias for within-Guidelines sentences."[188] The Commission's proposal creates that bias. It and other proposals may very well be constitutional, but a rigorous assessment of constitutionality is required, and the proposals largely lack constitutional analysis.

2. *Calls for Stronger Appellate Review*

Arguably, "*Booker* left unclear exactly how loose appellate review must be to satisfy the Sixth Amendment."[189] Some scholars maintain that Congress should require appellate courts to apply a sufficiency-of-the-evidence standard to judicial fact-finding.[190] Additionally, then-Professor Stephanos Bibas argues that Congress should create sentencing courts, prohibit defendants from waiving their rights to appeal, and require political diversity on appellate panels.[191] Professor Crystal S. Yang argues for courts to require a "heightened justification for more severe departures."[192]

These scholars are generally on the right track. The problem with the post-*Booker* system is the same as the main problem that afflicted indeterminate sentencing: the system lacks meaningful appellate review. But these scholars call for greater appellate review without fleshing out what the systems would look like or defending the constitutionality of the systems, which is critical because *Booker* struck down stronger appellate review and other scholars have identified constitutional problems with increasing appellate review.[193] Bibas, for example, readily acknowledges that precedent bars the sufficiency-of-evidence proposal.[194] Furthermore, none of these proposals are well-tailored toward reducing inter-judge disparity. Yang limits her form of appellate review to sentences outside guideline ranges, which would prevent her form of

---

the Guidelines."); Hessick & Hessick, *supra* note 99, at 3–4 (describing the balance as increasing judicial discretion to avoid the Sixth Amendment problem, but creating some appellate review to pursue uniformity).

187. Baron-Evans & Stith, *supra* note 24, at 1716–31 (arguing that *Booker* requires any Guidelines system to be advisory).

188. Hessick, *supra* note 24, at 1347.

189. Bibas et al., *supra* note 113, at 1372–73.

190. Bibas et al., *supra* note 113; O'Hear, *supra* note 113, at 752 (arguing that courts should create an "explanation review").

191. Bibas et al., *supra* note 184, at 1373.

192. Yang, *supra* note 107, at 1333.

193. Baron-Evans & Stith, *supra* note 24, at 1731.

194. Bibas et al., *supra* note 113, at 1396 ("[T]he Court's jurisprudence is deeply misguided. Binding guidelines and searching appellate review are needed to make sentencing more consistent and legitimate.").

review from considering inter-judge disparity that occurs within guideline ranges.

3. *Trailing Edge Guidelines and Statistical Curving*

Professor Mark Osler describes uniformity and discretion as historically antagonistic to each other. To solve this problem, he suggests that the government create a computer system through which judges enter specific sentencing data to see what sentences similarly-situated offenders received. This notably would achieve some of the same goals as this system—namely, making judges more aware of the sentences issued by other judges. Osler also suggests that appellate courts apply a heightened standard of review when sentences stray too far from sentences of those other judges.[195]

In Osler's system, however, the common sentencing practices of judges would *become* the Guidelines, and the Commission's role would be limited to defining mitigating or aggravating circumstances without placing numerical offense levels on those circumstances.[196] Although Osler's proposal has some appeal because the Commission claims to have created the Guidelines in the first place based on historical sentencing practice,[197] the proposal in this Article pursues a different tactic in response to information derived from a review of the sentencing data.

Foremost, this Article's proposal permits a greater degree of forward-looking modifications to the Guidelines. Osler's model seeks to institute a method that would provide feedback regarding sentencing practices,[198] but it also risks entrenching past practices by ensuring that previous sentences have a strong gravitational effect on future sentences. This Article's proposal leaves more room for the Commission to amend the Guidelines in numerical ways, such as the Commission did when it reduced the drug quantity table by two levels.[199]

Second, this Article's proposal more strongly adheres to the ability of the Guidelines to act as an anchor. Osler's proposal seeks to be more transformational, but by prohibiting the Commission from attaching numerical offense levels to aggravating and mitigating factors,[200] Osler's

---

195. Mark Osler, *The Promise of Trailing-Edge Sentencing Guidelines to Resolve the Conflict between Uniformity and Judicial Discretion*, 14 N.C. J. L. & Tech. 203, 207, 234 (2012).

196. *Id.* at 230, 243.

197. U.S. Sentencing Guidelines Manual Ch. 1(A)(1)(3) (U.S. Sentencing Comm'n 2015) (noting that the Commission started with an "empirical approach that used as a starting point data estimating pre-guidelines sentencing practice").

198. Osler, *supra* note 195, at 230.

199. U.S. Sentencing Comm'n, Amendments to the Sentencing Guidelines 29-36 (April 30, 2014), http://www.ussc.gov/sites/default/files/pdf/amendment-process/reader-friendly-amendments/20140430_RF_Amendments.pdf (displaying a two-level reduction from previous years in the base offense level assigned to various drug quantities).

200. Osler, *supra* note 195, at 229.

model would make it difficult for judges to compare aggravating and mitigating factors with each other because the computer system would show only the end sentence, not how a judge arrived at it. For instance, the current Guidelines dictate that a defendant's intent to promote terrorism should be weighed six times as heavily as physically restraining a victim,[201] and it is twice as mitigating when an offender is a "minimal" participant than when he is a "minor" participant.[202] This Article's proposal retains this useful information.

Finally, this Article's proposal expands on some of the limits of Osler's model by synchronizing the basic idea of computer-generated, data-driven sentencing with the actual data itself—something that was beyond the scope of Osler's initial article. It turns out that the data, while enormous, is not robust enough to apply Osler's model. As laid out in Part IV, below, the Commission's data files show that, as the number of compared factors increases, the sample sizes rapidly decrease. The dataset is too small for a meaningful analysis of sentencing patterns when the set is filtered by more than a few data points. Osler suggests that judges could input the specific offense characteristics associated with a crime into the computer.[203] But the data shows that this method would shrink the population sample beyond usefulness. For instance, the Guidelines category for basic forms of fraud (section 2B1.1) includes *nineteen* specific offense characteristics. The presence or absence of these offense characteristics is not random, but assuming for the sake of illustration that each characteristic is equally likely, only 1 in 524,288 offenders under section 2B1.1 would exactly match any other defendant[204] before general enhancements are even factored in. Only about 70,000 defendants are convicted each year for *all* Guidelines categories, much less for section 2B1.1. Of course, the presence of offense characteristics is not random, but even if including all those variables would sometimes produce a sufficiently large sample size, including more than a few variables risks overcomplicating the system enough to render it ineffective. A more complicated system would be less likely to be well-received and would increase the risk of error.

Another scholar recently published a proposal that bears some similarity to Osler's and to the proposal in this Article. Adi Leibovitch remarks that specialized courts and courts of varying geographic jurisdictions exhibit different sentencing patterns. She argues that those

---

201. U.S. SENTENCING GUIDELINES MANUAL §§ 3A1.3, 3A1.4 (U.S. SENTENCING COMM'N 2016).

202. *Id.* at § 3B1.2.

203. Osler, *supra* note 195, at 233 (including as one of the factors to be considered in the search for similarly situated offenders the fact that an offender robbed a bank).

204. Assuming that each characteristic is equally likely to be absent as it is likely to be present, the probability that two offenders would exactly match the distribution of nineteen different variables is $\frac{1}{2^{19}} = \frac{1}{524,288}$.

patterns arise because judges sentence offenders based on how blameworthy an offender is in relation to other offenders a judge sees.[205] Judges who sentence only relatively low-level offenders tend to sentence those offenders more harshly than do courts of general jurisdiction.[206] Leibovitch argues that courts should use computers to review the distributions of sentences from other courts and employ "statistical curving" to ensure that sentences correlate with those issued in other jurisdictions.[207]

Leibovitch's article, like Osler's, suggests that past sentences should effectively become the guidelines judges use to sentence offenders. This backward-looking proposal is distinct from a proposal that seeks to employ sentencing data at the appellate stage. By introducing data into the equation at the sentencing stage instead of only the appellate stage, Leibovitch's proposal runs the same risk as Osler's: that previous sentencing practices will anchor judges, causing defense attorneys and prosecutors to fight over that data. For that reason, Leibovitch's proposal would similarly risk freezing current sentencing in place. Although district court judges could look at sentencing data as part of the proposal in this Article—indeed, they can do that already—this Article's proposal makes review of sentencing data an official part of sentencing policy only at the appellate stage, so it avoids much of the risk of freezing and ensures that the Sentencing Commission can make policy changes. Although Leibovitch mentions that appellate courts could review sentencing information and press more weight against sentences that deviate significantly from mine-run sentences,[208] establishing the contours of what that appellate system would look like is beyond the scope of her article. This Article, in contrast, develops a methodology in Part IV and identifies that the abuse-of-discretion standard is significantly responsible for the increase in inter-judge disparity.

B. THE PROPOSAL IS SOUND POLICY

To date, nobody has proposed a reform that accomplishes what the proposal outlined in this Article would. Indeed, only Osler's and Leibovitch's come close to targeting inter-judge disparity, which was a central aim of the Sentencing Reform Act and continues to frame debates on sentencing today.[209] Following *Booker*, the Guidelines fall short of

---

205. Adi Leibovitch, *Punishing on a Curve*, 111 Nw. U. L. Rev. 1205, 1224–33 (2017).

206. *Id.* at 1207.

207. *Id.* at 1256, 1261.

208. *Id.*

209. Hofer, *supra* note 8, at 39. *But cf.* Judge Nancy Gertner, *How to Talk About Sentencing Policy—and Not Disparity*, 46 Loy. C. Chi. L.J. 313 (2014) (remarking that disparity is "far, *far* less important than issues of sentencing fairness, of proportionality, of what works to address crime" and that rampant sentencing disparity is a myth).

reducing inter-judge disparity, although they are not completely dissociated from these goals to the extent they anchor judges.

