Denver Law Review

Volume 85 | Issue 1 Article 4

January 2007

Empirical Questions and Evidence in Rita v. United States

Paul J. Hofer

Empirical Questions and Evidence in Rita v. United States

EMPIRICAL QUESTIONS AND EVIDENCE IN *RITA V. UNITED STATES*

PAUL J. HOFER†

INTRODUCTION

"Everyone is entitled to his own opinion, but not his own facts" according to a popular saying, variously attributed to Daniel Patrick Moynihan, James Schlessinger, or some unknown author. The line of cases beginning with *Apprendi v. New Jersey*[1] and continuing most recently with *Rita v. United States*[2] certainly confirms the first part of the quote. The various majority, concurring, and dissenting opinions in these cases express a wide range of theories regarding the roles of juries and judges in fact-finding and the constitutional requirements for facts used at sentencing. Perhaps the most extreme example of a split opinion, one can hope, was the Court's bifurcated ruling in *United States v. Booker*.[3] *Booker*'s remedial opinion, which made the Federal Sentencing Guidelines ("Guidelines") "advisory" rather than "mandatory,"[4] seems strangely irrelevant to the constitutional values underlying the merits opinion, which concerned the importance of the jury right in finding facts that increase punishment.[5] District and appellate courts are now engaged in defining what "advisory" Guidelines mean, and what procedural and substantive requirements still attend sentencing in the federal courts.

While opinions vary, facts have an independence and objectivity that the right empirical methods promise to reveal. Of course, consensus on the facts and what conclusions can safely be drawn from them is still often elusive, particularly when the facts at issue require statistical analysis and interpretation. Contrary to another popular saying, statistics are neither lies nor damn lies, but they can be used improperly and they can be too limited to reach certain conclusions. Fortunately for social and behavioral scientists working in law and public policy, statistics can often provide sound answers to important empirical questions that would otherwise be left to anecdote or opinion.

---

† Paul J. Hofer, B.A., University of Nebraska-Lincoln, 1980; J.D., University of Maryland School of Law, 1986; Ph.D., Johns Hopkins University, 1987. This work was undertaken while the author was Senior Research Associate at the United States Sentencing Commission. He currently conducts independent research and teaches at Johns Hopkins University. The views expressed in this article are those of the author and do not necessarily reflect any policies or positions of the United States Sentencing Commission.

1. 530 U.S. 466 (2000).
2. 127 S. Ct. 2456 (2007).
3. 543 U.S. 220, 229 (2005).
4. *Id.* at 245.
5. *Id.* at 243-45.

Statistical data have been presented to the Supreme Court since the introduction of the famous "Brandeis Brief" in 1908.[6] In *Rita*, the respondent Department of Justice and three *amici curiae*—the U.S. Sentencing Commission (USSC), the Federal Public and Community Defenders with the National Association of Federal Defenders (FPCD), and the New York Council of Defense Lawyers (NYCDL)—continued this Legal Realist tradition by providing the Court with data on an empirical question relevant to the case: Does an appellate presumption of reasonableness for sentences within the guideline range have an effect on the outcomes of appeals or on sentences imposed by the district courts?[7] Although the Court reached no conclusion on the question and suggested that its decision did not depend on any particular answer, the question is interesting in its own right. Knowing the effects of a presumption of reasonableness could help predict the consequences of the Supreme Court's decision to permit the courts of appeals to adopt such a presumption.

While not the subject of original data presented to the Court, the briefs in *Rita* also reviewed research relevant to other empirical issues. Most important among these were the questions of the Guidelines' effect on unwarranted disparity and whether the Guidelines have successfully helped to achieve the statutory purposes of sentencing. In addition, the decision itself makes a series of empirical claims and raises important research questions for the future. This article briefly reviews these other empirical questions after taking a closer look at the data presented to the Court on the effects of an appellate presumption of reasonableness.

I. THE NON-EFFECT OF THE FACTS ON THE LAW

United States v. Booker held that to avoid impinging the Sixth Amendment jury right, the Guidelines must be advisory, not mandatory.[8] This would appear to grant sentencing judges greater discretion, compared to before the decision, to sentence within or outside the guideline range. *Booker* also established a "reasonableness" standard of appellate review for sentencing decisions, which is emerging as largely equivalent to review for abuse of discretion and different from the *de novo* review that had been established by the PROTECT Act of 2003.[9] Prior to the

6. Muller v. Georgia, 208 U.S. 412 (1908). Justice Brandeis, who was then acting as a litigator, provided the Court with data on the health effects of long working hours on women. *See* Brief for Respondent, Muller v. Oregon, 208 U.S. 412 (1908) (No. 107), 1908 WL 27605.

7. *Rita*, 127 S. Ct. at 2465 (citing Brief for Petitioner at *28, Rita v. United States, 127 S. Ct. 2456 (2007) (No. 06-5754), 2006 WL 3740371; Brief for the United States at *34-39, Rita v. United States, 127 S. Ct. 2456 (2007) (No. 06-5754), 2007 WL 186288; Brief for the United States Sentencing Commission as Amicus Curiae Supporting Respondent at *15-16, Rita v. United States, 127 S. Ct. 2456 (2007) (No. 06-5754), 2007 WL 173622; Brief for the Federal Public and Community Defenders & the National Association of Federal Defenders as Amicus Curiae Supporting Petitioners at *12-15, Rita v. United States, 127 S. Ct. 2456 (2007) (No. 06-5754), 2006 WL 3760844; Brief for the New York Council of Defense Lawyers as Amicus Curiae Supporting Petitioner at *5-9, Rita v. United States, 127 S. Ct. 2456 (2007) (No. 06-5754), 2006 WL 3742254).

8. *Booker*, 543 U.S. at 245.

9. *Id.* at 261.

PROTECT Act, appellate courts had applied the abuse of discretion standard pursuant to the 1996 Supreme Court case *Koon v. United States*.[10]

Just how much discretion *Booker* actually returned to sentencing judges will depend on the substance and rigor of these new standards, and how they differ from those previously in effect.[11] It is natural to expect that the legal changes granting judges greater discretion would have an effect on actual judicial behavior, as *Booker* indeed did.[12] Conversely, one might expect that data concerning actual judicial behavior might have an effect on the Court as it gives definition to the new regime.

The petitioner in *Rita* and several of his *amici* assumed that appellate review that tended to encourage sentences within the guideline range would have the effect of decreasing the "advisory" quality of the Guidelines, potentially raising the same constitutional problems that had plagued the former "mandatory" Guidelines.[13] Standards of review and review procedures that increased the likelihood that courts of appeals would affirm within-range sentences would consequently increase the likelihood that sentencing judges would sentence within the guideline range.[14] Because that range will often have been increased by facts found by the judge, and not the jury, review that tended to increase sentences within the guideline range could raise Sixth Amendment problems.[15] One could thus assume that data showing that the presumption had the effect of increasing within-range sentences, and affirmance of those sentences on appeal, could be relevant to the constitutional analysis.

---

10. 518 U.S. 81, 91 (1996).

11. In *Booker*, the Court excised the previous statutory standard for departures outside the guideline range, which required judges to identify an "aggravating or mitigating circumstance of a kind, or to a degree, not adequately taken into consideration by the USSC in formulating the guidelines that should result in a sentence different from that described [by the guidelines]." 18 U.S.C.A. § 3553(b)(1) (2007). This standard remains, however, in the USSC's policy statement in the U.S. SENTENCING GUIDELINES MANUAL § 5K2.0 (2006), which sets out grounds for departure. After *Booker*, a new type of outside-the-range sentence has been recognized—a "variance"—which is not bound by this standard but only by the statutory provisions found at 18 U.S.C.A. § 3553(a) and whatever substantive standards emerge from reasonableness review.

12. *See generally* Paul J. Hofer, United States v. Booker *as a Natural Experiment*, 6 CRIMINOLOGY & PUB. POL'Y 433 (2007).

