Law & Society REVIEW

Questioning Blackmun's Thesis: Does Uniformity in Sentencing Entail Unfairness?

Ben Grunwald

It is commonly believed among criminal justice scholars that sentencing guidelines increase uniformity in sentencing at the cost of fairness. They reason that guideline systems rarely take all relevant case characteristics into consideration, and as a result, impose sentences in particular cases that are biased relative to the ideal or best sentence. This *bias effect* is one of the primary theoretical and practical challenges faced by courts and sentencing commissions in the last 30 years, and provides one of the strongest arguments against mandatory sentencing guidelines. This article identifies a second effect of guidelines on fairness, which has not been sufficiently acknowledged by the scholarly literature: the *variance effect* increases the fairness of sentences directly by increasing uniformity. This article uses statistical simulation to examine the relationship between the variance effect and the bias effect. The results provide substantial evidence that the variance effect is comparatively large, and that it may often outweigh the negative effects of bias. Under these conditions, sentencing guidelines will both increase uniformity and increase fairness.

Until the 1970s, judges in the United States enjoyed nearly unlimited discretion in assigning sentences to criminal offenders. Judges were free to adopt their own theory of punishment, to determine how that theory applied to the facts of a case, and to select the most appropriate punishment scheme on that basis. Typically, only very wide statutory ranges constrained judges' power to individualize sentences (Stith and Koh 1993). Early empirical research from the 1960s to the early 1980s revealed that the existing system produced large sentencing disparities, finding that similar defendants with similar convictions often received different punishments. Scholars theorized that these disparities arose from variations in judges' ideological background (Forst and Wellford 1981; Partridge and Eldridge 1974), from discrimination based on legally irrelevant characteristics such as race and ethnicity (Baldus, Pulaski, and Woodworth 1983), and from differences in local context (see Dixon 1995). Contemporary

Please direct all correspondence to Ben Grunwald, Department of Criminology, University of Pennsylvania; e-mail: Ben.Grunwald@gmail.com.

Law & Society Review, Volume 49, Number 2 (2015)

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

research confirms that these phenomena continue to impact sentencing today (Abrams, Bertrand, and Mullainathan 2012; Albonetti 1997; Bushway and Piehl 2001; Kaut 2002; Ulmer and Johnson 2004).

In response to evidence of disparity, federal and state legislatures established sentencing guideline systems to constrain judicial discretion. These systems typically used a limited number of variables to determine a sentence range for each defendant. Judges were then encouraged and in some systems required to impose a sentence within that range (Frase 1995). The level of discretion left up to the judge varied widely by jurisdiction, but no system was more restrictive than the Federal Sentencing Guidelines, which gave nearly all weight to crime severity and criminal history (Tonry 1993).

Almost immediately, the Federal Guidelines were criticized for promoting uniformity at the cost of fairness in individual cases (e.g., Ogletree 1987). Critics widely argued that the federal guidelines decreased the fairness of sentences by constraining judges' ability to take relevant case characteristics into consideration. I call this the *bias effect* of sentencing guidelines: guidelines can bias sentences away from the fairest or most appropriate sentence by limiting judges' ability to take all relevant case characteristics into consideration and to fully individualize punishment. While discussing disparities in the administration of the capital punishment, Justice Harry Blackmun articulated a well-known statement of the *bias effect*: "Experience has shown that … consistency and rationality … are inversely related to [fairness]. A step toward consistency is a step away from fairness" (*Callins v. Collins* 1994). Although Blackmun articulated this thesis in the context of the death penalty, I take his critique as a paradigmatic formulation of a broad and popular criticism of mandatory sentencing guidelines echoed by judges (Schwarzer 1991) and scholars (Alschuler 1991; Freed 1992; Ogletree 1987; Tonry 1993; Ulmer 1996), which persists until today (Kim 2004; O'Hear 2006; Osler 2003).

This article advances the sentencing literature in two ways. First, it develops a novel framework for conceptualizing the relationship between uniformity and fairness. Second, it tests Blackmun's Thesis by identifying and examining a second effect of sentencing guidelines that has not been acknowledged by the academic literature: increasing uniformity through guidelines has a second effect—a *variance effect*—of directly increasing the fairness of sentences on average.

This article uses simulation modeling to examine the relationship between sentencing guidelines, uniformity and fairness. I begin by defining an "ideal" sentence for a set of equivalent

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

criminal cases based on assumptions that are favorable to Blackmun's Thesis. I then randomly generate a preguideline distribution of sentences centered around the "ideal" sentence. Based on existing estimates from the literature, I introduce a sentencing guideline system, and posit a series of plausible bias and variance effects on the sentence distribution. The average distance of sentences in the preguideline and postguideline distributions from the "ideal" sentence is then compared to determine whether, under these assumptions, Blackmun's Thesis would hold. This approach provides an important methodological benefit. Unlike many other analytic methods in the sentencing literature, this article avoids the need to assume a thick normative theory about the purposes of punishment and the case characteristics that are relevant in sentencing.

The results of the analysis show that, under plausible conditions, the *variance effect* of sentencing guidelines is comparatively large and may often outweigh the negative effects of *bias*. When the bias effect is outweighed, Blackmun's Thesis does not hold, and the guidelines both increase uniformity and increase or maintain the existing level of fairness, thereby defusing one of the most formidable arguments against restrictive sentencing guideline systems.

The remainder of the article proceeds as follows. I begin with a brief history of sentencing guidelines, and a review of the empirical literature on sentencing disparity. Next, I develop a conceptual framework to clarify and explore the contours of Blackmun's Thesis, and then describe the basic design of the study. I conclude by reporting the results of the analysis, and by discussing the implications for the academic and policy debate on sentencing guidelines.

Sentencing Guidelines

From the late 1970s to 1990s, Congress and a number of state legislatures established sentencing commissions to design guidelines that would increase uniformity in sentencing. Congress, for example, established the United States Sentencing Commission through the Sentencing Reform Act of 1984. The commission was authorized to develop general rules to regulate the form (e.g., fine, probation, or imprisonment) and intensity of sentences on the basis of a list of variables including seriousness of offense, criminal history, age, education, vocational skills, mental and emotional condition, family responsibilities, community ties, and extent of participation in the offense. In 1987, the commission passed a guideline system based primarily on two of

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

those variables, seriousness of offense and criminal history. A limited number of additional case features were also given some weight (e.g., the amount of money stolen or the use of a weapon) (Ogletree 1987). Based on these characteristics, the Guidelines were designed to output a narrow range of sanctions from which judges were required to select the most appropriate punishment.

A number of states such as Minnesota, Pennsylvania, Virginia, and Massachusetts also enacted sentencing guidelines (Frase 1995). These guideline systems varied widely, but in general, they were less restrictive, complex, and controversial than the federal system (Tonry 1993). Unlike the presumptive or mandatory system enacted by Congress, many state legislatures adopted advisory guidelines that served as recommendations rather than binding rules (Frase 1995). Many state guideline systems, including that of Minnesota and Pennsylvania, have received approval from legal scholars and social scientists (Tonry 1993).