This proposal addresses inter-judge disparity without repeating the error the Act made when it drastically reduced judicial discretion. Before the Act, broad judicial discretion had been universal, rarely challenged, and explicitly upheld by the Supreme Court.[210] Congress attempted to reduce inter-judge disparity by significantly reducing discretion, but the problem was not in discretion alone; it was in the absence of meaningful appellate review. When Congress did increase appellate review, it did so only to enforce the mandatory nature of the Guidelines.[211]

As the studies explored above have now indirectly demonstrated, this lax appellate standard has proven ineffectual. *Booker* implemented an unusual form of appellate court review and then gutted its central aspects to afford district court judges extremely broad discretion.[212] Those appellate forms of review that remain are impotent. District courts can disregard Guidelines policies;[213] substantive reasonableness challenges are almost never successful;[214] and procedural reasonableness review is relatively unchallenging.[215]

As opposed to the proposals that seek a fundamental transformation of the Guidelines system, this proposal instead creates a system that supplements the Guidelines, makes judges more aware of the sentences issued by their peers without formally introducing the use of sentencing data at the initial stage of sentencing, and induces judges to provide greater justifications for their sentences. Judges already must "adequately explain the chosen sentence"[216] in a way that is at least somewhat proportional to the amount of departure from a guideline range.[217] But the Court's doctrine makes "adequate" mean hardly more than "minimal"[218] and does little to reduce inter-judge disparity. This reform increases the floor for what constitutes an "adequate" explanation for offenders who receive sentences that deviate from the norm.

---

210. *See supra* Part I.A.

211. 18 U.S.C. § 3742(e) (2006).

212. Hessick & Hessick, *supra* note 99, at 3, 14.

213. *See supra* notes 102–104 and accompanying text.

214. *See supra* note 107.

215. *See supra* notes 108–109 and accompanying text.

216. *See supra* notes 105–06 and accompanying text; *see also* 18 U.S.C. § 3553(c) (2012) (requiring sentencing judges to include a statement of reasons when sentencing an offender).

217. The holding in *Rita* requires sentencing judges to "set forth enough to satisfy the appellate court that he has considered the parties' arguments and has a reasoned basis for exercising his own legal decisionmaking authority." Rita v. United States, 551 U.S. 338, 356 (2007). The judge will "normally go further and explain why he has rejected" arguments to depart from the Guidelines, and "[w]here the judge imposes a sentence outside the Guidelines, the judge will explain why he has done so." *Id.* at 357. It is also "uncontroversial that a major departure should be supported by a more significant justification than a minor one" and that a judge must "adequately explain the chosen sentence." Gall v. United States, 552 U.S. 38, 50 (2007).

218. *See supra* Part I.B.

For example, in addition to requiring higher justifications for sentences in general, this proposal also requires higher justifications for inter-judge disparity that occurs within guideline ranges. In doing so, this proposal far exceeds anything the Court or Congress has required. Present doctrine assumes away any concern for this form of disparity; the Court has held that no explanation beyond relying on the Guidelines' internal reasoning will generally be warranted for these sentences.[219] Congress has also failed to recognize the existence of this form of disparity. The statute that covers appellate reversal for failure to provide a statement of reasons permits vacating a sentence only when the sentence lies outside the calculated guideline range.[220] Yet most judges sentence at the bottom of a calculated range,[221] causing disparity whenever a judge sentences elsewhere within the range. This form of disparity is particularly harmful because judges who sentence within a guideline range can rest easy knowing they will almost certainly be affirmed.[222]

This proposal gives greater effect to the Court's nominal requirement that judges adequately explain their sentences. *Booker* implied the standard of reasonableness review from the Sentencing Reform Act (after excising the provisions that made the Guidelines mandatory).[223] This proposal asks Congress to revamp that standard to put more pressure onto sentencing judges to adequately explain their sentences. As laid out below, increasing the burden on sentencing judges to adequately justify their sentences will have both ameliorative effects on inter-judge disparity and other incidental benefits that reduce general disparity.

1. *Effect on Judges*

The proposal in this Article would carry several benefits. Most notably, it would shift the conversation in sentencing by inducing judges to provide greater justifications for their sentences. This proposal permits district court judges to rebut the presumption of unreasonableness by creating a more robust explanation for why the defendant merits the sentence issued. The practice of including greater written justifications, even if unaccompanied by significant differences in sentencing behavior or reversal rates, would rectify at least some

---

219. *Rita*, 551 U.S. at 356–57.

220. 18 U.S.C. § 3742(f) (2012). The only exceptions are when the sentence is otherwise unlawful, is incorrectly implied, or where the offense has no applicable guideline range, limiting an appellate court to remand only if the sentence "is plainly unreasonable." *Id.* The presence of these exceptions has been criticized as violating *Booker*. United States v. Booker, 543 U.S. 220, 307 n.6 (2005) (Scalia, J., dissenting).

221. *See supra* notes 25–26 and accompanying text.

222. Gertner, *supra* note 209, at 319.

223. *Booker*, 543 U.S. at 261.

inter-judge disparity because it would create a written record that would establish that an offender is not similarly situated to other offenders. What's more, this proposal—including, as it does, a presumption of unreasonableness—would almost certainly increase the reversal rate. Judges who were at all risk averse and who doubted that they might be able to craft a sufficiently compelling justification for a sentence would likely moderate that sentence. And this proposal carries little risk that the gravitation of sentences will be so moderating that sentences become de facto uniform because this proposal largely targets harsher and more lenient judges.

A second and related major benefit of this proposal is the increased awareness judges would obtain of the sentencing practices of their peers. Many judges are largely unaware of sentences issued by other judges on the same court,[224] and rarely does information about the sentencing practices of other judges appear in the record. But under this proposal, outlier judges who do not already know that their sentencing practices deviate will quickly learn as much. And unlike Osler's and Leibovitch's proposals, which run the risk of placing too much focus on the sentences other judges issue, this proposal reaps the benefits of introducing sentencing data into the system without drawing too much focus away from the Guidelines. It does so by formally introducing sentencing data only at the appellate stage. Some sentencing judges may, of course, consult the sentencing data on their own—and judges are more likely to do so if they can decrease their chances of being reversed—but nothing suggests that consultation will shift the focus of sentencing from the Guidelines to sentencing data. Under current doctrine, judges will still be required to calculate guideline ranges, but nothing will require judges to consult the sentencing data.

The incidental effect this proposal would have on increasing judicial awareness of peer practices would also assist those judges with satisfying an obligation they already are required to meet but that is difficult to meet. When sentencing, judges must consider "the need to avoid unwarranted sentence disparities among defendants with similar records who have been found guilty of similar conduct."[225] In the light of the numerous studies that establish that inter-judge disparity has increased significantly over the last decade, it is clear this provision has minimal practical effect. One reason is that the sentences of similarly situated offenders are rarely in the record. The proposal in this Article, by increasing judicial awareness of other sentences, partly rectifies this problem.

Another benefit of this proposal is that it tackles inter-judge disparity indirectly. The indirectness of the proposal carries some

---

224. Kopf, *supra* note 2, at 50.
225. 18 U.S.C. § 3553(a)(6) (2012).

downsides: it bestows the benefit of enhanced appellate review only on those who receive the harshest and most lenient sentences even though inter-judge disparity also affects people who receive mine-run sentences. But the indirectness of this proposal is precisely why it should be effective. The proposal is geared toward aggregate, not individual, change. The Sentencing Reform Act employed a heavy-handed approach to try to directly eliminate inter-judge disparity. That approach created a host of other problems. But when judges provide greater justifications for their sentences, their actions will mitigate inter-judge disparity for *all* offenders. In targeting harsher and more lenient sentences, this proposal will disproportionately affect harsher and more lenient judges. These judges will either find themselves reversed with great frequency—which will convey important information to these judges—or will be deterred in a way that will cause them to give more attention to justifying their sentences. Judges who were initially unaware of—or underestimated—their harsher or more lenient tendencies will moderate.

For similar reasons, this proposal will also attack one area of inter-judge disparity scholars have so far neglected to address: reducing inter-judge disparity within guideline ranges. Although appellate courts are not permitted to presume that sentences issued outside guideline ranges are unreasonable, they are permitted to presume that sentences within guideline ranges are reasonable.[226] "[T]he presumption reflects the fact that . . . both the sentencing judge and the Sentencing Commission will have reached the same conclusion as to the proper sentence in the particular case. That double determination significantly increases the likelihood that the sentence is a reasonable one."[227] This presumption ordinarily makes it difficult to challenge inter-judge disparity that falls within a guideline range. No other scholar has yet proposed a solution that attacks within-range inter-judge disparity. This proposal does just that by removing the presumption of reasonableness for sentences that fall toward the tail ends of the distributions, regardless of whether those sentences fall within or outside of the guideline ranges.

Finally, this proposal would target the particular risks of inter-judge disparity associated with judges who have taken senior status. Because judges on senior status are able to tailor their dockets,[228] they are less likely to encounter a representative cross-section of cases. As Leibovitch has established, judges who sit on specialized courts tend to craft the severity of sentences in relation to the rest of their dockets, not in relation

---

226. Rita v. United States, 551 U.S. 338, 347 (2007).

227. *Id.*

228. 28 U.S.C. § 294(c) (2012) (requiring judges who have taken senior status to take on only as much work as they are "willing and able to undertake"). District courts can adopt more concrete regulations.

to all possible crimes.[229] Because judges on senior status can exclude certain kinds of cases from their dockets, the same effect extends to them. This proposal mitigates that effect.

2. *Minimal Effect on Anchoring*

The difficult task of an institution that crafts a sentencing regime that includes guidelines is discovering how to strike a balance between uniformity and individual treatment.[230] The mandatory Guidelines focused too much on narrow uniformity. The advisory Guidelines technically permit a sentencing judge to issue whatever sentence she wishes, subject only to the constraint that the judge consider the guidelines and that the judge not issue a substantively unreasonable sentence. But the evidence establishes that, in practice, the Guidelines have a similar anchoring effect on judges across the country.[231] This effect remains one of the strongest tools that prohibit inter-judge disparity from growing even worse, and this proposal does little to diminish that anchoring effect because it affords official use of sentencing data only at the appellate stage. By additionally limiting this system to individuals who have received a certificate of appealability,[232] this system will help ensure that the Guidelines maintain their current locus in federal sentencing policy. Instead of undermining the focus on the Guidelines, a risk entailed by other proposals that have suggested that judges reference sentencing data, this proposal chips away at the excesses of individual treatment that the advisory Guidelines produce.