13. Several commentators, including this author, have questioned the substance and accuracy of these labels. *See, e.g.*, Kevin R. Reitz, *The Enforceability of Sentencing Guidelines*, 58 STAN. L. REV 155, 156 (2005) (describing the terms as "legal jargon" and "distorted terminology" holding "talismanic power" for some justices, but having little inherent legal effect); *see also* Paul J. Hofer, *Immediate and Long-Term Effects of* United States v. Booker*: More Discretion, More Disparity, or Better Reasoned Sentences?*, 38 ARIZ. ST. L.J. 425, 437-38 (2006) (arguing that the Guidelines met the usual definition for "presumptive" guidelines prior to *Booker* and remain "presumptive" after, with the major difference being the standard for departure or "variance," which will emerge only as reasonableness review is given substance).

14. Rita v. United States, 127 S. Ct. 2456, 2465 (2007) (citing Brief for Petitioner, *supra* note 7).

15. *Id.*

It turns out that, as a legal matter, this assumption was wrong. Justice Breyer's majority opinion in *Rita* makes clear that "the presumption, even if it increases the likelihood that the judge, not the jury, will find 'sentencing facts,' does not violate the Sixth Amendment."[16] The constitutional question "is only whether the law *forbids* a judge to increase a defendant's sentence *unless* the judge finds facts that the jury did not find (and the offender did not concede)."[17] The presumption does not *require* a within-guideline sentence, nor does it "*forbid* the sentencing judge from imposing a sentence higher than the Guidelines provide for the jury-determined facts standing alone."[18] The Guidelines are still legally advisory, even if the presumption increases the frequency that they are followed, perhaps even to rates similar to those found in some periods under the previous "mandatory" system.[19] "*As far as the law is concerned*, the judge could disregard the Guidelines" and apply a longer sentence.[20]

Justice Scalia and other dissenters have not been impressed with these legalistic niceties since they were first introduced in *Booker,* and they remained unimpressed in *Rita.*[21] When engaging in reasonableness review, they argue, the question is under what circumstances it is permissible to sentence outside the guideline range. Justice Alito argued in *Cunningham v. California,*[22] an earlier 2007 case involving the California sentencing guidelines, that if reasonableness is to have any substance, there must be *some* circumstances in which it would be unreasonable to sentence above the guideline range without finding facts in addition to those found by the jury or conceded by the defendant.[23] Those facts would then raise the constitutional problem identified in the *Booker* merits opinion. Justice Scalia in *Rita* argued in dissent that, even though such circumstances were not present in the instant case, any reasonableness review with substance will necessarily raise lingering Sixth Amendment issues and thus should not prevail.[24] He proposed a purely procedural form of review to avoid such problems.[25]

Justice Souter, in dissent, showed the greatest concern with what a presumption of reasonableness means for actual judicial behavior, and whether it "would tend to produce guidelines sentences almost as regularly as the mandatory Guidelines had done."[26] He ultimately could not

16. *Id.*
17. *Id.* at 2466.
18. *Id.*
19. *See id.* at 2467.
20. *Id.* at 2466 (emphasis added).
21. *See id.* at 2474-77 (Scalia, J., concurring); *see id.* at 2487-88 (Souter, J., dissenting).
22. 127 S. Ct. 856 (2007).
23. *Id.* at 880 (Alito, J., dissenting).
24. *Rita*, 127 S. Ct. at 2476-84 (Scalia, J., dissenting).
25. *Id.* at 2476, 2483.
26. *Id.* at 2487 (Souter, J., dissenting).

endorse the presumption because it would contribute "gravitational pull" to the Guidelines and move the system back toward being mandatory.[27]

Justice Stevens, in concurrence with the majority, accepted the presumption because he found it consistent with review for abuse of discretion.[28] He shared Justice Souter's view that the guidelines should not be treated as mandatory now as they were before *Booker*. He doubted, however, Justice Souter's empirical prediction that the presumption would affect judicial behavior by exerting "gravitational pull."[29] To help weaken any possible pull, Justice Stevens emphasized the deference owed to all individualized sentencing decisions, whether inside the guideline range or outside it.[30] He particularly encouraged sentencing judges to consider individual characteristics that the USSC has deemed "not ordinarily relevant," wryly re-christening them "not ordinarily considered" under the Guidelines.[31]

These concerns about gravitational effects notwithstanding, data about the effect of the presumption on actual judicial behavior ultimately had no impact on the decision. Justice Breyer and the majority concluded that, even if the petitioner and the *amici* were correct that the presumption encourages judges to sentence within the Guidelines, "we do not see how that fact could change the constitutional calculus."[32]

II. THE NON-EFFECT OF THE LAW ON THE FACTS

Legally, it didn't matter whether the presumption actually affects judicial behavior. But factually, who was correct? Does an appellate presumption of reasonableness for within-guideline sentences affect affirmance and reversal rates by the appellate courts, or use of the guideline range by sentencing judges? This empirical question was addressed in the briefs and the amount of data concerning it that was presented to the court was daunting. Moreover, the various parties made conflicting claims about what the data showed. Both because the question itself is interesting, and because it serves as an example of the difficulties involved in drawing sound conclusions from statistics, close examination of the data in the briefs is worthwhile.

A. Appellate Decision Making

Petitioner-Defendant Victor Rita and his *amici* pointed to some striking facts. Data gathered by the NYCDL showed that, of 1,152 ap-

27. *Id.*
28. *Id.* at 2471 (Stevens, J., concurring).
29. *Id.* at 2473-74.
30. *Id.*
31. *Id.* at 2473.
32. *Id.* at 2467 (majority opinion).

peals of within-guideline sentences in the post-*Booker* period,[33] only sixteen had been vacated by appeals courts and only *one* of these was because the sentence was found substantively unreasonable.[34] Fifteen were vacated for procedural unreasonableness, usually because the sentencing court did not adequately explain its reasons.[35] Rita argued that the circuits seemed to be treating within-guideline sentences as *per se* reasonable, which would encourage sentencing judges to seek the safe haven of the Guidelines and discourage sentences outside the range.[36] Citing Sentencing Commission-sponsored reports, the NYCDL argued that sentences recommended by the Guidelines "are often greater than necessary" to achieve the § 3553(a) statutory purposes.[37] Suspicion should be raised, it argued, by a "pattern in which less than .08% of appealed within-guideline sentences are reversed as substantively unreasonable."[38]

1. Unusual Circumstances of the Sole Within-range Reversal—*United States v. Lazenby*[39]

The question of whether a presumption of reasonableness contributes to the high affirmance rate of within-guideline sentences was complicated, however, by the fact that the only circuit which has reversed a within-guideline sentence on substantive grounds[40]—the Eighth—had itself adopted a presumption of reasonableness several months earlier.[41] The unusual circumstances of this unique case, *United States v. Lazenby*, merit study to explore what has been needed to rebut the presumption.

Two female co-defendants had been involved in a methamphetamine manufacturing and distribution conspiracy for which they purchased precursor chemicals.[42] Both had a romantic relationship with the same, more culpable co-defendant, who was already serving a lengthy prison term.[43] They were sentenced one month apart by different judges.[44] One received a within-guideline sentence of 87 months;[45] the

33. Brief for the New York Council of Defense Lawyers as Amicus Curiae Supporting Petitioners, *supra* note 7, at *3. The NYCDL data were obtained through a keyword search of Westlaw. The FPCD used a slightly different search. Each search returned a different number of cases, which accounts for the differences in numbers in each category. Although no bias is obvious in any of the search procedures, it is difficult to determine which returned the most representative and complete results.

34. *Id.* at *5.

35. *Id.*

36. Brief for Petitioner, *supra* note 7, at *34.

37. Brief for the New York Council of Defense Lawyers as Amicus Curiae Supporting Petitioners, *supra* note 7, at *5.

38. *Id.*

39. 439 F.3d 928 (8th Cir. 2006).

40. *Id.* at 933-34.

41. United States v. Lincoln, 413 F.3d 716, 717 (8th Cir. 2005)

42. *Lazenby*, 439 F.3d at 929-30.