Today, all federal and state guideline systems are advisory as a result of two Supreme Court cases, *Blakely v. Washington* (2004) and *Booker v. United States* (2005). The tension between uniformity and fairness in sentencing guidelines, however, remains a live policy debate. First, empirical evidence suggests that advisory guidelines continue to influence judges' sentencing practices (Bushway, Owens, and Piehl 2012; Pfaff 2006). Second, the Supreme Court rendered sentencing guidelines advisory on a relatively narrow issue, and scholars have noted the availability of mandatory systems that would survive *Blakely* and *Booker* review (Chanenson 2004). Third, the United States Sentencing Commission has recently proposed legislative and appellate constraints on judicial discretion, which according to two scholars would "restore the Guidelines very nearly to the legal status they enjoyed before *Booker*" (Starr and Rahavi 2013: 9).

Empirical Research on Sentencing Disparity

I review the empirical literature on sentencing disparity for two purposes. First, a review of existing empirical methodologies helps clarify the strengths of the current study. Second, it also helps set plausible bounds on the parameters of the quantitative analysis by answering two key questions: what is the magnitude of sentencing disparity, and what is the effect of sentencing guidelines on disparity? Researchers have used two main methodological approaches to answer these questions: *comparable distributions* and *identical cases*. I discuss each literature separately.

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

The Comparable Distribution Approach

The comparable distribution approach assumes that certain groups of cases are statistically equivalent either by *controlling for observable variables* or by *exploiting random assignment* of cases to judges.

Studies that *control for observable variables* typically measure sentencing disparity by variation in the dependent variable. These studies find that total sentencing disparity decreases after the enactment of guidelines. Karle and Sager (1991), for example, report substantial reductions in the variation of sentences within broad categories of crime (e.g., robbery) after the enactment of the federal guidelines. A United States Sentencing Commission report examines narrower categories of crime (e.g., bank robbery of less than $10,000 with no criminal history), and finds that variation in sentences decreased by 15 to 60 percent for nearly all categories examined after the federal guidelines were enacted (USSC 1991). Stolzenberg and D'Alessio (1994) fit a linear regression model with four independent variables[1] and measured "total sentencing disparity unexplained by legally mandated sentencing factors." The authors estimate a relative reduction in sentencing disparity of roughly 60 percent after the enactment of the guidelines in Minnesota (Stolzenberg and D'Alessio 1994: 302).

As others have noted, sentencing studies that control on observable variables have several important methodological limitations. First, the models may not include all relevant independent variables (Baumer 2013). This is particularly problematic for studies that measure disparity based on unexplained variation in sentence lengths (e.g., Stolzenberg and D'Alessio 1994). In these studies, the omission of *any* relevant variable will bias the measure of disparity regardless of its correlational structure. Thus, in practice, these studies capture both unwarranted *and* warranted sources of disparity.

Second, omitted variable bias is particularly problematic in the sentencing disparity literature because there is little normative consensus among judges and scholars about the variables that are relevant to sentencing (Hofer, Blackwell, and Ruback 1999; Rhodes 1991). Some studies have attempted to address this problem by defining equivalent cases based on the categories of crimes defined by a sentencing guideline system (e.g., Rhodes 1991). But, unless one believes that the guideline system has correctly grouped similar cases, these studies merely confirm that

[1] Offense seriousness, criminal history, presence of a weapon, and whether the most serious conviction offense was a personal crime.

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

"post-Guidelines sentences are more likely to be in accordance with the Guidelines" (Anderson, Kling, and Stith 1999: 280; Tonry 1993). Indeed, testing whether sentencing disparity itself changed would require a normative theory about the case characteristics that are relevant at sentencing—a theory over which many readers are likely to disagree.

Third, most sentencing studies that control for observable variables focus exclusively on sentencing, and do not capture disparities arising from earlier phases in the criminal justice system (see Baumer 2013). This criticism applies to most studies in the literature, including those that adopt other methodological approaches. But, they are particularly relevant for studies that examine the effects of guidelines on disparity. Guidelines may introduce disparities into the charging process by increasing the power of prosecutors during plea negotiations (e.g., Miethe 1987; Starr and Rehavi 2013). As a result, when studies find that sentencing disparity has decreased after guideline enactment, it is often unclear whether the disparity has merely moved to an earlier stage in the process.

Some scholars have taken an alternative methodological approach by *exploiting the random assignment of cases to judges*. Since random assignment ensures that judges receive roughly equivalent caseloads, social scientists have compared the mean sentence of each judge in the same district to assess interjudge sentencing disparity. The earliest study to use this approach found that among six judges in one district, one judge imposed a median sentence of 6 months, while another imposed a median sentence of 12 (Gaudet, Harris, and John 1934).

More recent studies have used random assignment to estimate the effect of federal guidelines on interjudge disparity. In a study of three federal districts, Waldfogel (1991) estimates the mean absolute deviation—the average difference between the mean sentence length of each judge and the grand mean sentence length for all judges—before and after the enactment of the federal guidelines. Preguidelines, the mean absolute deviation was between 4.5 and 6 months, or between 12 and 26 percent of the mean sentence length, respectively.[2] After enactment, the mean absolute deviation doubled in two of the districts. Anderson, Kling, and Stith (1999) use a similar approach with data from 140 federal judges. The authors find that, prior to the guidelines, two judges imposed sentences that differed from each other by 16 to 18 percent on average. That estimate declined by 8 to 13 percent after the guidelines were adopted. Other studies using random

[2] The author reports the overall average sentence lengths in bar chart form (1991: 154).

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

assignment to judges have used an alternative measure of interjudge disparity. Rather than comparing the mean sentences for each judge, they consider the proportion of variation in sentences attributable to the judge assigned to the case. Applying this approach, Payne (1997) and Hofer, Blackwell, and Ruback (1999) find mixed results: disparity decreased for some crime categories after guideline enactment and increased for others.

Scholars have also examined the effect of changes to the Federal Sentencing Guidelines after their enactment. Most importantly, the United States Supreme Court rendered all sentencing guidelines effectively advisory in 2004 and 2005 (*Booker v. United States; Blakeley v. Washington*). Two later cases, *Gall v. United States* (2007) and *Kimborough v. United States* (2007), further expanded judicial sentencing discretion. Scott (2010) uses random assignment in three federal districts to show that the mean absolute deviation increased from 4.6 months before *Booker* (or 15 percent of the average sentence length) to 6.2 months after *Booker* (18 percent), and 9.0 months after *Gall/Kimbrough* (26 percent).[3] Yang (2014) reports substantively similar results with data from all federal districts. Several other studies have examined the effect of *Booker* specifically on racial disparities in sentencing, but have reached conflicting results (Fischman and Schanzenbach 2012; Starr and Rehavi 2013; Ulmer, Light, and Kramer 2011; USSC 2010).

Taken together, the random assignment literature suggests that prior to guideline enactment, the mean absolute deviation of judges' sentences were between 10 and 26 percent of the average sentence length (Anderson, Kling, and Stith 1999; Waldfogel 1991). The literature also suggests that guideline enactment and subsequent guideline policy changes may have resulted in changes in the mean absolute deviation that range between 0 and 26 percent (Anderson, Kling, and Stith 1999; Scott 2010).