3. *Effect on Non-Judicial Sources of Disparity*

This reform not only brings federal sentencing practice more in line with the central purpose of the Sentencing Reform Act; it also carries ancillary benefits. It is well established that judges are responsible for relatively little sentencing disparity compared to prosecutors.[233] But

---

229. Leibovitch, *supra* note 205.

230. Barkow, *supra* note 97, at 1619.

231. Barkow, *supra* note 97, at 1621.

232. *See supra* Part II.

233. WILLIAM J. STUNTZ, THE COLLAPSE OF AMERICAN CRIMINAL JUSTICE 6 (2011) (identifying the shift in power from judges and juries to prosecutors as one of the key reasons for current criminal justice problems); U.S. SENTENCING COMM'N, *supra* note 4, at 101–02 (finding that, pre-*Booker*, statutes and Guidelines accounted for 73% of sentencing disparity, with judges accounting only for 2.9% of sentencing disparity); Anderson et al., *supra* note 9, at 301 (detailing the significant impact regional prosecution policies through U.S. Attorney's Offices have on inter-district sentencing disparity); Bunin, *supra* note 19, at 81 (identifying prosecutors as the greatest source of disparity); Hofer, *supra* note 8, at 39 (identifying mandatory minimums, prosecutor decisions, and plea bargaining decisions as more significant sources of disparity); *cf.* STITH & CABRANES, *supra* note 4, at 130 (noting that prosecutors have gained greater relative control over sentencing than before the Sentencing Reform Act).

reducing inter-judge disparity has collateral effects on prosecutor-driven disparity.

Prosecutors have numerous avenues to create disparity. In determining what charges to bring and whether to stipulate certain facts, prosecutors have a comparatively larger role in framing an offender's record in relation to the actual offense. Prosecutor stipulations have increased since *Booker*, as have prosecutors' demands that offenders who enter pleas waive their rights to appeal.[234] And prosecutors pursue mandatory minimums at disparate rates. None of the eligible defendants in some districts were charged with mandatory minimums, but prosecutors in other districts charged three-quarters of eligible defendants with mandatory minimums.[235] Prosecutors also control whether motions for sentencing enhancements or reductions are filed—they have the sole authority to file § 851 enhancement motions of information, which double an offender's mandatory minimum sentence—and the vast majority of departures for substantial assistance are prosecutor-driven.[236]

The sources of prosecutor-driven disparity are voluminous, and a tailored reform like the one presented in this Article can address only some of them. Although judges are often randomly assigned to cases, prosecutors often are not,[237] which means prosecutors can tailor their conduct based on the judge who is assigned a case. It is this source of disparity that the reform presented in this Article can mitigate. For instance, prosecutors can tailor charges to a judge because they can often file a superseding indictment after a judge has been assigned.[238] And because prosecutors are repeat players, they have the incentive to tailor their conduct to help ensure that future interactions are beneficial.[239] But when a judge's sentencing decisions are constrained not by the force of guideline ranges but by the need to construct a compelling explanation justifying a significant deviance from the median sentence, prosecutors will have proportionally less influence on these judges.

This proposal also can rectify some racial disparity in sentencing. Although researchers disagree on the existence or extent of racial disparities in sentencing and whether those disparities are caused by

---

234. Ulmer & Light, *supra* note 11, at 338.

235. Yang, *supra* note 112, at 1323–24.

236. Although a judge can grant a departure for substantial assistance without a motion of the government, judges rarely do. Section 5K1.1 motions for substantial assistance from the government were recognized in 9482 cases last year. 2014 SOURCEBOOK, *supra* note 107, at tbl.30 n.1. But judges recognized substantial assistance without a government motion in just over 400 cases. 2014 SOURCEBOOK, *supra* note 107, at tbl.25, tbl.25A, tbl.25B. Prosecutors can also file for substantial assistance motions to reduce a sentence after an offender has been sentenced. FED. R. CRIM. P. 35(B).

237. Hofer, *supra* note 8, at 41.

238. Yang, *supra* note 112, at 1278 n.43.

239. *See generally* ROBERT AXELROD, THE EVOLUTION OF COOPERATION (1984) (arguing that cooperation can arise from self-interested motives alone).

sentencing decisions instead of the decisions made at the investigation or adjudication stages,[240] the Sentencing Commission has released a report that finds a substantial increase in racial sentencing disparity after *Booker*.[241] And the Acting Chair of the Commission recently reported that black male defendants are twenty-five percent less likely than white male defendants to receive downward variances *sua sponte*.[242] To the extent some racial disparities arise from sentencing decisions instead of earlier decisions—such as charging decisions—this Article's proposal ameliorates some of that racial disparity by narrowing the overall gap and incidentally the racial gap.

4. *Further Positive Effects*

Tying the identities of judges to their sentencing habits has proved controversial. Opinions vary on whether judges need identity protection given their Article III protection,[243] but this reform benefits from not requiring the disclosure of judicial identities, which may be a dealbreaker for some actors.

Additionally, recognizing that inter-judge disparity is a relatively small part of the disparity equation, this proposal avoids aggravating those other sources of disparity. The same cannot be said of some other proposals. For instance, because some proposals increase prosecutorial power—like the ABA proposal—they risk making disparity worse.[244]

This proposal could also spark the development of a common law of sentencing in federal courts. Because the doctrine of sentencing review promulgated after *Booker* has been so deferential, precedent includes almost no guidance as to what "reasonable" means in the context of reasonableness review.[245] The precise contours of what appellate review would look like does not lend itself to a ready explanation ex ante. Appellate review would instead be fleshed out through litigation in fact-specific scenarios. To the extent more sentences were reversed, the proposal in this Article would "creat[e] the building blocks of a common law of federal sentencing" that would define what is meant by a

---

240. Scott, *supra* note 4, at 716–18.

241. Scott, *supra* note 4, at 717.

242. Judge William Pryor, Remarks at Scalia Law School 7–8 (May 17, 2017), https://www.ussc.gov/sites/default/files/pdf/news/speeches-and-articles/20170517_Pryor-Remarks.pdf.

243. *Compare, e.g.*, Scott, *supra* note 5, at 22 (arguing that the Commission's policy of refusing to publicly disclose judge identifiers is rooted in a desire for judges to shield themselves from criticism–"an astonishing expectation for public officials who enjoy life tenure"), *with* Mosi Secret, *Wide Sentencing Disparity Found Among U.S. Judges*, N.Y. TIMES (Mar. 5, 2012), http://www.nytimes.com/2012/03/06/nyregion/wide-sentencing-disparity-found-among-us-judges.html (identifying Justice Rehnquist as worrying "that collecting data on judges' sentencing practices 'could amount to an unwarranted and ill-considered effort to intimidate individual judges.'").

244. Yang, *supra* note 112, at 1279.

245. *See* Laura I. Appleman, *Toward a Common Law of Sentencing: Gall, Kimbrough, and the Search for Reasonableness*, 21 FED. SENT'G REP. 3, 3–4 (2008).

"reasonable" sentence.[246] By requiring sentencing judges to increase the level of explanation for sentences that fall outside the mine-run, and by reviewing those explanations, a more robust common law of sentencing might begin to emerge.

5. *The Limits of the Proposal*

Not every subpart of a system must be oriented toward the system's overall goals. This proposal is limited to ameliorating inter-judge disparity, and it does so in a way that avoids impinging on the gains that *Booker* made[247] while attacking the inter-judge disparity that *Booker* introduced.

Because this proposal uses the appeals of individual sentences as a vehicle to enact broader, aggregate change, the proposal incurs some inequities. The proposal bestows the main ameliorative effects on all, but some individuals receive the additional, direct benefit of having access to a new appellate tool, and still others bear the burden of having to defend their lenient sentences. This inequity is small because this reform should largely lead to a shift in the sentencing conversation—it will encourage judges to issue greater justifications—not a dramatic increase in overturned sentences. The remaining inequities are acceptable because they enable the pursuit of broadly shared benefits that might be unobtainable without either this proposal or a much more severe proposal like mandatory guidelines.

This proposal is also limited because of its reliance on data the Commission collects. This data largely comes from a Statement of Reasons form judges are supposed to submit to the Commission within 30 days of sentencing an offender.[248] When information is left off the form or is otherwise unavailable, the Commission pulls data from the Pre-Sentencing Reports probation officers compile to relay facts about the offense and the offender and to assist the judge at sentencing.[249] This protocol can introduce inconsistencies.

Former-Judge Nancy Gertner has declared the entire data set worthless for measuring inter-judge disparity. She argues the forms are

---

246. *Id.* at 3 ("A sentencing common law was something originally envisioned by the U.S. Sentencing Commission.").

247. Baron-Evans & Stith, *supra* note 24, at 1682–1703; *see also* Bunin, *supra* note 19, at 82–83 (identifying judicial discretion as preventing more disparity than it causes in part because the Guidelines fail to take into account some mitigating factors).

248. 28 U.S.C. § 994(w) (2012).