43. *Id.*

44. *Id.* at 931.

45. *Id.*

other a sentence of twelve months and one day,[46] substantially below the applicable guideline range of 70-87 months.[47] The Eighth Circuit stated that "[t]he notable aspect of these appeals is the extreme disparity in the sentences imposed on two remarkably similar participants in the same criminal conspiracy. Moreover, a number of factors suggest that substantially greater leniency was afforded the *more culpable* defendant."[48]

The circuit reversed the below-range sentence as unreasonable, noting its prior ruling that "an extraordinary reduction must be supported by extraordinary circumstances," which it found lacking in this case.[49] The within-range sentence posed a more difficult question, since it was presumed reasonable and only "highly unusual circumstances will cause this court to conclude that the presumption has been rebutted."[50] The court found that the case was highly unusual based on several facts.[51]

First, even the prosecutor had stated at the sentencing hearing that the defendants were similar.[52] The difference in their guideline ranges was the result of different quantities of drugs stipulated in their plea agreements; according to the pre-sentence report, these different quantities did not reflect meaningful differences in their offense conduct.[53]

The defendant receiving the guideline sentence was the first of the co-defendants to plead guilty, and she agreed to cooperate with the government.[54] However, her testimony was not needed and no motion for a sentence reduction based on substantial assistance was made by the prosecutor, even though the circuit inferred that her cooperation played a role in the guilty pleas of others.[55] The circuit believed the sentencing judge had given "too much weight" to the prosecutor's statement that she was not authorized to seek a below-range sentence.[56] Most important, the circuit found that too little weight had been given to the "extreme disparity between the sentences imposed on two similarly situated con-

46. *Id.* at 930.
47. *Id.*
48. *Id.* at 932.
49. *Id.* (quoting United States v. Dalton, 404 F.3d 1029, 1033 (8th Cir. 2005)). The reasonableness of this standard is itself the subject of a pending appeal before the Supreme Court. *See* United States v. Gall, 446 F.3d 884, 889 (8th Cir. 2006), *cert. granted*, 127 S. Ct. 2933 (U.S. June 11, 2007) (No. 06-7949).
50. *Lazenby*, 439 F.3d at 933.
51. *Id.*
52. *Id.*
53. *Id.*
54. *Id.*
55. *Id.*; *see also* U.S. SENTENCING GUIDELINES MANUAL, *supra* note 11, § 5K1.1 (allowing for a departure from the Sentencing Guidelines based on a defendant's "substantial assistance in the investigation or prosecution of another person who has committed an offense . . . .").
56. *Lazenby*, 439 F.3d at 934.

spirators."[57] It thus reversed *both* the below-range and the within-range sentences and remanded for re-sentencing.[58]

This case shows that it is *possible* for within-range sentences to overcome the presumption of reasonableness and be reversed. But the highly unusual circumstances, with the two defendants treated in the same appeal, is likely to give little solace to defendants. And a single example cannot answer whether the presumption makes affirmance of within-guideline sentences *more likely* than they would be without a presumption. To answer this we need statistical comparisons of large numbers of cases.

2. Comparisons of Circuits With and Without the Presumption

The natural comparisons needed to assess the causal effects of a presumption of reasonableness are: (1) between circuits that do and do not have the presumption; and (2) within circuits before and after adopting the presumption. "Causal effects" is social and behavioral science lingo, and to some it may seem more appropriate for describing billiard balls hitting than the legal consequences of a court decision. Philosophers sometimes distinguish human reasons from physical causes, and object to analyzing human decision making in mechanistic terms. Moreover, some arguments raised by the parties and *amici* did not necessarily imply that the presumption *caused* judicial behavior, but instead cast the presumption more as a *symptom* of "institutional resistance" to *Booker*[59] or of a "culture of guideline compliance."[60] Nevertheless, an important underlying empirical claim in Rita's argument was that judicial behavior would be affected by the presumption, and this implies some type of causation. Thus, any change in behavior caused by adoption of the presumption should be reflected in affirmance and reversal rates.

Clean comparisons to test this hypothesis were hard to find, however. Because all but sixteen within-range sentences had been affirmed,[61] these numbers were too small for meaningful comparisons across circuits or time. Analysts had to rely on affirmance and reversal rates for outside-the-range sentences.[62] But for these, a factor more important than whether a presumption was applied is whether the sentence was above or below the guideline range. Potentially complicating matters even further, the identity of the party filing the appeal might also

57. *Id.*

58. *Id.*

59. Brief for the Federal Public and Community Defenders & the National Association of Federal Defenders as Amicus Curiae Supporting Petitioners, *supra* note 7, at *1.

60. Brief for Petitioner, *supra* note 7, at *29 (quoting Douglas A. Berman, *Tweaking* Booker: *Advisory Guidelines in the Federal System*, 43 HOUS. L. REV. 341, 349 (2006)).

61. *See* Brief for the New York Council of Defense Lawyers as Amicus Curiae Supporting Petitioners, *supra* note 7, at *5.

62. *See id.* at *5-6 (reviewing statistical analysis of affirmance and reversal rates for sentences outside the Sentencing Guidelines).

have an effect. Some below-range sentences are appealed by the defense, although none of these proved successful at achieving a reversal.[63]

Despite these difficulties, the NYCDL argued that "formal adoption of a presumption of reasonableness for within-guidelines sentences seems to affect the review of even those sentences not within the guidelines."[64] Its data showed that in circuits that had adopted the presumption, 47 of 51 below-range sentences appealed by the government were vacated as unreasonable, a reversal rate of 92 percent.[65] In circuits not adopting the presumption, 13 of 20 were vacated, a reversal rate of 65 percent.[66] The FPCD data were slightly different, but with the same overall pattern: 86.2 percent reversal in government-appealed below-range sentences in presumption circuits and 60 percent in non-presumption circuits.[67]

Lost in these comparisons, however, is whether the differences between these circuits were any less *prior* to adoption of the presumption by some of them. Without these comparisons, the data do not reveal whether the presumption itself was a contributing cause of the higher reversal rate of below-range sentences or whether the differences preexisted adoption of the presumption, as was argued by the Department of Justice.

3. Comparisons of Reversal Rates for Above- and Below-Range Sentences

Rita also argued that, contrary to the Supreme Court's holding in *Booker*, the courts of appeals were not applying the same reasonableness standard of review "across the board."[68] This was particularly so regarding sentences above the guideline range appealed by the defense compared to below the guideline range appealed by the government. The NYCDL framed these data in a larger argument that courts were not complying with the first statutory command at 18 U.S.C. § 3553(a), which directs courts to "impose a sentence sufficient, but not greater than necessary" to comply with the statutory purposes.[69] The sentencing literature calls this the "parsimony provision," and the NYCDL argued that

63. *Id.* at *6 n.6 (noting that 138 below-range sentences were appealed by defendants asking for a greater reduction).

64. *Id.* at *6.

65. *Id.* app. at 5a.

66. *Id.* app. at 6a.

67. Brief for the Federal Public and Community Defenders & the National Association of Federal Defenders as Amicus Curiae Supporting Petitioners, *supra* note 7, at *13.

68. Rita v. United States, 127 S. Ct. 2456, 2488 (2007) (Souter, J., dissenting) (citing United States v. Booker, 543 U.S. 220, 223 (2005)).

69. Brief for the New York Council of Defense Lawyers as Amicus Curiae Supporting Petitioner, *supra* note 7, at *4.

"empirical data revealed non-parsimonious patterns in the application of reasonableness review."[70]

The NYCDL data showed that 78.3 percent of below-range sentences appealed by the government were reversed, compared to 3.5 percent of above-range sentences appealed by the defense. The FPCD reported the same rates, and reviewed case law to argue that below-range sentences receive review akin to *de novo*, while review of above-range sentences is "perfunctory."[71]

The problem with the statistical comparisons is that not all outside-the-range sentences are equally likely to be appealed, and the merits of these appeals may not be equal. Different incentives bear on defense and prosecuting attorneys and their selection criteria for appeals are very different. Even though there are about seven and a half times *more* non-government sponsored below-range sentences than above-range sentences, Rita's data showed that government appeals of below-range sentences are far *less* frequent—just slightly more than half—than defendant appeals of above-range sentences.[72] In other words, a much smaller portion of below-range sentences are appealed by the government, and they are likely selected for merit as well as for policy considerations.[73]

The difficulty of making clean inferences from appellate affirmance and reversal rates led the USSC to exclude any comparisons of these data from its *amicus* brief.[74] It simply wasn't clear that the data addressed the empirical issue before the Court. The same was not true, however, of rates of within and outside-the-range sentences imposed by district judges.