Prior work has acknowledged several limitations in the random assignment approach. First, as Hofer, Blackwell, and Ruback (1999) note, random assignment captures only the "tip of the iceberg" because it cannot capture disparities arising from sources other than judge assignment. Second, by focusing on the *mean* sentence of judges, the random assignment approach measures only a small fraction of all interjudge disparity. Even if they share the same mean, judges' sentences may have different functional forms or standard deviations. And judges may impose the same sentence on average, without imposing the same sentence in particular cases. Third, random assignment does not ensure that a court receives an equivalent caseload over time. As a result, these

[3] Scott uses a different measure of interjudge disparity, but provides sufficient data to calculate the mean absolute deviation (2010: 61).

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

studies do not rule out the possibility that changes observed after guideline enactment are caused by secular trends in the criminal justice system.

The Identical Case Approach

The identical case approach measures sentencing disparity in *individual cases* rather than in *distributions of comparable cases*. The typical study provides judges with an identical set of real or hypothetical cases and requests a sentence recommendation for each. Unfortunately, no identical case studies in the literature are longitudinal. They provide insight on the magnitude of disparity at a given moment, but cannot estimate the effect of guideline changes over time.

Two early research initiatives (Seminar and Institute 1962; Sentencing Institute and Joint Council 1962) asked federal judges to assign sentences for a diverse set of hypothetical cases. The recommended sentences varied widely. One tax evasion case drew recommendations as lenient as a 6-month suspended prison sentence, and as harsh as a 5-year prison sentence with a $20,000 fine. The recommended sentences for an embezzlement case ranged from probation to 5 years in prison. Similarly, Partridge and Eldridge (1974) conducted a study in which fifty federal district court judges were given complete presentence reports, and were asked to recommend a sentence for each. The authors find evidence of "substantial" disparities in the judges' sentencing recommendations.

Forst and Wellford (1981) distributed hypothetical bank robbery and fraud cases to 264 federal judges. Each hypothetical provided a limited number of facts about the case. Their results present evidence of large, and sometimes, huge disparities in recommended prison length. On average, the bank robbery cases received a sentence length of 8.7 years, and an average standard deviation of 5.2. Similarly, the fraud cases received an average sentence length of 5.2 years and a standard deviation of 3.3. Thus, the interjudge disparity was roughly 60 percent of the average sentence length.

Scholars have also applied the identical case approach to state court judges. Austin and Williams (1977) distributed descriptions of five hypothetical minor felony and misdemeanor offenses to 47 Virginia district court judges. The case descriptions "conveyed . . . defendant's name, the criminal charge and a synopsis of the testimony" (Austin and Williams 1977: 307). The authors find that even "when legal cases are equalized within offense categories, judges still show substantial disparity." (Austin and Williams 1977: 309).

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

Diamond and Zeisel (1975) explore disparity through sentencing councils in New York and Chicago. In these councils, judges sought advice from each other on sentencing decisions for real cases on their dockets. Colleague judges were informed of the facts of each case through presentence reports prepared by the probation department. The authors find that two judges on the same council "differ[ed], on the average, by between one-third and one-half of the mean sentence" (Diamond and Zeisel 1975: 122).

Taken together, the best studies using the identical case approach observe distributions of sentence recommendations with standard deviation that are 30 to 60 percent of the mean sentence length.

Scholars have identified some important limitations on the validity of data gathered through the identical case approach. First, judges may understand the aim of the study and "deviate from their normal sentencing practice" to "dispel an unwanted reputation" (Diamond and Zeisel 1975: 116). The implication is that more extreme judges might recommend sentences closer to the average leading to a conservative estimate of the true magnitude of disparity.

Second, scholars have argued that judges receive less information for hypothetical cases than real criminal cases, and that less information will encourage judges to use their imaginations to fill in the "gaps," and inflate the estimate of disparity. This is a limitation, but it seems just as likely that information gaps would deflate the estimate. As Johnson (2003) has argued, hypothetical fact patterns with few details abstract away controversial case features over which there is little consensus, and which might generate great differences in recommended sentences.[4]

Third, Diamond and Zeisel (1975) note that judges may recommend more severe sentences than in true criminal cases because it is easier to imprison a hypothetical defendant than a living person. These concerns are somewhat addressed by sentencing council data. Sentencing councils provide an optimal window to examine disparity because judges receive detailed presentence reports and "know that their recommendations can and often do have a real impact on the sentence actually imposed

[4] Moreover, most criminal cases are disposed by plea bargain. Thus, judges typically do not benefit from the wealth of information brought out at trial, and instead receive only limited information through a presentence investigation report and a brief sentencing hearing (Johnson 2003). This depth of information is similar to the information judges receive in at least some of the studies that use the identical case approach (e.g., Partridge and Eldridge 1974).

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

. . . [N]o more realistic arrangement can be devised that will allow several judges to sentence one offender" (Diamond and Zeisel 1975: 116).

Conceptual Framework

Blackmun's Thesis asserts that imposing uniformity through sentencing guidelines introduces a bias into sentencing that leads to an overall reduction in fairness. The following section defines terms and concepts to clarify the precise contours of Blackmun's Thesis.

Definitions

The first key concept is that of the *ideal sentence*. The ideal sentence is the best sentence, or the fairest sentence in a particular case. It is the sentence that a defendant would receive in a perfect criminal justice system. Of course, views about the ideal sentence in particular cases will vary from person to person depending on their normative theory of punishment (Rossi, Berk, and Campbell 1997; U.S. Sentencing Commission 1987), and sentencing commissions have faced great theoretical and organizational challenges in choosing between existing theories (Savelsberg 1992). Under a retributive theory of punishment, for example, the purpose of sentencing is to give offenders what they deserve (Tonry 2006). The ideal sentence, then, depends on a limited set of case characteristics related to the severity of the crime and the culpability of the defendant.[5] In contrast, under utilitarian theories of punishment, the purpose of sentencing is to promote social utility. Utilitarian theories often emphasize rehabilitation or specific deterrence. Under these theories of punishment, the ideal sentence may depend on a much larger group of case characteristics that correlate with recidivism, including age, gender, employment, and family.[6]

The current study does not attempt to resolve this longstanding debate about the correct theory of punishment or the case characteristics that are relevant to sentencing. Instead, it attempts to examine Blackmun's Thesis and the relevant tensions between bias and uniformity—to the extent possible—without

[5] Scholars sometimes note that retributive theories of punishment result in greater *equality* in sentencing specifically because retributive theories often recognize relatively few case characteristics as relevant to sentencing.

[6] Scholars sometimes associate rehabilitation-based theories of punishment with greater *individualization* because they recognize a very large number of case characteristics as relevant to sentencing decisions.

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

assuming any particular substantive theory of sentencing.[7] It does so by making assumptions about the numeric value of the ideal sentence that are favorable to Blackmun's Thesis, and by examining the underlying mathematical relationships of the concepts of bias, uniformity and fairness. I leave the definition of the ideal sentence vague to be inclusive of diverse normative views about sentencing.