249. To the extent possible, information is first taken from the Statement of Reasons. In addition to a few other forms, the Presentence Investigation Report prepared by the probation officers under FED. R. CRIM. P. 32 is used to fill gaps in the data obtained by the Statement of Reasons. U.S. SENTENCING COMM'N, VARIABLE CODEBOOK FOR INDIVIDUAL DEFENDERS 2 (2015), https://www.ussc.gov/sites/default/files/pdf/researchandpublications/datafiles/Individual_Codebook_FY99_FY14.pdf.

too "simplistic" and were designed only to help the Commission "monitor Guideline enforcement, principally departures." She also contends that courtroom deputies or probation officers, not judges, routinely complete the forms.[250]

To be sure, the sentencing data betrays inaccuracies and shows that not all individuals who submit data do so in a uniform manner. For example, the vast majority of judges who find no applicable mandatory minimum on the statutory form appropriately report the "statutory minimum" on the Statement of Reasons as "0." But out of the more than 10,000 drug trafficking cases in 2014 where the judge reported no applicable mandatory minimum, the data for more than 800 offenders inexplicably listed a non-zero value for the variable that shows a "statutory minimum."[251] The data for one offender showed he was subject to a statutory minimum of life but included a contradictory variable that indicated that "no count of conviction carries a mandatory sentence." Despite this, another variable stated that a "mandatory minimum sentence [was] imposed." Additionally, the data for 229 offenders exhibits that no counts of conviction carried a mandatory sentence, but the data also exposes that the judges somehow departed from mandatory minimum sentences.[252]

Despite these errors, Gertner's analysis may be overstated. Her criticisms derive from her experience on a court that may be unrepresentative of national practice. Further, data inaccuracies are inherent in any reasonably substantial dataset, and the Commission takes measures to mitigate data inaccuracy. Biannually, all data is compared with the sentencing data file held by the Administrative Office of the U.S. Courts to ensure accuracy, and sentences that fall outside a guideline range are reviewed by hand.[253]

But even accepting all of Gertner's criticism, the data remains the best source for defining categories of similarly situated offenders, and the Commission uses this data for its congressionally authorized research

---

250. Gertner, *supra* note 155, at 47.

251. This Article controls for the following factors to obtain this figure: the offender was sentenced under section 2D1.1 of the U.S. Sentencing Guidelines Manual, the judge issued a non-zero value for the variable that shows whether the offender was convicted under a statute bearing a statutory minimum, and a contradictory separate variable was present that signaled that "no count of conviction carries a mandatory sentence."

252. This Article controls for the following factors to obtain this figure: the offender was sentenced under section 2D1.1 of the U.S. Sentencing Guidelines Manual, the judge issued a zero value for the variable that shows whether the offender was convicted under a statute bearing a statutory minimum, and a contradictory separate variable was present that signaled that "one or more counts of conviction *carry mandatory* [minimums] but the court determined it does not apply."

253. Lou Reedt et al., *Effective Use of Federal Sentencing Data*, U.S. SENT'G COMM'N 9 (2013), https://www.ussc.gov/sites/default/files/pdf/research-and-publications/datafiles/20131122-ACS-Presentation.pdf.

purposes.[254] The shortcomings are relatively rare in a scheme where more than 70,000 offenders are sentenced each year. Especially considering that high precision is not necessary for this proposal, data that is good enough for the Commission ought to be good enough for this proposal, even if it is less than ideal.

That leaves the practical limits of this appeal. As with any system that calls for new forms of information to be presented to judges, some logistical issues arise as to how data will be made accessible. But those issues are almost entirely administrative in nature, and computer accessibility places this specific reform within reach in a way not true in earlier decades. As explained in detail in Part IV, calculating whether a sentence should be presumed unreasonable is straightforward and would require no work from judges if administrative officials within the courts calculated sentences.

C. The Proposal Is Constitutional

The critical weakness in many of the proposals considered above is that there are good arguments that the proposals are unconstitutional and that none of the proposals sets out to defend itself on constitutional grounds.[255] The constitutional concerns are two-fold. First, the Court may have crafted a delicate balance: the Court promoted judicial discretion to avoid the Sixth Amendment issue but also instituted a gutted form of appellate review to promote some uniformity.[256] If, as some say, this balance is refined and delicate, then reducing judicial discretion to issue a non-Guidelines sentence might be constitutionally problematic.[257] The second point of concern is that a constraint, even if not tied to the Guidelines, might still render an indirect effect on the Guidelines that makes them de facto mandatory.

If a proposal reduces a judge's discretion and ability to avoid a non-Guidelines sentence (the first point of concern), that proposal must wade through the *Booker* progeny. The case friendliest toward altering the forms of appellate review is *Booker* itself. In drawing up a remedy, the *Booker* majority—which included Justice Breyer, an architect of the original Guidelines from his days on the First Circuit[258]—explicitly sought to divine "what 'Congress would have intended' in light of the Court's constitutional holding."[259] From Congress's decision to orient the Sentencing Reform Act toward promoting uniformity, the Court found

---

254. 28 U.S.C. § 995(a)(12), (14)–(16) (2012).

255. Baron-Evans & Stith, *supra* note 24, at 1716–41.

256. Hessick & Hessick, *supra* note 99, at 3–4, 14.

257. Hessick, *supra* note 24, at 1348; *see also* Gall v. United States, 552 U.S. 38, 47 (2007) (prohibiting a presumption of unreasonableness for sentences outside the Guidelines).

258. Linda Greenhouse, *Guidelines on Sentencing Are Flawed, Justice Says*, N.Y. Times (Nov. 21, 1998), http://www.nytimes.com/1998/11/21/us/guidelines-on-sentencing-are-flawed-justice-says.html.

259. United States v. Booker, 543 U.S. 220, 246 (2005).

that Congress had implied a reasonableness review.[260] The Court made its view clear when it declared, "Ours . . . is not the last word: The ball now lies in Congress's court."[261] These statements indicate that Congress has at least some flexibility to alter the Court's reasonableness standard or to abolish review for abuse of discretion, and the Court's later cases are essentially interpretations of the "reasonableness review" it found Congress had created by statute.

But the majority of movement has occurred on the bench, not in Congress. In *Gall*, the Court prohibited courts from adopting "rigid" mathematical rules that required sentencing judges to justify their sentences in proportion to the degree of departure. Doing so, according to the Court, would come "too close to creating an impermissible presumption of unreasonableness for sentences outside the [g]uideline range."[262] The Court also prohibited a heightened standard of review for sentences lying outside the guideline range.[263] But unlike in *Booker*, the Court dropped its deferential language, making it unclear how much of the holding was constitutionally necessary and how much was just a common law exposition on the reasonableness review the Court created in *Booker*. In *Rita*, the Court held that courts of appeals could presume that sentences were reasonable if the sentences fell within guidelines ranges.[264] Because the presumption was non-binding, the Court rejected the Petitioner's argument that "plac[ing] an additional burden on the district court to justify non-Guidelines sentences" is impermissible.[265]

The statement in *Booker* that "[t]he ball now lies in Congress's court" strongly suggests that all holdings about reasonableness review that follow *Booker* are not required by the Constitution. At least one Justice has expressly adopted the understanding that reasonableness review amounts to nothing more than common law reasoning. Justice Thomas has repeatedly refused to join the rest of the Court's more recent pronouncements about reasonableness review on the basis that "the Court's decisions in this area are necessarily grounded in policy considerations rather than law" because the *Booker* remedy "has no basis in the statute."[266]

Even if the Court's doctrinal pronouncements are tied to the Constitution, the proposal in this Article does not impinge that doctrine. The doctrine permits presumptions if the presumptions do not bind judges to the Guidelines (*Rita*) but not if a sufficient *risk* exists that the

260. *Id.* at 260, 263.

261. *Id.* at 265.

262. Gall v. United States, 552 U.S. 38, 47 (2007).

263. *Id.* at 50–51.

264. Rita v. United States, 551 U.S. 338, 347 (2007).

265. Brief for the Petitioner at 33, Rita v. United States, 551 U.S. 338 (2007) (No. 06-5754).

266. Kimbrough v. United States, 552 U.S. 85, 114 (2007) (Thomas, J., dissenting); *accord, e.g.*, Pepper v. United States, 562 U.S. 476, 518 (2011) (Thomas, J., dissenting).

presumptions will (*Gall*). More importantly, the Court seems to care tremendously about whether an appellate standard creates presumptions that depend on whether the sentence is inside or outside the guideline range. In *Rita*, the presumption applied only because the sentence fell within a guideline range, and in *Gall*, an appellate court applied a rigid proportionality justification and a heightened form of review because it determined that the sentence was outside the guideline range. The Court refused to permit that holding.

The proposal in this Article does not impinge on this doctrine because it does not tie itself to the Guidelines. Under this proposal, judges would not consider whether a sentence falls inside or outside a guideline range. And given the common, but not universal, practice of judges to sentence at the bottom of guideline ranges, this proposal would implicate sentences both inside and outside the guideline ranges. Although this Article uses data derived from sentences under the Guidelines, it does so only to define categories of similarly situated offenders, not to directly demarcate the boundaries between unreasonable and reasonable sentences, as occurred in *Gall*.

This proposal also overcomes the second constitutional concern: whether the reform indirectly makes the Guidelines de facto mandatory. Admittedly, this proposal draws some parallels to the *de novo* form of review the Court excised in *Booker,* but those parallels are shallow. Before it was excised, § 3742(e) explicitly directed an appellate court to review whether a sentence departed from a guideline range in a way not authorized by the Guidelines.[267] But for this proposal, whether a sentence is inside or outside a guideline range is immaterial. And although any form of stricter appellate review constrains judges somewhat, Congress is constitutionally capable of increasing constraints on judicial discretion. Congress, could, for example leave judges with no room for discretion by tying specific, determinate sentences to every possible offense.