B. District Court Decision Making

Petitioner Rita cited rates of within-, above-, and below-guideline sentences to argue that "[a] culture of guideline compliance has persisted after *Booker*."[75] Data on these rates are released quarterly by the USSC and have provided the key measure of *Booker*'s effects on sentencing.[76] Because of the large numbers of cases involved and the long time frame

---

70. *Id.* at *5-9.

71. Brief for the Federal Public and Community Defenders & the National Association of Federal Defenders as Amicus Curiae Supporting Petitioners, *supra* note 7, at *13, *16-19.

72. *See* Brief for the New York Council of Defense Lawyers as Amicus Curiae Supporting Petitioners, *supra* note *7, app. 5a-6a.

73. A more telling comparison would be the *likelihood of reversal* of all above and below-range sentences. However, data to make this comparison is not available in either the data gathered by the *Rita amici* or in the regular releases of the USSC.

74. *See* Brief for the United States Sentencing Commission as Amicus Curiae Suppporting Respondent, *supra* note *7.

75. Brief for Petitioner, *supra* note 7, at *29, (citing Berman, *Tweaking* Booker, *supra* note 60, at 349).

76. Brief for the United States Sentencing Commission as Amicus Curiae Supporting Respondent, *supra* note 7, at *18 (citing UNITED STATES SENTENCING COMMISSION PRELIMINARY QUARTERLY DATA REPORT 38 (Dec. 2006)).

for which data are available, these rates can be used to assess the causal effects of court decisions,[77] legislation,[78] and other historical factors. And they can be used to test the effects, if any, of the adoption of a presumption of reasonableness. Again, comparisons are needed across circuits that have or have not adopted the presumption and within a circuit before and after adoption of the presumption. The *amici* briefs of the FPCD and the USSC set out to directly test the effects of the presumption on sentencing judges by making these comparisons. Both briefs used USSC data, yet they drew very different conclusions.

1. Same Facts, Different Conclusions

The USSC brief first restated the petitioner's hypothesis—that a presumption of reasonableness renders the Guidelines effectively mandatory—but concluded that "the Commission's data refute that argument."[79] Indeed, the USSC noted, "[i]f petitioners' theory were correct, one would expect to see over time a significant and widening gap between the rate of below-guideline sentences in circuits with a presumption of reasonableness and the rate of such sentences in circuits without a presumption."[80] The brief then switched to a comparison of within-system[81] sentences, which are inversely related to below-range sentences, stating: "On the contrary, the rate at which sentencing judges impose a sentence [within the system] in circuits that apply the presumption is quite close to the rate in circuits that apply no presumption . . . . And the difference in those rates has remained virtually unchanged."[82]

The FPCD brief concluded the opposite:

---

77. Hofer, *supra* note 12. *See generally* Paul J. Hofer ET AL., *Departure Rates and Reasons After* Koon v. United States, 9 FED. SENT'G REP. 284 (1997).

78. Max Schanzenbach, *Have Federal Judges Changed Their Sentencing Practices? The Shaky Empirical Foundations of the Freeney Amendment*, 2 J. EMPIRICAL LEGAL STUD. 1 (2005) (exploring the empirical assumptions of the Freeney Amendment to the Prosecutorial Remedies and Tools Against the Exploitation of Children Today Act of 2003, Pub. L. 108-21 (codified at 117 Stat. 650 (2003))).

79. Brief for the United States Sentencing Commission as Amicus Curiae Supporting Respondent, *supra* note 7, at *16.

80. *Id.*

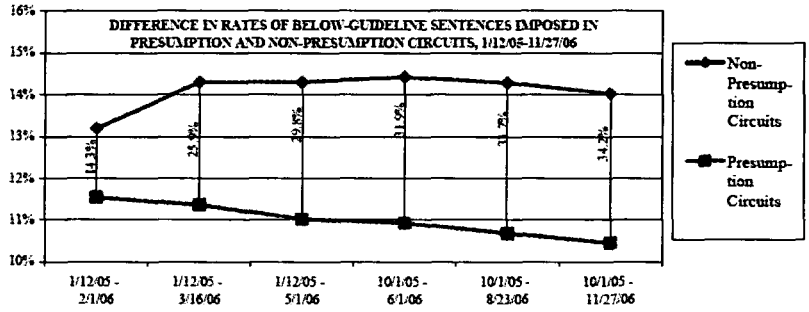
81. In post-*Booker* data releases and testimony, the USSC has described sentences within the guideline range combined with government-sponsored sentences below the guideline range as "conformance with the guidelines." *See Oversight Hearing on* United States v. Booker*: One Year Later – Chaos or Status Quo? Before Subcomm. on Crime, Terrorism, and Homeland Security of the H. Comm. on the Judiciary*, 109th Cong. 11 (2006) (statement of Hon. Ricardo H. Hinojosa, Chair, United States Sentencing Commission). The government sponsors below-range sentences for offenders who substantially assist in the prosecution of other persons, or participate in a fast-track program by pleading guilty quickly, or otherwise reach a plea agreement for a below-range sentence.

82. Brief for the United States Sentencing Commission as Amicus Curiae Supporting Respondent, *supra* note 7, at *16. The Department of Justice also argued that differences among the presumption and non-presumption circuits predate *Booker*. *See* Brief for the United States, *supra* note 7, at *38. Limitations in the Commission's ability to distinguish government-sponsored from other below range sentences make comparisons of rates prior to 2003 problematic, however.

> The gap between the rate of below-guideline sentences imposed by district courts in circuits that have returned the Guidelines to their presumptive status and that in circuits that have declined to do so has steadily widened. As of November 27, 2006, the rate in non-presumption circuits exceeded that in presumption by 34.2%.[83]

Both the USSC and the FPCD included graphs in appendices to their briefs. Figure 1 reproduces the FPCD graph displaying a widening in below-range rates between presumption and non-presumption circuits. Figure 2 shows the USSC's graph, with below-range rates in presumption and non-presumption circuits at the bottom and within-range rates at the top. No widening gap is apparent in either set of lines. It is embarrassing for social science when different empirical conclusions are drawn from the same data, and it is worth trying to figure out why. As usual with statistical evidence, the devil is in the details.