A *relevant case feature* is a feature of a criminal case that impacts the ideal sentence. The use of a weapon to facilitate a crime, for example, is likely a relevant case feature because it increases the ideal sentence by some length of time.

A sentence is *unfair* to the extent it differs from the ideal sentence. Sentence unfairness is, thus, some function of the difference between the ideal sentence and the actual sentence. An analysis of the first variable in Blackmun's Thesis, *fairness*, thus, involves a comparison of average sentence unfairness in a specific sentencing system before and after guidelines are introduced. The preguideline system is fairer than the postguideline system if it has lower average sentence unfairness.

To analyze *uniformity*, the second variable in Blackmun's Thesis, some concept is needed to identify cases that are similar, and thus, deserve the same treatment. A *set of morally equivalent cases* refers to all criminal cases that share the same set of relevant case features, and as a result, share the same ideal sentence. In a perfect criminal justice system, the judge would impose the same sentence in these equivalent cases. For example, imagine that one particular set of relevant case features is: (1) a robbery, (2) with a gun, (3) that is unloaded, (4) committed by an offender who has two prior felony convictions for aggravated robbery. For this hypothetical, all crimes that involve these four relevant case features, and no other relevant case features, are *morally equivalent*.[8]

Clearly, however, these cases will not all receive the same sentence. Discrepancies between sentences actually received may arise from any number of sources including differences in the ideological beliefs of the judge, the competence or resources of the prosecutor, and other unrelated variables. *Sentencing variation* among morally equivalent cases is measured by the standard deviation of the sentences imposed in those cases. Thus, an

[7] Indeed, Blackmun does not invoke any particular theory of sentencing in relation to Blackmun's Thesis. Throughout his opinion in *Callins v. Collins*, Blackmun often refers to the "fair" sentence in a particular case as the "appropriate" sentence (e.g., 1994: 1149–50). Thus, another way to frame Blackmun's Thesis is: a step toward consistency is a step away from the most appropriate sentence, *however your normative theory of punishment defines it*.

[8] Importantly, a *set of morally equivalent cases* is not defined by the law in any particular jurisdiction. Rather, it is a collection of cases that would be deemed morally equivalent by an idealized or perfect criminal justice system.

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

analysis of the second variable in Blackmun's Thesis, *uniformity*, involves a comparison of the standard deviation of sentences from equivalent cases before and after the introduction of sentencing guidelines.

Defining Blackmun's Thesis

I now frame Blackmun's Thesis in the terminology defined above. The primary goal of sentencing guidelines is to increase uniformity in sentences among similar cases, that is, to decrease their standard deviation. Figure 1 provides an illustration of this process. Recall the unloaded gun robbery hypothetical discussed above. The black line in Figure 1 represents the preguideline sentences for these cases. The gray line represents postguideline sentences—the sentences those cases would have received under a guideline system. Both distributions share the same mean, but the postguideline sentences have a smaller standard deviation because guidelines increase uniformity.

Proponents of Blackmun's Thesis assert that this reduction in the standard deviation is not the only effect of guidelines: imposing greater uniformity also has a *bias effect*. Once again, imagine that the solid black and gray lines in Figure 2 represent the preguideline and postguideline sentences, respectively. Imagine further that the ideal sentence for these hypothetical cases,

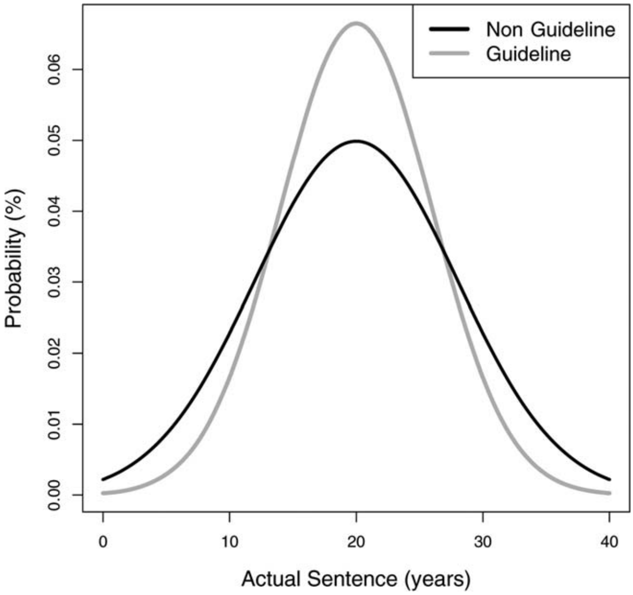

Figure 1. Sentences for Equivalent Cases Before and After Guidelines Decrease Standard Deviation.

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

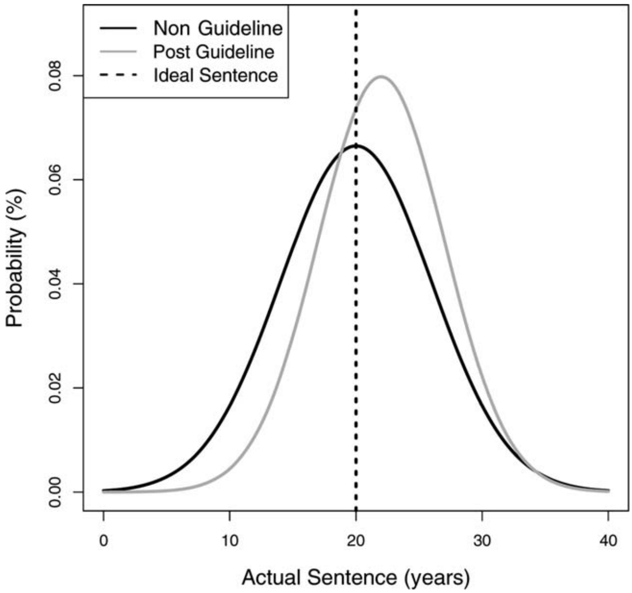

Figure 2. Sentences for Equivalent Cases Before and After Guidelines Increase Bias and Decrease Standard Deviation.

represented by the dotted line, is the mean of the preguideline distribution. As expected, the standard deviation for the postguideline distribution is smaller. But, the mean has also increased, representing a bias relative to the ideal sentence.

There are good reasons to expect that guidelines exert this kind of bias on the mean. Before the guidelines are enacted, judges have more discretion in identifying and weighing a wider range of case features during the sentencing process. The introduction of guidelines can bias the sentences imposed by judges by prohibiting them from considering case features they would have otherwise. To see this more clearly, suppose that preguidelines, most judges mitigate the sentence in our unloaded gun robbery hypothetical by roughly 2 years because the fact that the gun is unloaded indicates diminished culpability. If the legislature adopts a guideline system that recognizes a firearm as an aggravator, but does not distinguish between loaded and unloaded firearms, then judges can no longer mitigate the sentence. The postguideline distribution would, as a result, be biased upward by 2 years. Blackmun's Thesis asserts that, due to this bias effect, sentencing guidelines decrease fairness in individual cases on average.

However, there is a second potential effect of guidelines on fairness not considered by proponents of Blackmun's Thesis.