Under this Article's proposal, the effective range in which judges could sentence would narrow, but the proposal would not introduce the core Sixth Amendment problem underlying *Booker*, which was that courts were required to issue certain sentences based on facts not found by the jury. Indeed, the whole scheme of reasonableness review (especially substantive reasonableness) expressly permits decreasing judicial discretion based on judicial fact-findings. When a court of appeals finds that a district judge's sentence was substantively unreasonable, the court holds that the facts the judge found bounded the possible sentences the judge could issue. Even if the Court has carefully crafted within reasonableness review a balance between discretion and

---

267. 18 U.S.C. § 3742(e)(3)(B) (2012).

uniformity, *Booker* states that Congress could supplant reasonableness review altogether. The Court's holding in *Rita*—that a presumption of reasonableness is not unconstitutional in part because it is not binding—is instructive here. While this Article's proposal constrains judges, it does so only to the extent it requires greater attention to explanations of sentences. It does not bind judges to the guideline ranges.[268]

The strongest argument against this proposal is that any reduction in the range of sentences a judge can impose is unconstitutional—that is, a court can review sentences only for procedural reasonableness. While some scholars support this position,[269] the Court rejected this argument 7-2 when Justice Scalia (joined by Justice Thomas) asserted it,[270] and the position has failed to gain traction since. What's more, the proposal in this Article arguably functions not as a substantive constraint, but as a valid procedural constraint. Some lower courts might construe this constraint as substantive,[271] but the Court's holding in *Gall* explicitly describes as "procedural error" a judge's "failing to adequately explain the chosen sentence."[272] Boiled down, this Article's proposal merely asks Congress to require more explanation for outlier sentences—that is, this proposal seeks to alter the goal posts for what is "adequate."

IV. METHODOLOGY FOR DEFINING CATEGORIES OF SIMILARLY SITUATED OFFENDERS

This part establishes a methodology the Commission could use to define categories of similarly situated offenders, and by wading through the variables the Commission uses to code the data, it shows that enough of the right kind of data exists to meaningfully create these categories. This proposal operates as a new appellate structure outside the Guidelines, but it depends on data drawn from a Guidelines-centered system. Fortunately, the Commission's trove of sentencing data provides a ready opportunity to create these categories. Within thirty days of sentencing an offender, judges must provide detailed sentencing data to

---

268. Other scholars have argued that the Court has created a balance between discretion and uniformity and that a proposal is unconstitutional when it "shift[s] the delicate balance the Court has struck after *Booker* away from district court discretion towards the Guidelines." Hessick, *supra* note 24, at 1343, 1348. Hessick identifies in the jurisprudence a uniformity ideal. But this uniformity ideal is derived from the Sentencing Reform Act, not the Constitution. By adopting the proposal from this Article, Congress would slightly shift its ideals away from those expressed in the Sentencing Reform Act.

269. *E.g.*, Sarah M. R. Cravens, *Judging Discretion: Contexts for Understanding the Role of Judgment*, 64 U. MIAMI L. REV. 947, 966 & n.73 (2010).

270. Rita v. United States, 551 U.S. 338, 370 (2007) (Scalia, J., concurring in part and dissenting in part) (arguing that no substantive component to reasonableness review can exist and that "district courts must be able . . . to sentence to the maximum of the *statutory* range").

271. *See* United States v. Borho, 485 F.3d 904, 911 (6th Cir. 2007).

272. Gall v. United States, 552 U.S. 38, 51 (2007).

the Commission.[273] Data files of individual and corporate sentences since 2002 are publicly available.[274] The compliance rate is also good. Courts gave the Commission every Judgment and Conviction order and 98.8% of the Statement of Reason forms in 2014.[275] Courts are required to use these forms to record the reasons they issued a specific sentence,[276] and these forms are the primary documents the Commission uses to code sentencing data. Although courts were somewhat less compliant with providing Pre-Sentence Reports (handing over about ninety-two percent), the Commission collects those reports largely to fill data gaps in the Statement of Reason forms.

Even where researchers have discerned judges' identities, the studies often suffer from an inability to control for case characteristics that affect sentencing.[277] This Article's method benefits from not needing judge identifiers, and it allows for control of case characteristics because it uses the Commission's data, which includes thousands of data points for each individual offender. Among other things, the data includes the Guidelines code representing the statute of conviction (for example, section 2D1.1 for drug trafficking), the presence or absence of the numerous factors that go into calculating a guideline range, the circuit and courthouse of the sentencing proceedings, and 264 possible reasons a district court judge can give for departing from a guideline range.[278]

Although the data set is based on Guidelines that are themselves an inexact tool for classifying like offenses, no better data set exists, and the size of this set allows one to tailor the variable choices to diminish the internal biases in the Guidelines. For instance, the Guidelines arguably place too much emphasis on drug quantity, skewing the comparison between drug and non-drug cases. But the data points allow one to avoid that bias by isolating drug offenders from other offenders.

A. LIMITING THE PROPOSAL TO SPECIFIC OFFENSE CATEGORIES

Measuring inter-judge disparity is already difficult; it is made practically impossible if researchers take steps that decimate population sample sizes. The more variable qualifiers used to define a category of similarly situated offenders, the smaller the sample size becomes. The proposal accordingly should be used only for appeals of offenses that are common enough to draw a decent sample size: notably drug, firearm, and

---

273. 28 U.S.C. § 994(w) (2012).

274. *Commission Datafiles*, U.S. SENT'G COMM'N (2017), https://www.ussc.gov/research-and-publications/commission-datafiles.

275. 2014 SOURCEBOOK, *supra* note 107, at tbl.1.

276. 28 U.S.C. § 994(w)(1)(B) (2012).

277. *See supra* note 125.

278. U.S. SENTENCING COMM'N, *supra* note 249, at A-11–A-16; U.S. SENTENCING COMM'N, VARIABLE CODEBOOK (2014), https://www.ussc.gov/sites/default/files/pdf/research-and-publications/datafiles/Organizational_Codebook_FY00_FY14.pdf.

white-collar cases, which comprise about fifty-six percent of the federal criminal docket.[279] Although immigration offenses make up a large part of the federal docket (twenty-nine percent in fiscal year 2015[280]), those offenses are excluded because, despite their number, the offenses create comparatively little opportunity for inter-judge disparity. Congress's authorization of nationwide fast-track sentencing—significant downward departures given to some immigration offenders who enter guilty pleas—has significantly decreased judicial discretion.[281] All remaining offense categories constitute only about fifteen percent of the federal docket. Not enough people commit these crimes to create sufficiently large population samples against which an individual's offense can be measured. That the proposal in this Article is inapplicable to certain offenses does not dampen the promise of the proposal. This proposal does not seek an all-or-nothing revolution to make all sentences just; it instead adopts the ideal that sentencing some defendants in a more just manner is better than nothing. This proposal, moreover, works in part by making judges more aware of sentencing patterns within their districts. To that extent, this reform should have positive effects on the entire criminal system.

B. Sample Size

Certain offense categories are excluded because of sample size issues, but even offenses that are included must not be considered if no sufficient sample size of like offenders exists. The Commission can tailor the sample size to specific offenses or even to specific districts. For instance, the Commission might allow a smaller sample size in districts that have historically suffered worse inter-judge disparity. But a sample size of twenty-five similarly situated offenders should ordinarily be sufficient. This sample size ensures that some measure of sentencing pattern can be drawn in a district. A larger sample size would risk eliminating many smaller districts from consideration. Even this sample size eliminates some districts—although the opportunity for inter-judge disparity in those districts is comparatively smaller.

The sample should be taken from the previous three years of offenders for which data is available. This compromises between a need

---

279. U.S. Sentencing Comm'n, 2015 Sourcebook of Federal Sentencing Statistics fig.A (2016) https://www.ussc.gov/research/sourcebook-2015.

280. *Id.*

281. PROTECT Act, Pub. L. No. 108-21 § 401(m)(2)(B), 117 Stat. 650 (2003) (The Commission shall promulgate "a policy statement authorizing a downward departure of not more than 4 levels if the Government files a motion for such departure pursuant to an early disposition program authorized by the Attorney General and the United States Attorney."); Memorandum from James M. Cole, Deputy Attorney Gen., to all United States Attorneys 2 (Jan. 31, 2012), https://www.justice.gov/sites/default/files/dag/legacy/2012/01/31/fast-track-program.pdf (permitting fast-track sentencing in all districts).

to reach a sufficient sample size and the recognition both that sentencing patterns are dynamically tied to changes in judges, Guidelines, and statutory law and that permitting too much of a backward-bias makes it harder to induce future sentencing changes. Additionally, this compromise recognizes that the choice of year introduces complexities; Guidelines amendments ordinarily take effect on November 1 instead of at the beginning of a regular calendar year.[282] Where major amendments have been added to the Guidelines, the Commission can make exceptions or create a formula by which offenders in previous years can be compared to current offenders.[283]

C. CHOICE OF FACTORS

The strength of this proposal lies not only in the more-abstract call for more stringent appellate review, but also in the ability to view the data and establish that the data could be used to define categories of similarly situated offenders. For that purpose, this Subpart outlines suggested factors the Commission could use to craft those categories. But the proposal does not depend on the acceptance or rejection of each of these suggested factors. This Subpart suggests each factor largely for prudential reasons. This Article defines categories of similarly situated offenders as those offenders in the past three available years who have the following identical characteristics:[284]

---

282. A possible way around this problem would be to include the variables that detail the sentencing date, but even that solution is imperfect. The sentencing date does not necessarily detail the offense date, and defense litigation often focuses on which Guidelines Manual applies. Traveling down this rabbit hole risks drowning the usefulness of this proposal in complexity. This proposal strikes a balance between pragmatic simplicity and the precision that comes from considering more variables.

283. For example, after the changes to the drug quantity tables took place on November 1, 2014, many drug offenders became eligible for adjusted sentences based on guideline ranges two levels below what the same offenders previously received. In comparing current offenders to earlier offenders, the Commission could keep the analysis the same—that is, modern offenders would be compared with those receiving the same guideline ranges pre-2014 but who had lower drug quantities. But if there is good evidence, for example, that judges across the board found the drug penalties too harsh and that the drug quantity table was reduced to reflect *current* sentencing patterns, not a policy change, then it will be appropriate to compare modern offenders with earlier offenders who had *higher* guideline ranges.

284. For a list of the Commission's variables that correlate with these traits, see *infra* Appendix.

- District of sentencing

- Guideline section dictating which category of offense conduct was used to sentence the offender[285]

- Applicable guideline range in months, reflecting all adjustments minus acceptance of responsibility

- Criminal history score

- Acceptance or nonacceptance of responsibility; and

- Application or nonapplication of a mandatory minimum.