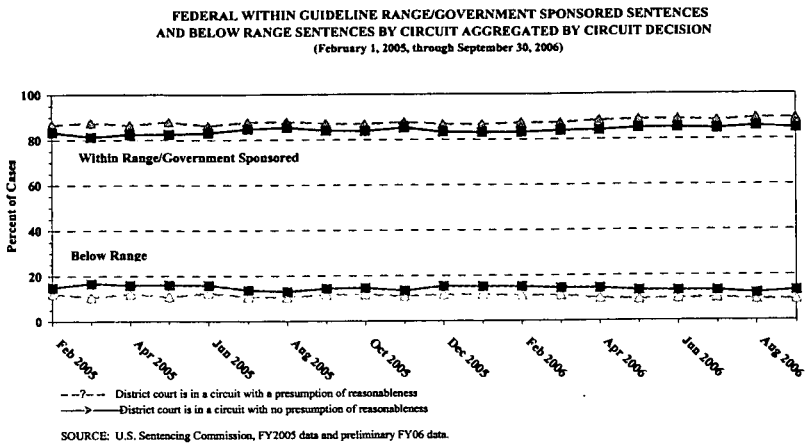
Figure 1: Data from FPCD Brief



The data used to prepare this chart was taken from the following reports: U.S. Sentencing Commission, Post-Booker Sentencing Update, published Feb. 14, 2006, Mar. 30, 2006, May 24, 2006, July 11, 2006; Quarterly Sentencing Update, published Aug. 30, 2006 and Dec. 6, 2006, available at http://www.ussc.gov/bf.htm. The lower line represents the number of below-guideline sentences imposed divided by the number of cases sentenced in circuits that had adopted a presumption of reasonableness before or during the respective period. The upper line represents the number of below-guideline sentences imposed divided by the number of cases sentenced in circuits that had not adopted a presumption of reasonableness before or during the respective time period. The percentage shown on the vertical line at each time period is the percentage by which the rate of below-guideline sentences for the circuits that had not adopted a presumption exceeded the rate of below-guideline sentences for the circuits that had adopted a presumption. The seven circuits to adopt a presumption did so in the following cases: *U.S. v. Mares*, 402 F.3d 511 (5th Cir. Mar. 4, 2005), *U.S. v. Lincoln*, 413 F.3d 716 (8th Cir. July 7, 2005), *U.S. v. Ameline*, 415 F.3d 606 (7th Cir. July 7, 2005), *U.S. v. Williams*, 436 F.3d 706 (6th Cir. Jan. 31, 2006), *U.S. v. Green*, 436 F.3d 449 (4th Cir. Feb. 6, 2006), *U.S. v. Knight*, 437 F.3d 1050 (10th Cir. Feb. 17, 2006), *U.S. v. Dorcely*, 454 F.3d 366 (D.C. Cir. July 21, 2006).

83. Brief for the Federal Public and Community Defenders & the National Association of Federal Defenders as Amicus Curiae Supporting Petitioners, *supra* note 7, at *12.

Figure 2: Data from USSC Brief



2. Different Methods Lead to Different Conclusions

The differences between the two graphs, and in the conclusions drawn from them, can be explained largely by differences in the methods used to create them.[84] The FPCD was burdened with several serious data limitations that the USSC could avoid since the USSC collects the data and has complete access. First, since the USSC's publicly-released data is quarterly and not monthly, the FPCD could not provide monthly data that could more precisely pinpoint possible effects. Second, the USSC's quarterly data releases are cumulative, i.e., they do not break out the rates in mutually exclusive time periods. Thus, there is overlap in the periods on the FPCD graph that makes clean comparisons impossible.

Another methodological difference concerned how to classify circuits. Some circuits had a decision adopting the presumption of reasonableness, others had a decision explicitly declining to adopt such a presumption, and others (and all at the beginning) had no decision either way. As reported in the footnote to its graph, the FPCD compared circuits that had adopted a presumption, represented by the lower line in the graph, with all others.[85] As reported in the Methodology section of the Appendix, the USSC compared circuits that had adopted the presumption with circuits that had held that sentences within the guideline range are

84. One difference concerned the law. The FPCD found that the Fifth Circuit became a presumptive circuit in the case *United States v. Mares*, 402 F.3d 511 (5th Cir. 2005). The USSC did not classify that circuit as presumptive until several months later in *United States v. Alonzo*, 435 F.3d 551 (5th Cir. 2006).

85. Brief for the Federal Public and Community Defenders & the National Association of Federal Defenders as Amicus Curiae Supporting Petitioners, *supra* note 7, at app. 1.

not presumed reasonable.[86] Circuits that had reached no conclusion were excluded from the graph.[87] It is not immediately obvious which comparison is superior. One might expect any effect of the presumption to be strongest in a comparison among circuits that had explicitly chosen one way or the other. Yet the Sentencing Commission's data, which made that comparison, showed no effect.

Another methodological issue makes drawing conclusions from the FPCD analysis difficult. The various circuits have markedly different below-range rates, and have had these rates for a long time. These differences predate the adoption of the presumption and reflect a myriad of institutional, cultural, political, and other factors. The USSC graph divides circuits according to whether they explicitly adopted or declined to adopt the presumption *at any time.* The rates for these circuits are then compared across the entire period. The same circuits are compared in each month. The FPCD graph divides the circuits based on whether they had adopted the presumption by that *particular time.* Thus different circuits are compared at different times on the graph.

Whenever a circuit is added to the mix, it adds to the overall rate both any influence of the presumption and also all the other influences that make some circuits have higher rates than others. An extreme example helps make this clear. If the most "conservative" circuits have the lowest below-range rates and are also the most eager to declare within-guideline sentences presumptively reasonable, the rates earliest on the line representing presumptive circuits would be suppressed by those circuits' overall conservatism. If less conservative circuits adopt the presumption later, the overall rate could well go up, even if the presumption does cause rates to be suppressed to some extent. If one accepts this example as plausible, the method used by the FPCD actually stacked the deck against supporting its claim. But in fact, the data do not fit this extreme example. It is difficult, however, to decipher from the FPCD chart how the timing of the circuit decisions may have influenced the rates, and whether the presumption itself had an independent effect. In short, and in social science lingo, the effects of other differences among the circuits are confounded with any possible effects of the presumption.

3. Examining Individual Circuits

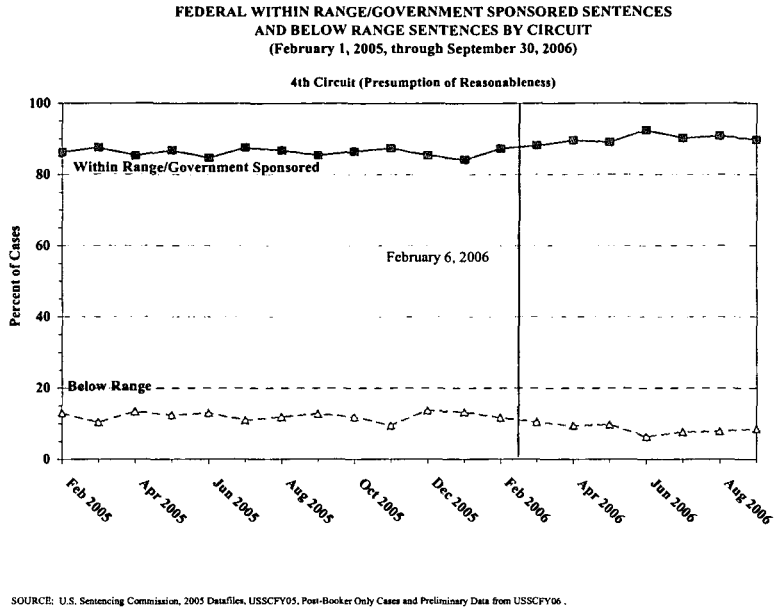
One way to avoid this problem is to look at each circuit individually. The USSC brief included separate graphs for each circuit, showing their monthly within-system and below-range rates along with a horizontal line indicating the month the circuit adopted the presumption or de-

86. Brief for the United States Sentencing Commission as Amicus Curiae Supporting Respondent, *supra* note 7, at app. 14a.

87. *Id.* at app. 14 (including data from all circuits except for the Federal Circuit).

clined to adopt the presumption.[88] Two of these graphs are included here: the Fourth Circuit (Figure 3), from which *Rita* was appealed and consistent among the circuits as having the lowest below-range rate; and the Second Circuit (Figure 4), which has the highest below-range-rate. Examination of these graphs and those for the other circuits included in the brief[89] demonstrates that it remains difficult to draw clear conclusions regarding the effects of a presumption.