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

Increasing uniformity in sentences among equivalent cases, on average, directly increases their fairness. Figure 3 depicts three possible distributions of postguideline sentences. Each distribution has a 2-year bias in the mean relative to the ideal sentence but has a different standard deviation. The black line has a standard deviation of 6 and represents sentences with the most unfairness. These sentences are, on average, 5 years away from the ideal sentence. The dark gray line has a standard deviation of 5 and is less unfair. These sentences are, on average, 4 years away from the ideal sentence. The light gray line has a standard deviation of 4 and is the fairest. These sentences are, on average, 3.5 years away from the ideal sentence. The implication is that increasing uniformity diminishes the average distance to the ideal sentence, and thus, increases fairness.[9] Blackmun's Thesis does not hold when the negative effects of *bias* are outweighed by the positive *variance* effects of uniformity. The remainder of this article explores the potential relative sizes of these effects subject to a wide range of possible conditions.

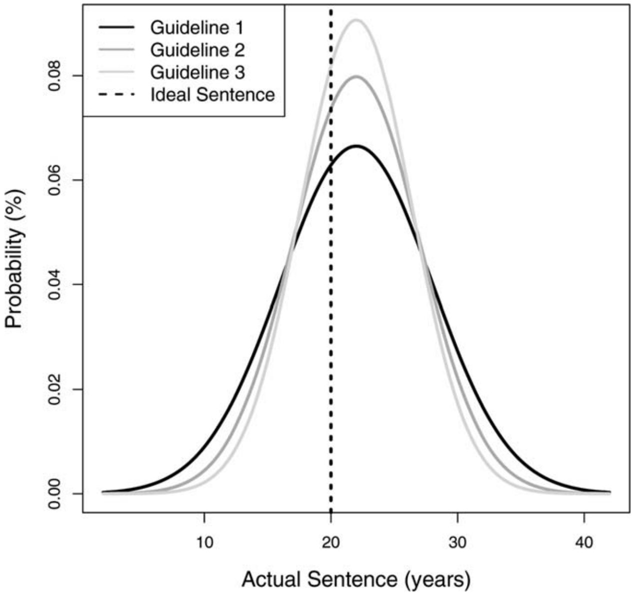

Figure 3. Sentences for Equivalent Cases with Equal Bias and Varying Levels of Uniformity.

[9] Assuming that sentences are symmetrically distributed, increasing uniformity will increase fairness as long as there is at least one sentence that falls on both sides of the ideal sentence. The relationship between uniformity and fairness will hold as long as judges have not completely missed the mark.

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

Before proceeding, two additional clarifying points are in order. First, the tradeoff between bias and variance is not limited to the context of sentencing. It is similar, for example, to related tensions in the selection of statistical estimators. Like the proponents of Blackmun's Thesis, social scientists often emphasize bias in the selection of statistical estimators. Some scholars have noted, however, that overemphasizing bias may lead to unbiased but inefficient estimators that are "worse" than more efficient but biased estimators (Lynch and Western 2004; Stolzenberg and Relles 1997). This article raises a related argument in the context of sentencing where scholars have not appreciated the importance of variance for fairness.

Second, in addition to criticizing the federal sentencing guidelines for promoting uniformity at the expense of fairness, scholars also commonly criticize the guidelines for prescribing sentences that are, across the board, too severe (Stith and Cabranes 1998). This critique is distinct from Blackmun's Thesis. An increase in severity is not a necessary consequence of sentencing guidelines. Indeed, scholars have noted that many state guidelines successfully maintained historical sentencing averages (Tonry 1993). Moreover, the Federal Sentencing Commission deliberately increased sentences due to a perception that the federal courts were insufficiently punitive (Stith and Cabranes 1998). The severity of the federal system is not due to tension between uniformity and fairness in guideline systems, but rather, to a separate policy decision to increase severity through guidelines and mandatory minimums.

Analytic Method and Design

An analysis of Blackmun's Thesis involves four key parameters: (1) the magnitude of sentencing disparity among an equivalent set of cases prior to guideline enactment, (2) the effect of the guidelines on disparity, (3) the bias effect of the guidelines (i.e., effect on the average sentence length), and (4) the ideal sentence. A simulation modeling approach does not require perfect knowledge of these parameters. Indeed, they would no doubt vary by jurisdiction, time, and crime category anyway. Instead, I look to the empirical literature for guidance on plausible bounds on the parameters, and then test for all values between these bounds.

First, I rely on estimates from the identical case approach for guidance on the magnitude of preguideline disparity. Diamond and Zeisel (1975) and Forst and Wellford (1981) observed sentence recommendations with standard deviations ranging from 30 to 64 percent of the mean sentence length. In the preliminary

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

analysis below, I assume the preguideline disparity is 25, 50, or 75 percent of the average sentence length. A subsequent analysis then relaxes this assumption by testing a far greater range of values including 0.

Second, I rely on the random assignment literature for guidance on the effect of sentencing guidelines on disparity because no identical case studies in the literature were conducted longitudinally. The random assignment literature observes postenactment changes in the mean absolute deviation of judges' sentences that range from 0 to 26 percent (Anderson, Kling, and Stith 1999; Scott 2010; Waldfogel 1991). I, therefore, test for effects that range from 0 to 30 percent.

Third, the empirical literature provides no guidance on the magnitude of the bias effect of sentencing guidelines. To be conservative, I assume wide bounds for bias effects that range from 0 to 50 percent of the average preguideline sentence length.

Finally, it is necessary to operationalize the *ideal sentence* for a set of morally equivalent cases. There is no nonarbitrary method to peg the value of the *ideal sentence*. Underlying Blackmun's Thesis, however, is the assumption that the preguideline average sentence for equivalent cases is closer to the ideal sentence than the postguideline average. I, therefore, make the charitable assumption that preguidelines, judges on average get it exactly right: I assume that the *ideal sentence* equals the mean of the preguideline distribution. Unfairness in both the preguideline and postguideline sentences is measured relative to that benchmark.

A few further notes about this definition of the ideal sentence are in order. First, this specification favors Blackmun's Thesis, and leads to a *conservative* test. The mean of the preguideline distribution is the value that minimizes unfairness for that distribution.[10] Second, perceiving that historical sentences were insufficiently severe (or overly severe) for particular offenses, some sentencing commissions have aimed not only to decrease the standard deviation of sentences, but also to increase (or decrease) the mean. This reflects a belief that the ideal sentence is higher (or lower) than historical practice. By assuming the ideal sentence is the mean of the preguideline distribution, I cannot account for this kind of "prescriptive" change (Tonry 1993). But, if the sentencing commission is correct that the ideal sentence is higher (or lower) than historical practice, then the test of Blackmun's Thesis is once again rendered more *conservative* because

[10] It is not unreasonable to expect that the true ideal sentence is somewhere between the mean of the preguideline and postguideline distributions. If so, the argument against Blackmun's Thesis is even stronger than the evidence presented in this article.

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

the true ideal sentence will be closer to the postguideline distribution than the ideal sentence assumed in the analysis.