Because this reform is premised on shifting the location of the conversation, stricter precision—which would require a greater number of variables—is unnecessary. Indeed, including more variables would complicate the calculations and decimate sample sizes.

1. Factors Included

Of the six factors stated above, five require little explanation. Their relevance is easy to see. For instance, the Guidelines offense category ensures that appellate courts compare offenders who have similar criminal conduct, and the variable for acceptance of responsibility ensures that offenders with similar culpabilities are compared. (Acceptance of responsibility is not included in the variable for the calculated guideline range.) The researchers in the study about the Southern District of California lumped offenders into the same category if they committed drug trafficking crimes, were subject to mandatory minimums, and had received the benefits of the safety valve (thus ensuring that they had the same criminal history).[286] This Article adopts a similar comparison but also considers acceptance of responsibility and narrows the population samples to offenders who have the same calculated guideline ranges.[287]

Limiting consideration of data to the specific district of sentencing requires greater explanation. This reform is oriented toward the narrow purpose of reducing inter-judge disparity, not transforming the entire sentencing system. The ideal variable regarding geography should therefore be tailored as much as practicable to include sentences issued by any judge who could ordinarily sentence the specific offender. It is

---

285. This variable is admittedly imperfect because the conduct of some offenders requires calculating multiple different Guidelines sections and choosing the highest calculation. Thus, the conduct of offenders with the same calculation might diverge more than might be immediately apparent.

286. Mason & Bjerk, *supra* note 125, at 191.

287. This proposal does lose some information used in the California Study. The existence of a safety valve does not merely clue a judge into the criminal history of an offender; it also signals to the judge whether the offense was violent and the offender's relative role in the offense, 18 U.S.C. § 3553(f)—factors this proposal considers only indirectly.

only with respect to those judges that an individual can experience inter-judge disparity because judges ordinarily cannot sentence offenders outside the district in which an offense was committed.[288]

Descriptive and practical reasons favor limiting geography by district and not further narrowing the scope to specific courthouses. First, offenders in some districts can be randomly assigned to any courthouse within a district.[289] Second, using a variable covering the entire district helps ensure that a sufficiently large population sample can be drawn. Finally, the location of a crime within a district is sometimes as arbitrary as the assigning of an individual offender to a specific judge. It makes some difference, but not much, that a police officer might find drugs during a routine traffic stop close to the ultimate delivery location in, say, Hartford, Connecticut, than merely in route in New Haven or Bridgeport. This criticism is, of course, applicable where somebody is apprehended in New York for an eventual delivery to Bridgeport, but interstate travel—because it requires intrastate travel—is necessarily less common.

One could increase the population sample sizes for categories of similarly situated offenders by considering, in the alternative, all offenders across all districts. But that consideration is better tailored toward reducing inter-*district* disparity, which does not necessarily reduce inter-judge disparity. For instance, the disparity between two districts can be zero even if judges within each district diverge from other judges on the same court. Because this Article seeks to reduce inter-judge disparity, limiting review district-by-district is more appropriate for this proposal.

Furthermore, some inter-district disparity is healthy to a sentencing regime, even if the Sentencing Reform Act has rejected that value, whereas no affirmative direct benefit to inter-judge disparity exists. For instance, Guidelines-identical firearms crimes do not have the same level of harm when committed in rural Wyoming than in Manhattan, and it is understandable that two judges might sentence those offenders differently. It is precisely because not all statutorily identical crimes are equal that the Guidelines prescribes ranges in the first place, yet most judges ignore the purpose of those ranges by almost always sentencing at the low end of the guideline range. Moreover, even if the harm of a crime is identical in different districts, the harm to different localities of

---

288. FED. R. CRIM. P. 5(C)(3)(D). One notable exception to this rule is the opportunity for judges who have taken senior status to sit by designation in other districts. 28 U.S.C. § 294(d) (2012). But while those judges may bring their sentencing practices and biases from their home districts, their sentences should be compared with other in-district sentences.

289. D. CONN. L. R. CRIM. P. 50 (2009) (requiring only that cases be assigned "in accordance with a general policy . . . in the interest of the effective administration of justice"). The District of Minnesota also has assigned cases "without regard to the division in which the case arose." D. MINN. ORDER FOR ASSIGNMENT OF CASES 7 (Dec. 1, 2008).

incarceration changes between districts. The incidental harms of incarceration are arguably worse in areas where the incarceration rate is already high, and federal law has underappreciated that phenomenon in its haste to make a federal case of many issues that used to fall within the sole province of the states.[290] Pursuing too vigorously an ideal of inter-district uniformity also removes one of the largest checks on the severity of the federal system and undermines state policy.[291]

2. Factors Excluded

Here, too, this Article's choice reflects a compromise between precision and practicality. The drug quantity table often dictates the same offense level for two offenders who have wildly varying drug quantities (for example, fifteen kilograms of cocaine versus forty-nine),[292] but judges may reasonably sentence such Guidelines-identical offenders differently. One could further narrow the categories of similarly situated offenders by controlling drug quantities more precisely, but this line-by-line comparison of every factor the Commission collects would do nothing more than to reveal that each crime is unique—a useless exercise when trying to determine which offenders are similar.

That comparison would ensure that an appropriate sample size never arose. So, too, would an approach that considered every enhancement or reduction. For example, the offense category for theft and embezzlement includes nineteen specific offense characteristics,[293] and requiring offenders to match each of those factors would rapidly diminish the population sample.

Instead, this proposal recognizes that the internal logic of the Guidelines asserts that offenders are similarly situated based on the calculated guideline ranges alone. Because this reform, at its core, accepts the Guidelines logic, Congress could more feasibly adopt this proposal because it departs less from current policy. The conversion of these variables into a common currency is admittedly a legal fiction. In reality, the Guidelines place inordinate weight on certain factors—notably drug quantity and loss amount.[294] But a lot of these

---

290. *See, e.g.*, Michael M. O'Hear, *National Uniformity/Local Uniformity: Reconsidering the Use of Departures to Reduce Federal-State Sentencing Disparities*, 87 Iowa L. Rev. 721, 723 (2002).

291. *Id.* at 753–65 (arguing that courts have too readily pursued an ideal of national uniformity at the expense of an ideal of local uniformity, which is equally valid under federal law).

292. U.S. Sentencing Guidelines Manual § 2D1.1(c) (U.S. Sentencing Comm'n 2015).

293. *Id.* at § 2B1.1(b).

294. For economic crimes, the Guidelines have the potential to place far more weight on the economic loss of a crime than for an offender's relative role in the offense. In *United States v. Adelson*, 441 F. Supp. 2d 506, 507 (S.D.N.Y. 2006), the court rejected the government's request for a Guidelines life sentence. Noting that the practically inevitable decline in stock prices that follows from the revelation of fraud can have a tremendously large loss impact when that decline happens across

biases are contained to specific sentencing categories (such as drug quantity). To get rid of these biases, this Article does not fully accept the assertion of the Guidelines that any two offenders with the same calculated range are similar. It instead accepts that assertion only when two offenders committed the same statutory offense. This Article further limits the sample with five other elements, which mitigates the internal biases of the Guidelines.

A precise comparison between offenders also is not needed for a proposal that reduces inter-judge disparity. By making certain sentences subject to unreasonableness presumptions, this proposal will induce judges to include greater justifications for their sentences and to become more familiar with the sentencing practices of other judges on the same court. By nature, determining who is a victim or beneficiary of inter-judge disparity is difficult because no baseline "correct" sentence exists from which to measure a departure. But the strength of this proposal is that it will ameliorate sentences in general by increasing awareness of district-wide sentencing practices.

The methodology for this proposal omits other variables in part because even the Guidelines contain no method for comparing those variables. One such area is the hundreds of reasons judges may choose for departing from a guideline range, none of which are quantified and

---

millions of shares, the court stated that the loss table "may lead to guidelines offense levels that are, quite literally, off the chart" and are "patently absurd." *Id.* at 509, 515.

The loss table is located at section 2B1.1 of the U.S. Sentencing Guidelines Manual, but in a sense of rigid display, the Guidelines apply the loss table to a host of rather diverse crimes, and it is far from obvious that a loss amount in one crime applies to an equivalent fiscal loss amount in another. For instance, the loss table comes into play for the "volume of commerce attributable" to a defendant broadcasting obscene material, section 2G3.2, or merely transporting it, section 2G3.1. But it also applies to insider trading *gains*, section 2B1.4, and the destruction of fish, other wildlife, and plants, section 2Q2.1. Furthermore, it is not even clear that loss amount is an appropriate measure to begin with. Embezzlement of one million dollars that destroys a small company and puts thirty people completely out of work may be more devastating than a loss ten times as high that is spread loosely among ten million passive investors. Also controversial is how a loss is calculated in the first place. Some courts have applied a market-capitalization measure of damages, looking at the average value decrease per share multiplied by the number of shares. *Cf.* United States v. Olis, 429 F.3d 540, 542 (5th Cir. 2005) (reporting that the district court had sentenced according to the loss in stock price to a shareholder, without regard to any other factor that might affect stock price). This method of loss calculation is overly simplistic; it does not take into account the price at which shares were purchased or the multitude of factors that affect share price. Two-thirds of the loss in *Olis* occurred either before or more than a week after the problems were revealed. *Id.* at 548. Because market capitalization can drive a sentence, some have called for a separate guideline for those offenses. *The Reform of Federal Sentencing for Economic Crimes*, AM. BAR ASS'N. CRIMINAL JUSTICE TASK FORCE 9 (Nov. 10, 2014), https://www.americanbar.org/content/dam/aba/publications/criminaljustice/economic_crimes.au thcheckdam.pdf. The Commission recently amended the Guidelines to help ameliorate this problem. Previously, the Guidelines required courts to presume that market capitalization loss provided a reasonable estimate of the actual loss. Now, the Guidelines allow courts to use "any method that is appropriate and practicable." U.S. SENT'G COMM'N, AMENDMENTS TO THE SENTENCING GUIDELINES 25, 30 (Apr. 30, 2015), https://www.ussc.gov/sites/default/files/pdf/amendment-process/official-text-amendments/20150430_Amendments_0.pdf.

many of which are inherently incomparable. One variable asks whether a sentence "[a]fford[s] adequate deterrence to criminal conduct."[295] That variable will mean two very different things for different offenders—especially if "deterrence" can refer to specific deterrence, not only general deterrence.