Figure 3: Fourth Circuit

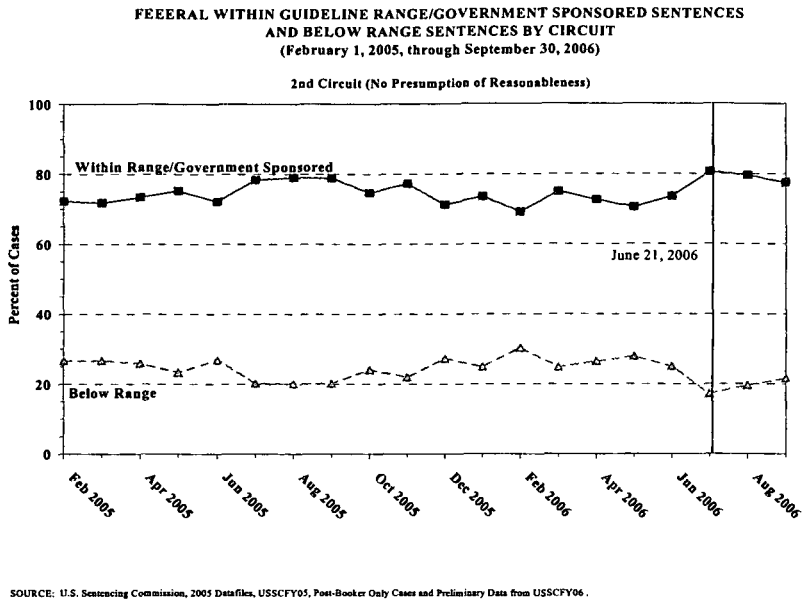


SOURCE: U.S. Sentencing Commission, 2005 Datafiles, USSCFY05, Post-Booker Only Cases and Preliminary Data from USSCFY06.

88. *Id.* at app. 1-13.

89. The full set of graphs for each circuit, along with the rest of the USSC's brief, can be found at http://www.ussc.gov/sc_cases/USSC%20Amicus%20BriefNos06-5618_06-5754.pdf.

Figure 4: Second Circuit



SOURCE: U.S. Sentencing Commission, 2005 Datafiles, USSCFY05, Post-Booker Only Cases and Preliminary Data from USSCFY06.

In the Fourth Circuit, the rate declines after the presumption was adopted.[90] In the Second Circuit, it appears lower after the presumption was *rejected.*[91] In the Third Circuit, not shown here, there is a similar counter-hypothesis trend.[92] In others there is no obvious effect at all, or the post-presumption rate appears to continue a pre-existing downward trend, or the data are too few or too volatile to draw any conclusions. The most precise thing one can say is that there is no obvious effect.

4. The Advantages of a More Rigorous Statistical Test

So-called "graphical analysis"—just looking at the graphs to see if anything jumps out at you—can sometimes work to draw sound causal conclusions.[93] This is especially true if the effect one is looking for is very strong and thus obvious. Where the effect is weak and different circuits show different tendencies, any overall effect of the presumption can easily be missed. It would be a more powerful test of the hypothesis if data from all the circuits could somehow be examined at once.

90. *See* Brief for the United States Sentencing Commission as Amicus Curiae Supporting Respondent, *supra* note 7, at app. 5.

91. *Id.* at app. 3.

92. *Id.* at app. 4.

93. *See* Hofer, *supra* note 12.

The statistical method multivariate regression can disentangle any effect of the presumption from other pre-existing differences between the circuits and also from any unknown influences that may have affected rates in all circuits month-to-month. It also compares all the circuits at once, and tests whether any difference between presumption and non-presumption circuits is sufficiently great to rule out chance as a possible explanation. Unfortunately, from a scientific perspective, no party or *amici* used this method in their submissions to the Court. From a legal and practical perspective, this may be understandable because most lawyers and judges are unfamiliar with the method and uncomfortable relying on its results. A knowledge gap exists between the statistical skills of most lawyers and judges and the skills needed to apply and understand the best methods for answering empirical questions embedded in the law.[94]

A multivariate analysis was not presented to the Supreme Court, and the justices were left with competing claims and less-than-ideal data and methods to resolve them. Perhaps it is best that the majority decided, as discussed in Part I, that the actual effects on judicial behavior of adoption of the presumption were of no constitutional significance. Hopefully, the conflicting and complicated statistical findings on this empirical issue did not themselves influence the Justices to reject the significance of the question. Legal Realists, who pioneered empiricism and the Brandeis Brief, have always believed that legal decisions are best made in light of the facts.

5. The Effect and Non-Effect of Legal Doctrine on Sentencing Behavior

Some lawyers are reluctant to believe that Supreme Court decisions on apparently important legal questions sometimes do not make any difference in practice. Such was the case with an earlier Supreme Court decision that set a nationwide standard for review of sentences outside the guideline range, *Koon v. United States*.[95] *Koon* established the "abuse-of-discretion" standard for appellate review of guideline departures.[96] Attorneys in the Department of Justice subsequently blamed *Koon* for an increase in downward departures and argued that re-establishment of a *de novo* standard—which was accomplished in the PROTECT Act but undone by *Booker*—was needed to bring the non-government sponsored below-range rate back down.[97] In fact, although

94. For an early and comprehensive exploration of this issue, see generally John Monahan & Laurens Walker, *Social Authority: Obtaining, Evaluating, and Establishing Social Science in Law*, 134 U. PA. L. REV. 477 (1986).

95. 518 U.S. 81 (1996).

96. *Id.* at 99-100.

97. *Child Pornography & Abduction Prevention, Before the Subcomm. on Crime, Terrorism, and Homeland Security of the H. Comm. on the Judiciary*, 108th Cong. 17-18 (2003) (statement of Daniel P. Collins, Associate Deputy Att'y General). *See generally* William W. Mercer, *Assessing*

the downward departure rate did increase after *Koon*, the data show that the increase was the continuation of a long-term trend that predated *Koon* and was likely based on a variety of factors and little influenced by *Koon* itself.[98]

Looking at the history of below-range rates under the Guidelines, it is apparent that sometimes legislation and court decisions have affected below-range rates and sometimes they have not.[99]

There was a long-term and gradual increase in below-range sentences for the first dozen years after implementation of the Guidelines.[100] Then, remarkably, the trend was reversed around 2001.[101] This reversal took place *before* enactment of the PROTECT Act, the sentencing sections of which were specifically designed to decrease non-government sponsored below-range sentences. The PROTECT Act doubtless contributed to the downward trend in below-range sentences that continued until the decision in *Booker*, but it was itself a reflection of an institutional shift at the Department of Justice that occurred with the change in presidential administrations. For better or worse, the Department of Justice and U.S. Attorneys around the country set out to enforce the guideline range more vigorously, and these efforts were reflected in the sentencing data.[102]

Unlike *Koon*, *Booker* had an immediate and dramatic effect on sentences. The rate of both non-government sponsored below-range sentences and above-range sentences approximately doubled compared to the PROTECT Act era, restoring the below-range rate to a level similar to earlier in the guideline era.[103] In the most recent quarters, the rate has held steady been twelve percent.[104] The rate of above-range sentences in the most recent quarters is 1.5 percent.[105]

Compliance with the U.S. Sentencing Guidelines: The Significance of Improved Data Collection and Reporting, 16 FED. SENT'G REP. 43 (2003) (observing that non-substantial assistance downward departures increased significantly after *Koon*).

98. Mark T. Bailey, Note, *Feeney's Folly: Why Appellate Courts Should Review Departures from the Federal Sentencing Guidelines with Deference*, 90 IOWA L. REV. 269, 295-300 (2004) (noting that the government rarely appealed downward departures after *Koon* but was highly successful when it did, and arguing that *Koon* had little effect on departure rates); Paul J. Hofer ET AL., *supra* note 77, at 284-86, 290.

99. *See* Hofer, *supra* note 13 at 427-29, 431-34 (for a graph and further discussion of long-term trends).

100. *Id.* at 428-29 (including Figure 1: Monthly Rates of In-Range and Out-of-Range Sentences).

101. *Id.*

102. For a discussion of the Department of Justice's institutional relationship with the Sentencing Guidelines, see Frank O. Bowman III, *Mr. Madison Meets a Time Machine: The Political Science of Federal Sentencing Reform*, 58 STAN. L. REV. 235 (2005).