The Design: Statistical Simulation and Computation

This study uses statistical simulation to explore the relationship between uniformity and fairness proposed by Blackmun's Thesis.[11] I begin by examining this relationship for a particular preguideline distribution, and later relax this assumption by exploring a wider range.

I begin with a preguideline distribution of sentences with a mean of eight and a standard deviation of 6. This ratio between mean and standard deviation is relatively large, but it is consistent with findings in the empirical literature. Some cases are likely subject to wide sentencing variation, particularly those involving multiple characteristics over which little consensus exists on their relevance to sentencing (e.g., young age, psychological disorders, history of family abuse, child dependents, stable employment, positive contributions to the community). The first step is to estimate unfairness in this preguideline distribution. A random sample of 10,000 elements is generated from a normal distribution with a mean of 8 and standard deviation of 6.[12] Each of the 10,000 elements represents a sentence from one of 10,000 hypothetical morally equivalent criminal cases. Because a sentence cannot take on a negative value, any of the 10,000 elements that are less than zero are set to zero.[13] Next, to calculate sentence unfairness, the absolute value of the difference between each of

[11] The methodology can be understood as a kind of Monte Carlo simulation that approximates the integral of the function of sentence unfairness (i.e., actual sentence minus some constant representing the ideal sentence) for a distribution of morally equivalent cases that is multiplied by X.

[12] To address instability in some estimates, the final analysis actually draws 1,000 samples with 10,000 elements in each. This translates into an effective "sample" of ten million sentences per distribution.

[13] This deviation from the normal distribution is of only minor concern. Relatively few negative sentences were actually generated. Sensitivity analyses reveal that resetting those negative values to zero has little substantive effect on the findings of the study. Parallel analyses were conducted with precisely the same baseline values, but using means that were far enough from zero that no negative values were randomly generated. Little substantive change to the findings of the study were observed. Moreover, setting negative sentences to zero favors Blackmun's Thesis, and thus renders the test more conservative. Since the ideal sentence is always equal to or greater than zero, negative values are always further from the ideal sentence than zero. Thus, resetting negative values to zero increases the fairness of a distribution. As the bias in the mean of the postguideline distribution is increased, fewer negative values are randomly generated. In contrast, the proportion of values set to zero in the preguideline distribution remains constant. The decreasing number of randomly generated negative values (reset to zero) in the postguideline distribution decreases the fairness of the postguideline distribution relative to the preguideline distribution, thereby increasing the difficulty of disproving Blackmun's Thesis.

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

the 10,000 sentences and the ideal sentence is computed. In this case, the mean of the preguideline distribution is eight, and so the ideal sentence is also eight. A sentence of 5 years, is therefore, associated with unfairness of $|5-8|=3$ years. Finally, a summary of total sentence unfairness is calculated by computing the mean unfairness among the 10,000 sentences.

Calculating sentence unfairness for postguideline sentences is more complicated because sentencing guidelines can have a wide range of possible effect sizes on uniformity and bias. Thus, *all* plausible effect sizes in one-percent increments are estimated. First, I estimate the effect on fairness of a guideline system that exerts no change in the standard deviation or the mean. Ten thousand sentences are randomly generated from a normal distribution with the same parameters as the preguideline distribution (i.e., mean of 8; standard deviation of 6). Of course, aside from small differences due to random chance, this postguideline distribution will share the same level of unfairness as the preguideline distribution. Next, the effect on unfairness of a guideline system that biases the mean by 1 percent and exerts no change in the standard deviation is tested. Ten thousand sentences are randomly generated from a normal distribution with a mean of $8+(0.01*8)=8.08$ and a standard deviation of 6. Once again, mean unfairness relative to the ideal sentence is computed. Next, the effect on fairness of a guideline system that biases the mean by 2 percent and exerts no change on the standard deviation is tested. Ten thousand sentences are randomly generated from a normal distribution with a mean of $8 + 0.02*8 = 8.16$ and standard deviation of 6, and mean unfairness is computed. This process continues by one-percent increments until the mean is 50 percent higher than its baseline value ($0.50*8 = 12$). Next, I perform precisely the same procedure but with a standard deviation that is 1 percent lower than the preguideline standard deviation ($6-6*0.01 = 5.94$), and so on. This iterative procedure is repeated until mean unfairness is computed for a postguideline distribution with a mean that is 50 percent larger than the baseline mean of 8 (i.e., 12), and a standard deviation that is 35 percent smaller than the baseline standard deviation of 6 (3.9).

At this point, we have estimates of the average unfairness resulting from a wide range of effect sizes on the mean and standard deviation of a preguideline distribution with a mean of eight and a standard deviation of 6. For all postguideline distributions that have the same or less average unfairness than the preguideline distribution, Blackmun's Thesis does not hold. For all postguideline distributions that have higher average unfairness than the preguideline distribution, Blackmun's Thesis holds.

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

To explore how these effect sizes vary, I provide similar results for two other preguideline distributions with a smaller standard deviation of 4 and 2. But, we are not only interested in Blackmun's Thesis for three particular preguideline distributions. If we assume that preguidelines, a distribution of morally equivalent cases can have a mean between 0 and 60 years, and a standard deviation between 0 and 40, then there are 2,400 possible preguideline distributions. I perform the same analysis for all 2,400 pairwise combinations.

Assumptions of the Design

Having outlined the basic design of the study, I take this opportunity to make explicit several important methodological assumptions. The first assumption is that sentences for morally equivalent cases are distributed normally. The validity of this assumption will no doubt vary by context. Equivalent cases with average sentences that are close to zero, for example, cannot be normally distributed due to left-side censoring. As another example, it is possible that guidelines do not merely reduce the standard deviation of sentences but also change the functional form. The normal distribution was selected, however, for a number of desirable properties.

First, unlike other probability distributions, the parameters of the normal distribution are defined by a mean and standard deviation. Thus, the parameters of the normal distribution map well onto the parameters of Blackmun's Thesis, and allow for their direct manipulation.

Second, the normal distribution has relatively thin tails, meaning it generates few outliers that dominate the results. This feature is perhaps more important than the exact functional form. Indeed, the results of the analysis are not intended to be exactly right. Rather, they are intended to shed light on general trends in the tension between uniformity and fairness. Similar analyses can, at least in principle, be conducted for a wider range of distributional forms.