Finally, this Article groups together all offenders who received a mandatory minimum sentence and, separately, all who do not, regardless of the reason the court did not apply a mandatory minimum. This latter choice respects that the Commission considers offenders who were never subject to a mandatory minimum to be similarly situated to those who were simply granted a reprieve.[296]

3. The Complexity of Guideline Amendments

One noticeable omission from this Article's choice of factors is any consideration of recent Guidelines amendments. Major amendments can make comparing sentences of offenders over a course of years difficult. This proposal begins at a base level of generality and calls for the Commission to complicate the analysis only where necessary. Considering amendments at the general stage risks overcomplicating this reform. For one thing, fashioning a rule for all amendments would be overbroad. For another, amendments ordinarily take effect on November 1 instead of at the beginning of a calendar year. If amendments were taken into account, then additional variables (such as date of sentencing) would need to be considered, further complicating this proposal.

Some amendments can be ignored altogether if the effect of the amendments is minute, because offenders sentenced before and after the amendment will often still be similar. But for other amendments, the Commission should draft special policy statements that instruct judges on how those judges should consider specific amendments.

V. CONCRETE EXAMPLES

Previous proposals, notably Mark Osler's, have suggested inserting the sentencing data into a computer-based algorithm to determine what sentence should issue. But providing concrete examples was beyond the

---

295. U.S. SENTENCING COMM'N, *supra* note 249, at A-11.

296. This choice is a bit imperfect. The main reasons for departing from a mandatory minimum—substantial assistance and the safety valve—are not completely comparable. Congress has stated that imposing a lower sentence than the mandatory minimum is "general[ly] appropriate[]" when an offender has rendered substantial assistance. 28 U.S.C. § 994(n) (2012). In contrast, a court "shall impose" a non-mandatory minimum sentence if the offender is eligible for the safety valve. 18 U.S.C. § 3553(f). But *Booker* effectively nullified any distinction between "may" and "shall," diminishing the distinction between the two categories. The only real difference is that, with substantial assistance, the government sometimes suggests a departure amount, and the judge sometimes obliges because judges are supposed to consider "the government's evaluation of the assistance rendered." U.S. SENTENCING GUIDELINES MANUAL § 5K1.1(a)(1) (U.S. SENTENCING COMM'N 2015).

scope of those articles. The following examples demonstrate how the proposal in this Article would work in practice.

A. DISTRICT OF MASSACHUSETTS

The three graphs below represent sentencing patterns in the District of Massachusetts from 2011 to 2014. These charts represent, for three guideline ranges, offenders sentenced in that district who were sentenced under section 2D1.1 (drug trafficking), had criminal history scores of 1, received a three-level sentence reduction due to acceptance of responsibility, and did not receive a mandatory minimum sentence.

Figure 2: Guideline Range: 37-46 Months, D. Mass.
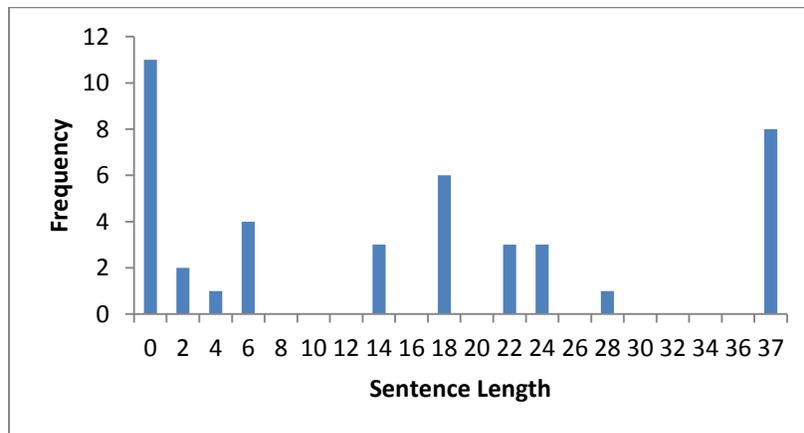
Figure 3: Guideline Range: 46-57 Months, D. Mass.
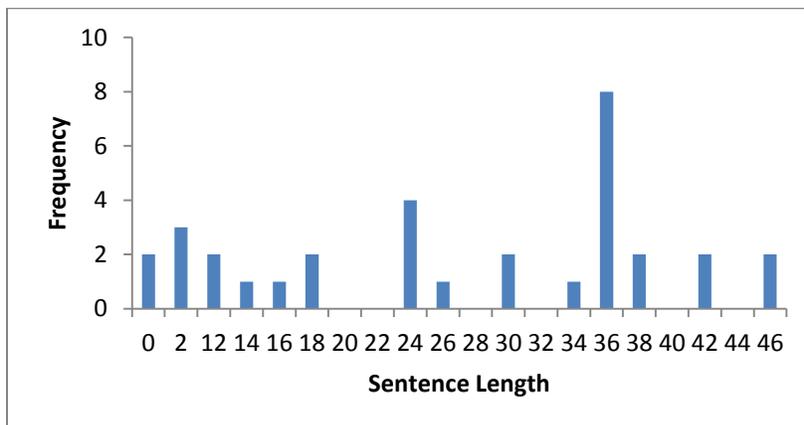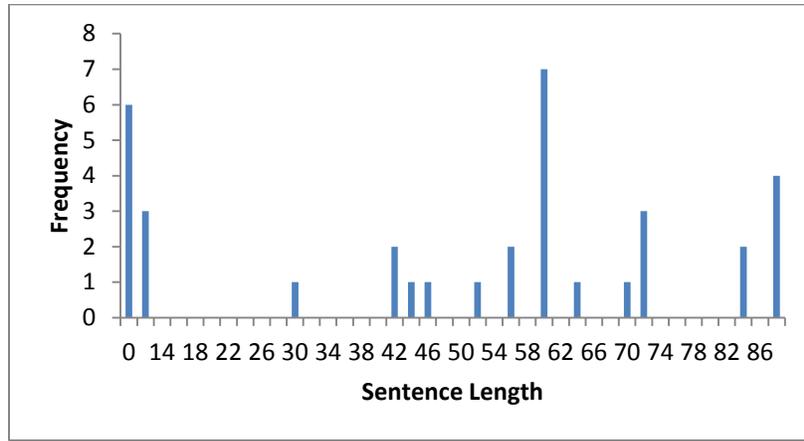
Figure 4: Guideline Range: 87-108 Months, D. Mass.

Consistent with earlier remarks on the topic, sentences tend to aggregate toward the bottom boundary of a guideline range. In each of these cases, not one offender received higher than the range minimum—indicating that there may be less inter-judge disparity within guideline ranges in this district. However, some judges frequently sentence elsewhere in a guideline range, meaning that an offender can receive a noticeably harsher intra-guideline range sentence based entirely on judicial identity. This further establishes the need to develop a system that can wrestle with disparity that occurs within a calculated guideline range.

Several spikes appear in each of these graphs. The spikes in both Figure 2 and Figure 4 appear at the guideline range minimum, at 0 (representing a non-prison sentence), and in between. Additionally, no discernible correlation exists between the mid-point spike and an identifiable trend in downward departures or variances. One could understand a mid-point spike if it correlated with a reduction in an offense level, but the absence of a correlation with the bottom of a guideline range suggests that some judges are issuing specific sentences without discernible input from the pertinent guideline ranges. For instance, the mid-point spike in Figure 3 is 24 months, but that number does not represent a range minimum or maximum from the sentencing table. Only the spike in Figure 2 correlates with an actual guideline range, but it represents a six-point reduction in the calculated guideline range. Guidelines policies, however, typically provide for a two- or three-point reduction.

The graphs also demonstrate how this reform would work. Figure 2 includes no sentences that would be presumed unreasonable because the lowest and highest sentence (0 and 37 months) reflect 19 of the 42

sentences, meaning no category of sentence falls wholly within the bottom or top fifteen percent of sentences. But in Figure 3, the court of appeals would presume that the sentences of 0, 1, 2,[297] 41, or 46 months were unreasonable. Sentencing judges could rebut these presumptions by providing sufficiently compelling justifications, which themselves would have ameliorating effects on the sentences given to other offenders. For Figure 4, the court of appeals would presume that any sentence of 0 months or 87 months was unreasonable.

B. WESTERN DISTRICT OF MISSOURI

These charts represent, for four guideline ranges, offenders sentenced in in the same three-year span in the Western District of Missouri who were sentenced under section 2D1.1 (drug trafficking), had criminal history scores of 1, received a three-level sentence reduction because of acceptance of responsibility, and did not receive a mandatory minimum sentence.

Figure 5: Guideline Minimum: 30 Months, W. D. Mo.