103. Hofer, *supra* note 13, at 433.

104. U.S. SENTENCING COMMISSION PRELIMINARY QUARTERLY DATA REPORT (2ND QUARTER RELEASE PRELIMINARY FISCAL YEAR 2007 DATA THROUGH MARCH 31, 2007) 1 (tbl. 1) (2007), http://www.ussc.gov/sc_cases/Quarter_Report_2Qrt_07.pdf (last visited September 15,

While *Booker* increased the rate of below-range sentences based on mitigating factors identified by the sentencing judge, the most common reasons for sentencing outside the guideline range have always been government-sponsored. In the early years of the Guidelines, this was due to a government motion for a sentence reduction to reward an offender's substantial assistance in the prosecution of other persons,[106] or for other reasons included as part of a plea agreement with the defendant. The PROTECT Act added a new reason—the defendant's participation in an "early disposition program," which rewards an offender for pleading guilty quickly and foregoing other procedural rights.[107] In the most recent quarter, 25.1 percent of offenders received a below-range sentence for one of these government-sponsored reasons.[108] Clearly, the most important influences on within-range rates are not the legal standards governing appellate review of judge-initiated departures, but the policies and programs of the Department of Justice.

6. The Likely (Non-) Effect of *Rita*

Is the effect of *Rita* likely to be substantial like *Booker* or weak like *Koon*? The best guess is weak. In part, this is supported by the absence of an obvious effect of the presumption in the circuits. In part, it is because of the weakness of the decision itself. The court majority appeared determined to underscore the unimportance of the presumption, describing it as "non-binding" and contrasting it with evidentiary presumptions that shoulder one party with a burden of persuasion or proof. Justice Stevens' concurring opinion also makes clear that he endorses the presumption because it is consistent with the deference owed to the decisions of sentencing judges. He emphasizes, however, that similar deference is owed to their decisions to sentence *outside* the guideline range, raising a question of what exactly the presumption means for substantive review.

The decision in *Rita* seems weak and somewhat strange in another way: While the petition for certiorari was granted in the face of a conflict in the circuits, the decision doesn't appear to resolve the conflict. The decision states that a court of appeals *may* apply a presumption of reasonableness.[109] But the courts do not appear *required* to do so. One commentator has noted that this lack of resolution of the circuit conflict

---

2007). The Commission releases new quarterly figures on its website, www.ussc.gov, as soon as they become available.

105. *Id.*

106. U.S. SENTENCING GUIDELINES MANUAL, *supra* note 11, § 5K1.1; *see also* Stephanos Bibas, *Criminal Law: The Feeney Amendment and the Continuing Rise of Prosecutorial Power to Plea Bargain*, 94 J. CRIM. L. & CRIMINOLOGY 295, 299-300 (2004) (noting that the largest class of downward departures consists of § 5K1.1 motions).

107. U.S. SENTENCING GUIDELINES MANUAL, *supra* note 11, § 5K3.1.

108. U.S. SENTENCING COMMISSION PRELIMINARY QUARTERLY DATA REPORT, *supra* note 104, at 1 tbl.1.

109. Rita v. United States, 127 S. Ct. 2456, 2462 (2007).

the case was chosen to address seems unusual, if not unique, in the annals of Supreme Court jurisprudence.[110] It is even more odd given that reducing unwarranted disparity is the primary goal of the sentencing Guidelines. As it stands, *Rita's* effect is likely to be similar to *Koon* in another way: it will perpetuate marked differences among the circuits in their rates of below-range sentences.

CONCLUSION

The Court's decision to permit, but not require, a presumption of reasonableness for a within-guideline sentence is likely to have little effect. But this is not to say there would have been no effect if the decision had come out the other way. Indeed, the empirical question briefed in *Rita* was in some respects beside the point. The important question was not the effect of adopting the presumption in the circuits, but the effect of the Supreme Court adopting or rejecting the presumption. The drug would not have equal dosage regardless of the doctor, nor is its strength necessarily equal whether it is given or taken away.

If the Court in *Rita* had rejected a presumption of reasonableness for within-guideline sentences, it would have rejected the arguments the government and the USSC had made in support of it. Chief among these were that the Guidelines remain central in federal sentencing and that they deserve substantial weight because they reflect the Sentencing Commission's considered judgment of the best way to achieve the statutory purposes of sentencing found in 18 U.S.C. §3553(a).[111] Included in this general argument are a number of specific empirical questions that are ultimately of far greater consequence than the effects of a presumption of reasonableness because they concern the foundations of the Guidelines and their success at achieving the purposes of sentencing and the goals of sentencing reform.

A. Empirical Bases of Guideline Development and Amendment

One of these specific questions concerns the empirical basis for the Guidelines. Justice Breyer was, of course, present at the creation as a Sentencing Commissioner. In *Rita* he repeats his oft-told tale of the philosophical dilemmas that caused problems for the original Sentencing Commission until it turned to an "empirical approach" to Guideline development. Based on a statistical analysis of 10,000 cases sentenced in the years immediately preceding their work, the USSC set the initial guideline ranges on the sentence that had been imposed for various categories of offenses and offenders.[112] In this manner, they expected the

110. Posting of Peter Goldberger to SCOTUSblog, http://www.scotusblog.com/movabletype/archives/2007/06/another_view_of.html (June 21, 2007, 08:48 PM).

111. *Rita*, 127 S. Ct. at 2463.

112. Stephen Breyer, *The Federal Sentencing Guidelines and the Key Compromises upon Which They Rest*, 17 HOFSTRA L. REV. 1, 7-8 (1988); *see also* Paul J. Hofer & Mark H. Allenbaugh,

Guidelines to largely reflect average past practice. Though the goals of *uniformity* and *proportionality* often conflict, the Sentencing Commission developed a practical system that it believed sensibly reconciled the two ends. Because of this history, Justice Breyer writes in *Rita* that "[t]he Guidelines as written reflect the fact that the Sentencing Commission examined tens of thousands of sentences and worked with the help of many others in the law enforcement community."[113] He goes on to explain that the "Commission's work is ongoing," and that "[t]he statutes and the Guidelines themselves foresee continuous evolution helped by the sentencing courts and courts of appeals in that process."[114] He describes how the amendment process will proceed through examination of judges' stated reasons for departure, circuit court review of those reasons, and through "advice from prosecutors, defenders, law enforcement groups, civil liberties associations, experts in penology, and others."[115]

A lot has happened since Justice Breyer left the Sentencing Commission. Absent from his description of the Commission's work is any discussion of the role played by mandatory minimum penalty statutes, specific directives from Congress to the Sentencing Commission to increase penalties or set them at particular levels, or the many other ways that Congress has shaped the present Guidelines.[116] This absence leaves unanswered important questions about how judges are to treat guideline ranges that do not reflect past practice or any of the laudatory guideline

---

The Reason Behind the Rules: Finding and Using the Philosophy of the Federal Sentencing Guidelines, 40 AM. CRIM. L. REV. 19, 35-51 (2003) (discussing how the philosophy of the Guidelines was shaped, and obscured, by the guideline development process).

113. *Rita*, 127 S. Ct. at 2464.

114. *Id.*

115. *Id.* To support his account of the Guidelines' development and amendment process, Justice Breyer cites § 1B1.10(c) of the Guidelines manual, which lists 24 amendments that the USSC has chosen to apply retroactively. *Id.* To date, the USSC has promulgated over 700 amendments, a large proportion in response to acts of Congress. *See* UNITED STATES SENTENCING COMMISSION, FIFTEEN YEARS OF GUIDELINES SENTENCING (AN ASSESSMENT OF HOW WELL THE FEDERAL CRIMINAL JUSTICE SYSTEM IS ACHIEVING THE GOALS OF SENTENCING REFORM) (2004) app. B (B-1 to B-9), *available at* http://www.ussc.gov/15_year/15_year_study_full.pdf. Justice Breyer also cites § 1A1.1, commentary, which contains an account of the empirical guideline development process similar to that described by Justice Breyer in *Rita*. 127 S. Ct. at 2464 (citing U.S. SENTENCING GUIDELINES MANUAL, *supra* note 11, § 1A1.1). It is worth noting, however, that as part of the amendments made pursuant to the PROTECT Act, this comment was moved to a unique new section, "Historical Review of Original Introduction." This was done in part because the note no longer reflected the realities of the present Guidelines. *See* U.S. SENTENCING GUIDELINES MANUAL, *supra* note 11, § 1A1.1.