Third, normality is empirically plausible. The identical case approach likely provides the most useful evidence on this point. Yet, studies in the literature are over 30 years old, and thus, little data is still available. Fortunately, Partridge and Eldridge (1974) provide sufficient information to reconstruct some of the data from their study. The authors provided complete sentencing reports to 50 federal judges in the 1970s prior to the enactment of the federal sentencing guidelines. The left panel of Figure 4 depicts the sentences recommended for a male defendant with five prior convictions, two periods of incarceration and no history

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

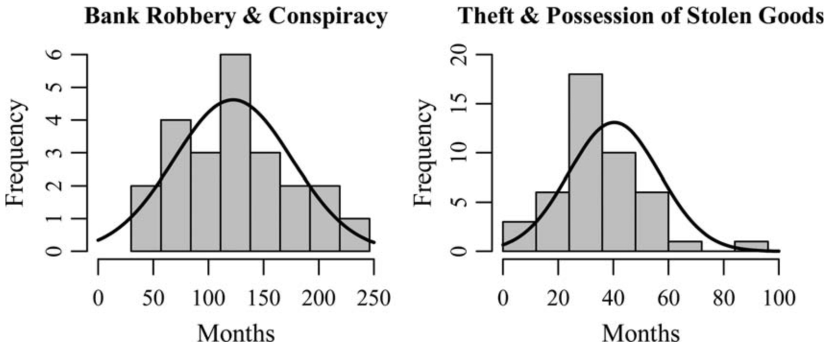

Figure 4. Histograms of Sentences from Equivalent Cases.[14]

of drug use, and who was convicted at trial of four counts of bank robbery and conspiracy. While the sample size is small ($n = 24$), the distribution appears to approximate normality. The right panel of Figure 4 depicts sentences recommended for a male defendant who was convicted at trial of two counts of theft and possession of goods from an interstate shipment, and who had no prior convictions, but did have other felony charges pending for acts committed after the instant offense. With a somewhat larger sample size ($n = 45$), the distribution also appears to approximate the normal distribution. Together, these distributions provide evidence that the assumption of normality is plausible for at least some sets of equivalent cases prior to the enactment of guidelines.

The second main assumption of the analysis is the conception of an ideal sentence. That conception involves three key components. First, as noted earlier, the design presupposes that every crime has an ideal sentence. Moral skeptics may wonder about this ontology, but Blackmun's Thesis itself invokes a comparison between the fairness of sentences under discretionary and guideline regimes. Such a comparison presupposes some ideal sentence against which to compare. Second, the design assumes that the ideal sentence is a discrete point value (e.g., 9 years exactly), rather than a range of values (e.g., 8 to 10 years). Third, the value of the ideal sentence is the mean of the preguideline distribution. As noted earlier, this is a charitable assumption that favors Blackmun's Thesis, and renders the test more conservative.

Finally, we need to draw some assumptions for calculating the unfairness of deviations from the ideal sentence. My analysis begins with the assumption that unfairness is both linear and

[14] The data derive from Partridge and Eldridge (1974).

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

symmetric. If so, a 2-year or 3-year deviation from the ideal sentence is twice or three times more unfair than a 1-year deviation. This assumption appears plausible particularly when examining smaller margins such as a 10 or 20 percent change in bias. In a subsequent analysis, I relax the assumptions of linearity and symmetry. Given the difficulty of examining an unlimited number of possible nonlinear and asymmetric functions, I explore several illustrative examples. First, I relax nonlinearity by modeling fairness as (X^2), $(X^2)/3$ and $(X^2)/5$, where X represents the distance of a sentence from the ideal sentence. As I discuss in greater detail below, relaxing the assumption of linearity has only minor effects on the results of the analysis. Second, I also relax the assumption of symmetry by applying a different functional form for sentences below the ideal sentence than for sentences above. It is not possible to test all possible functional forms,[15] but these additional analyses provide added insight on the behavior of Blackmun's Thesis if unfairness increases at different rates above and below the ideal sentence.

Results and Analysis

I present my results in three stages. I begin by presenting what I refer to as *conditional margin graphs*. These graphs present the range of possible effects of sentencing guidelines conditional on a specific preguideline distribution. They provide granular information on whether Blackmun's Thesis holds subject to a wide range of plausible effects on uniformity (standard deviation) and bias (mean). I then present additional results in *unconditional margin graphs* (Figures 7 and 8), which summarize the conditional margin graphs for all 2,400 possible preguideline distributions that have means ranging from 0 to 60, and standard deviations ranging from 0 to 40. Unsurprisingly, Blackmun's Thesis behaves predictably across different preguideline distributions. Finally, I illustrate how the results of the analysis are affected when we relax the assumptions of linearity and symmetry for fairness (Tables 1–3).

[15] For example, one blind reviewer noted that the pain or discomfort of an additional year of prison likely diminishes over a prisoner's tenure behind bars. As a result, a pure retributivist might believe that if a sentence over the ideal is lengthened, the marginal impact of each additional year on unfairness becomes smaller, and that as a sentence under the ideal is shortened, the marginal impact of each additional year becomes greater. The exponential functions of fairness I implement can account for the latter, but cannot account for the former. I merely note that such a nonlinear function of fairness would favor Blackmun's Thesis, and thus, render my test more conservative. As the mean of the postguideline distribution increases, the fairness of a larger proportion of sentences in the distribution would be computed based on the above-ideal-sentence function.

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

Conditional Margin Graph Analysis

Figure 5 is a *conditional margin graph* for a preguideline distribution with a mean of eight and a standard deviation of 6. Each dot in the figure represents a comparison of fairness between the preguideline distribution and a postguideline distribution with a percentage change in the mean and standard deviation. The X-axis represents a percent change in the mean relative to the value of the preguideline mean of eight, and the Y-axis represents a percent change in the standard deviation relative to the value of the preguideline standard deviation of 6. In other words, the point at 0.10 on the X-axis and −0.10 on the Y-axis represents a comparison between the preguideline distribution (mean = 8; SD = 6), and a postguideline distribution on which sentencing guidelines have exerted a 10 percent increase in the mean and a 10 percent decrease in standard deviation (mean = 8.8; SD = 5.4). A gray dot signifies that the guidelines have decreased fairness, and thus, that Blackmun's Thesis holds. A black dot signifies that the guidelines have maintained or increased fairness overall, and that Blackmun's Thesis does not hold. Accordingly, the black dot at the point (0.1, −0.1) indicates that a guideline system that decreases the standard

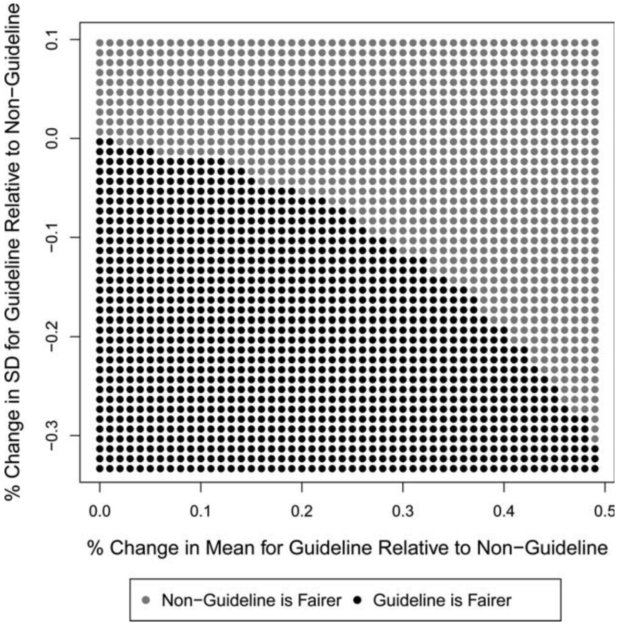

Figure 5. Conditional Margin Graph, Mean = 8, SD = 6.
Fairness of Preguideline and Postguideline Sentences Subject to Percentage Shifts in Mean and Standard Deviation of Postguideline Sentences.