---
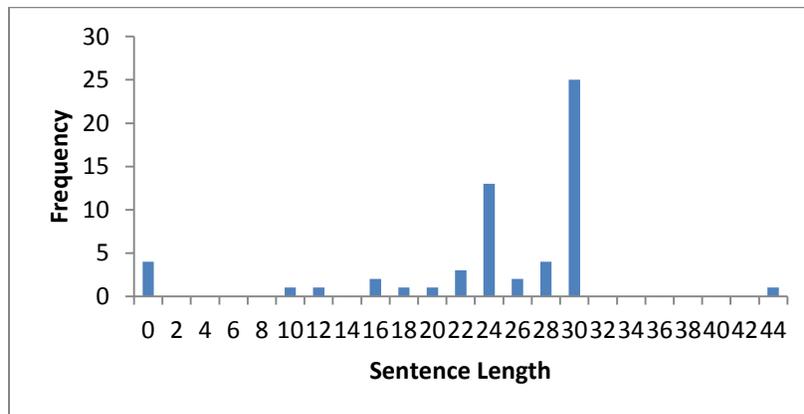
297. To reduce the complexity of the horizontal axis, this Article organizes the data in bins. The 2-month bin includes all sentences greater than 0 months but less than or equal to two months.
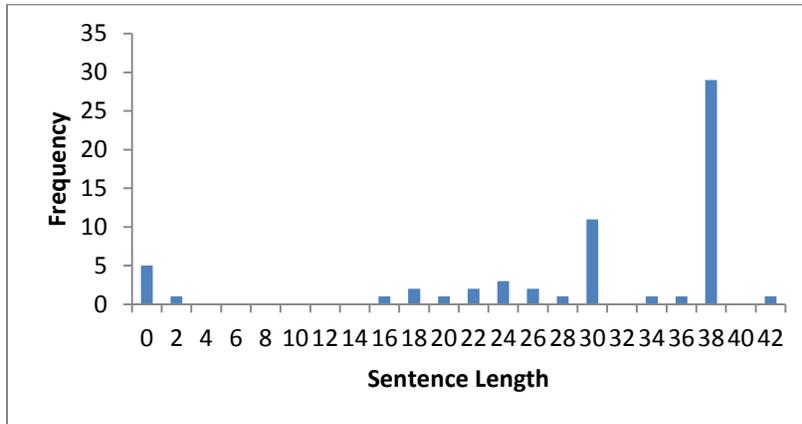
Figure 6: Guideline Minimum: 37 Months, W. D. Mo.
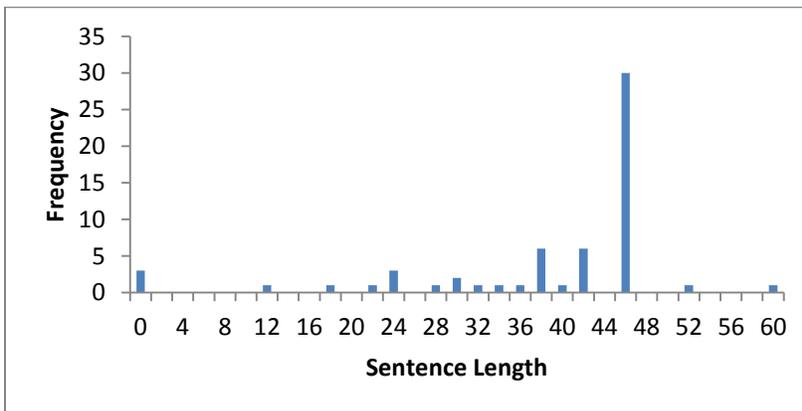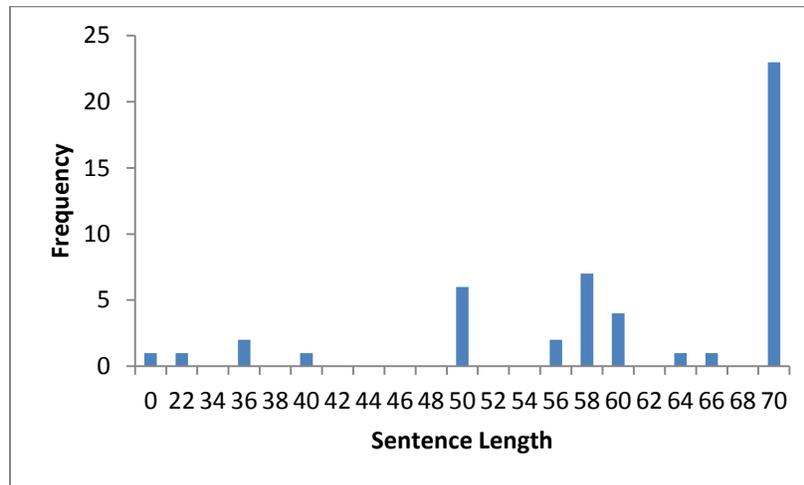

Figure 7: Guideline Minimum: 46 Months, W. D. Mo.

Figure 8:[298] **Guideline Minimum: 70 Months, W. D. Mo.**

In each of these four figures, one spike relates to the guideline range minimum, and another relates to the minimum for a two-level reduction from the applicable range minimum. Thus, unlike the data from the District of Massachusetts, the data indicates that some judges apply a de facto two-level reduction after calculating the guideline ranges, perhaps under a belief that the Guidelines are two levels too harsh. And no common variable is apparent among the offenders who received de facto two-level reductions. Additionally, in each of Figures 5, 6 and 7 at least one judge deviated from the norm that offenders should be sentenced at the bottom of a range, indicating possible inter-judge disparity within the guideline range.

It is easy to see how the proposal in this Article would play out in these situations. The court of appeals would presume that the sentences up to and including sixteen months in Figure 5 were unreasonable. The same is true for the forty-three month sentence (but not the thirty month sentences, as that sentencing bucket, although above the eighty-fifth percentile, also includes percentages below the eighty-fifth percentile). In Figure 6, the court of appeals would presume that everybody sentenced to eighteen months or less, as well as the individual sentenced to forty one months, received an unreasonable sentence. In Figure 7, offenders sentenced to twenty-four months or less would have the opportunity to appeal, as well as the offender sentenced to fifty-two

---

298. The sentences in the 58-month bin of Figure 8 (all sentences greater than 56 months but less than or equal to 58 months) were all equal to 57 months, which represents the new range minimum that would result if a judge were to apply a two-level reduction to the applicable range minimum.

months and the offender sentenced to sixty months. The offenders sentenced to forty-six months would not be able to appeal because that bucket falls both above and below the eighty-fifth percentile. In Figure 8, the sentences up to and including forty months would be presumed unreasonable.

CONCLUSION

Studies consistently show a significant increase in inter-judge disparity after *Booker*. While *Booker* removed many of the worst qualities of the Sentencing Reform Act, it also removed the aspect of the Act that most effectively reduced inter-judge disparity. Decisions of the Supreme Court after *Booker* have gutted any meaningful appellate review, which has left the federal sentencing system inept at curtailing inter-judge disparity. This problem is likely to grow worse as newer judges—less anchored to the Guidelines than those who sentenced under the mandatory Guidelines—join the bench.

Scholars have floated many proposals to change the Guidelines, but these proposals have been poorly tailored to reducing inter-judge disparity. Some of the proposals also make the same mistakes the Sentencing Reform Act did, and they have neglected to address constitutional questions raised by their proposals. Instead, reducing inter-judge disparity requires not a new set of Guidelines, but a heightened form of appellate review, and the Commission's data set provides a prime opportunity for creating a form of review unimaginable at the time Congress passed the Sentencing Reform Act.

APPENDIX

All data referenced in this Article is publicly available on the Commission's website.[299] Courts are required to provide the Commission with detailed data regarding a sentence within thirty days of that sentence.[300] The Commission then strips this data of identifying information (for example, judge's and offender's identities, case number) but still leaves up to thousands of different variables regarding the sentences. The Commission derives this data from a number of sources, including the Statement of Reason form, the Pre-Sentence Report (if necessary), the Judgment and Conviction Order, and in some instances the plea agreement.[301]

SPSS or SAS software can extract the information, although SAS cannot extract all the variables, including potentially relevant factors such as the offender's age, sentencing date, the loss amount from a crime, the date an offense of conviction began and ended, and (critically) whether a non-incarceration sentence was issued.[302]

After extracting the data, appropriate sentencing characteristics were isolated to find sentencing patterns:

- CIRCDIST: the district in which the offender was sentenced.

- GDLINEHI: the sentencing guideline, showing the section of the Guidelines under which the offender's highest range was calculated.

- XCRHISSR: the criminal history score of the offender.

- GLMIN and GLMAX: the calculated guideline range, including all chapter adjustments but excluding acceptance of responsibility.

- STATMIN: the existence and amount of any applicable mandatory minimum.

- ACCTRESP: the existence, and offense level reduction, attributable to acceptance of responsibility.

- MAND_: whether a mandatory minimum was applicable and was applied.

- SENTTOT0: The total monthly prison sentence, including probation as "0" values.

- SENTMON & SENTYR: the month and year of sentencing.

---

299. U.S. SENTENCING COMM'N, *supra* note 274.
300. 28 U.S.C. § 994(w) (2012).
301. U.S. SENTENCING COMM'N, *supra* note 249, at 2.
302. U.S. SENTENCING COMM'N, *supra* note 249, at 9, 29, 34, 41–42.

- REAS_: The reason(s) given for a sentence imposed outside the range.

- BOOKERCD: The generalized description (twelve possible values) for whether a sentence fell outside the range and for what generalized reason.[303]

After isolating the variables, the data was filtered into common factors. Similarly situated offenders were defined as those who were sentenced in the same district ("CIRCDIST"), under the same Guidelines section ("GDLINEHI"), with the same criminal history calculation ("XCRHISSR"), whose calculations fell within the same range ("GLMIN") and ("GLMAX"), who had the same status and applicable offense level reduction for acceptance of responsibility ("ACCTRESP"), and who either had a mandatory minimum applied or who did not ("MAND_").[304] The remaining variables, for example ("REAS_") were used to discern whether the disparities and spikes in the sentencing data were explainable by some other readily identifiable factor. For instance, after discovering consistent twin spikes in Figures 5-8, the REAS_ factors were reviewed, which come in almost 300 varieties,[305] to make sure no discernible pattern suggested that one or more factors contributed to disparately spiked sentences. Although a judge could list dozens or even hundreds of reasons for a departure, judges overwhelmingly issued either no reason at all or just one.

---

303. The underscore ("_") denotes a placeholder. For instance, an offender's data can include hundreds of possible reasons for a departure under the variable REAS_ (for example REAS1, REAS2, REAS3).

304. This variable was used in conjunction with the variable STATMIN. If an offender had a mandatory minimum applied, relying on both these variables helps determine whether the sentence was the mandatory minimum or whether a sentence above the mandatory minimum was applied.

305. U.S. SENTENCING COMM'N, *supra* note 249, at A-11–A-16.

***