116. Justice Breyer is consistent in neglecting the influence of Congress. His seminal article on the development of the Guidelines devotes several pages to the USSC's decision to increase penalties for white collar offenses above average past practice. *See* Breyer, *supra* note 112. These levels have long since been increased further, most recently by Congressional directives in the Sarbanes-Oxley Act. But he mentions only in passing, and in a footnote, that the penalties for drug trafficking offenses were set far above average past practice in order to incorporate the mandatory minimum statutory penalties enacted by Congress in the Anti-Drug Abuse Act of 1986. *See id.* at 24 n.121. The decision to incorporate the mandatory minimum penalties in the manner that it did had a more profound effect on the federal prison population, and on the lives of many tens of thousands of defendants, than any other decision of the original USSC. This impact was dramatically underestimated by the USSC at the time the Guidelines were implemented. *See id.* at 24.

amendment processes envisioned in the Sentencing Reform Act, but instead the will of Congress expressed through the medium of the sentencing Guidelines. Next term's case involving the Guideline for trafficking in crack cocaine, *Kimbrough v. United States*,[117] will give the Court an opportunity to weigh in on this crucial question.

B. Empirical Evaluation of the Guidelines

The ultimate questions—part empirical, part policy—involve the relationship between the Guidelines, the statutory purposes of sentencing, and the goals of sentencing reform. Much of the argument in *Rita* concerned if and how the USSC incorporated these purposes, and how well the Guidelines are achieving them.

Sentencing reform was undertaken to a large degree to reduce unwarranted disparity. The *Rita* briefs summarize research to evaluate the Guidelines' success at achieving this goal. But again, different parties looked at the same evidence and reached different conclusions. Petitioner Rita cited the USSC's own research on the effects of presentencing stages on sentencing disparity, on regional disparity, and on the adverse impact on black offenders of particular guideline provisions such as that concerning crack cocaine.[118] He concluded that the Guidelines had "only partially achieved" the reduction of unwarranted disparity.[119] The Department of Justice and the USSC, however, cited different sections of the same report to argue that the Guidelines had reduced disparity from the source at which they were primarily targeted—judicial discretion.[120]

At this point, the USSC's argument took a turn away from empiricism and toward a matter of definitions. "In focusing on disparities like the crack/powder ratio, petitioners misapprehend the intended purpose of § 3553(a)(6),"[121] which is the statutory provision directing judges to consider "the need to avoid unwarranted sentence disparities among defendants with similar records who have been found guilty of similar conduct." Disparity that is mandated by Congress, the brief argued, cannot by definition be unwarranted. Moreover, regional or other disparities caused by prosecutorial decisions, such as charge bargaining, "cannot be considered *unwarranted* disparity within the meaning of §

117. 127 S. Ct. 2933 (2007), *cert. granted*, 60 U.S.L.W. 3661 (U.S. June 11, 2007) (No. 06-6330).

118. *See* Brief for Petitioner, *supra* note 7, at *36-37 (quoting UNITED STATES SENTENCING COMMISSION, FIFTEEN YEARS OF GUIDELINES SENTENCING, *supra* note 115, at 135, 141-43).

119. *Id.* at 36.

120. *See* UNITED STATES SENTENCING COMMISSION, FIFTEEN YEARS OF GUIDELINES SENTENCING, *supra* note 115, at 94-99.

121. Brief for the United States Sentencing Commission as Amicus Curiae Supporting Respondent, *supra* note 7, at *27.

3553(a)(6)"[122], even though the USSC has taken measures to reduce such disparity (with limited success).

With these arguments, the debate has been removed from the realm of empiricism and placed firmly on the ground of semantics. Even if the objective facts about the Guidelines' effects on disparity are clear, no consensus on the Guidelines' success can emerge until agreement is reached on the meaning of "unwarranted." Nor can "disparity" be defined without consensus on the purposes of punishment and the priorities among them.[123] Judging from the briefs in *Rita*, this consensus is as far away as ever.

C. Empirical Questions About Individual Guidelines

Finally, the majority opinion in *Rita* may contain an invitation for judges to entertain arguments and take evidence on a type of challenge to the Guidelines that raises empirical questions at the heart of sentencing practice. Defendants and prosecutors have argued, with some success, that extraordinary circumstances can make the guideline range ineffective or disproportionate in a particular case, and that an outside-the-range sentence is needed.[124] *Rita* seems to invite a more general challenge—that a particular guideline is ineffective or disproportionate in a wide range, or even the majority, of cases to which it linguistically applies. In short, that the Guideline itself represents unsound policy.

To date, the courts of appeals have not looked favorably on these "categorical challenges." But Justice Breyer may be asking them to look again. He lists the kinds of arguments a judge might hear that the guideline sentence should not apply to a case:

> [P]erhaps because (as the Guidelines themselves foresee) the case at hand falls outside the "heartland" to which the Commission intends individual Guidelines to apply . . . , *perhaps because the Guidelines sentence itself fails properly to reflect the § 3553(a) considerations*, or perhaps because the case warrants a different sentence regardless.[125]

Later, the opinion describes the kinds of reasons sentencing judges need to articulate for the sentences they impose: If it is clear that a judge rests her decision upon the USSC's own reasoning, she need say very little.[126] But, if "a party contests the Guidelines sentence generally under § 3553(a)—that is, argues that the Guidelines reflect an unsound judg-

---

122. *Id.* at 28 (emphasis in original).
123. Hofer, *supra* notes 12, 13, and 77.
124. *See generally* Regina Stone-Harris, *How to Vary from the Federal Sentencing Guidelines Without Being Reversed,* 19 FED. SENT'G REP. 183 (2007).
125. Rita v. United States, 127 S. Ct. 2456, 2461, 2465 (2007) (emphasis added).
126. *Id.* at 2468.

ment, or, for example, that they do not generally treat certain defendant characteristics in the proper way,"[127] then more explanation is needed.[128]

Could it be that in *Rita* the Court is inviting the lower courts to hear evidence that particular Guidelines are not working to achieve the statutory purposes? Only time, and perhaps the upcoming opinion in *Kimbrough*, will tell. If it turns out the Court is now open to categorical challenges of the type many commentators have long encouraged,[129] *Rita* may be just the beginning of empirical evidence on sentencing questions presented to the Court.

127. *Id.*

128. All this makes more intriguing the opinion's description of the effects of a presumption of reasonableness for the guideline sentence: "Nor does the presumption reflect strong judicial deference of the kind that leads appeals courts to grant greater factfinding leeway to an expert agency than to a district judge." *Id.* The facts being referred to in this passage do not sound like facts about the case, but facts of the type also relevant to the USSC. In other words, empirical facts about the effects of the Guidelines and its success or failure at achieving its purposes.

129. For the earliest example of this "administrative law" approach to departures, see Ronald F. Wright, *Sentencers, Bureaucrats, and the Administrative Law Perspective on the Federal Sentencing Commission*, 79 CAL. L. REV. 1, 50-55 (1991); *see also* United States v. Jaber, 362 F. Supp. 365, 373 (D. Mass. 2005) (arguing for judges to compensate for the USSC's exclusion from the Administrative Procedures Act by developing a common law of sentencing that correlates particular guidelines with the statutory purposes); Joseph Luby, *Reining in the "Junior Varsity Congress": A Call for Meaningful Judicial Review of the Federal Sentencing Guidelines*, 77 WASH U. L.Q. 1199, 1239-40, 1255-67 (1999). For an argument that a more radical change in sentencing guidelines is needed, with a different type of judicial review, see Kate Stith & Karen Dunn, *A Second Chance for Sentencing Reform: Establishing a Sentencing Agency in the Judicial Branch*, 58 STAN. L. REV. 217, 224-29 (2005).