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

deviation of the preguideline distribution by 10 percent will increase or maintain fairness even if the guidelines add 10 percent bias.

The implications of Figure 5 are rather striking. Point (0.28, −0.1) reveals that, for a preguideline distribution with mean of 8 and standard deviation of 6, a sentencing system that decreases the standard deviation by 10 percent and causes a 28 percent bias in the mean will increase or maintain the preguideline level of fairness. At this cut point, standard deviation has a positive effect on fairness that is almost three times larger than the negative effect of bias. Similarly, point (0.4, −0.2) reveals that, a sentencing system that decreases the standard deviation by 20 percent and causes a 40 percent bias in the mean will increase or maintain fairness overall. At this cut point, standard deviation has a positive effect on fairness that is almost twice as large as the negative effect of bias in the mean. This is evidence against Blackmun's Thesis, as it suggests that decreases in uniformity can play a larger role in fairness than bias: even where guidelines increase the severity of an entire distribution of cases, small reductions in standard deviation may nonetheless deliver an overall increase in fairness.

Figure 6 is a conditional margin graph for a preguideline distribution with a mean of eight and a standard deviation of 4. The implications of Figure 6 are similar. Point (0.2, −0.1) reveals that, for this distribution, a guideline system that decreases the standard deviation by 10 percent and causes a 20 percent bias in the mean will nonetheless increase or maintain the preguideline level of fairness. At this cut point, standard deviation has a positive effect on fairness that is twice as large as the negative effect of bias. Similarly, point (0.28, −0.2) reveals that a sentencing system that decreases the standard deviation by 20 percent and causes a 28 percent bias in the mean will increase or maintain fairness. At this cut point, standard deviation has a positive effect on fairness that is almost 50 percent larger than the negative effect of bias in the mean.

Figure 7 is a conditional margin graph for a preguideline distribution with a mean of 8 and standard deviation of 2. Even where there is little sentencing variation to begin with, the story is similar: decreasing the standard deviation continues to have a substantial effect on fairness. Point (0.1, −0.1) reveals that, for this preguideline distribution, a sentencing system that decreases the standard deviation by 10 percent and causes a 10 percent bias in the mean will increase or maintain the preguideline level of fairness. At this cut point, standard deviation has a positive effect on fairness that is equal to the negative effect of bias. Point (.15, −0.2) reveals that a guideline system that decreases the

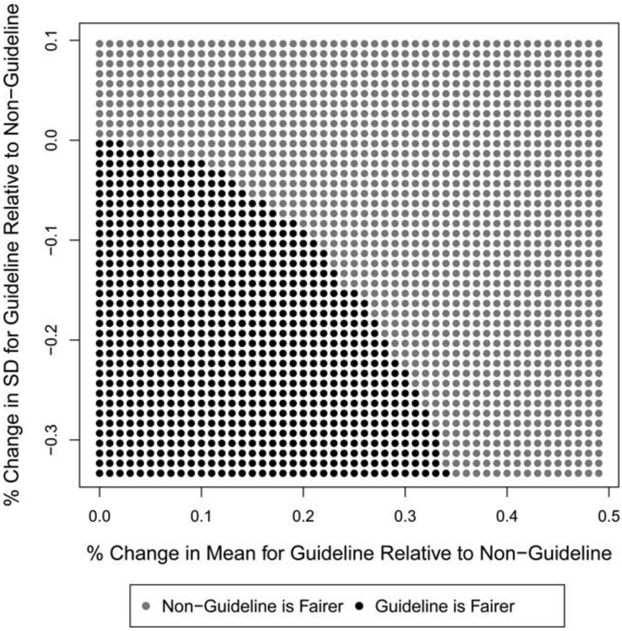

Figure 6. Conditional Margin Graph, Mean = 8, SD = 4. Fairness of Preguideline and Postguideline Sentences Subject To Percentage Shifts in Mean and Standard Deviation of PostGuideline Sentences.

standard deviation by 20 percent will increase or maintain fairness overall, even if it also causes as large as a 15 percent bias in the mean. At this cut point, standard deviation has a positive effect on fairness that is 75 percent the size of the negative effect of bias.

One clear generalization across Figures 5 through 7 is that small reductions in the standard deviation can have a large effect on fairness when compared to the effects of bias. A second generalization is that reductions in standard deviation have diminishing marginal returns: the first 10 percent reduction in standard deviation has a larger effect on fairness than a second 10 percent reduction.

Unconditional Margin Graph Analysis

Conditional margin graphs are "conditional" because they represent the relationship between uniformity, bias and fairness conditioned on a preguideline distribution with a specific mean and standard deviation. All possible 2,400 conditional margin graphs are summarized in two *unconditional* margin graphs, Figures 8 and 9.

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

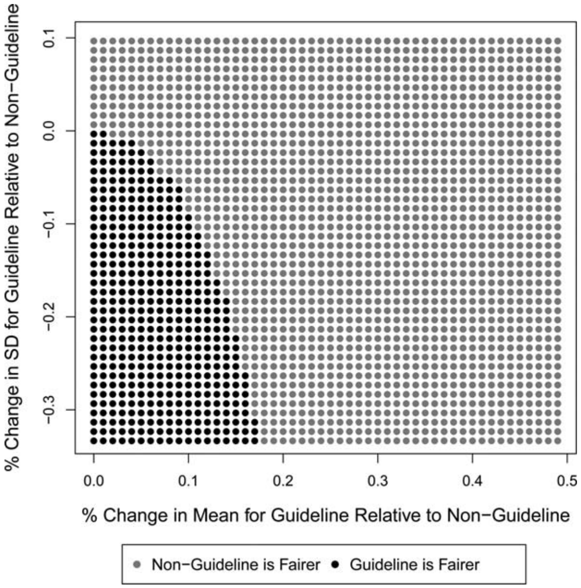

Figure 7. Conditional Margin Graph, Mean = 8, SD = 2. Fairness of Preguideline and Postguideline Sentences Subject To Specified Percentage Shifts in Mean and Standard Deviation of PostGuideline Sentences.

Figure 8 is an unconditional margin graph that answers the question, does standard deviation or bias in the mean have a greater impact on fairness? Each of the 2,400 points in the figure represent a summary of the key findings from the possible preguideline distributions. For example, the point at 8 on the X-axis, and 6 on the Y-axis summarizes the findings of the conditional margin graph for a preguideline distribution with a mean of 8 and a standard deviation of 6. A white point indicates that a guideline system that decreases standard deviation by 10 percent and increases the mean by 10 percent would decrease fairness overall. Thus, the white point at (8, 1) indicates that, for a preguideline distribution with a mean of eight and a standard deviation of 1, a guideline system that decreases the standard deviation by 10 percent and causes a 10 percent bias in the mean will decrease fairness overall. A light gray point indicates that a guideline system that decreases the standard deviation by 10 percent and increases the mean by 10 percent would increase or maintain fairness overall. Thus, the light gray point at (8, 2) indicates that, for a preguideline distribution with a mean of 8 and standard deviation of 2, a guideline system that decreases the standard deviation by 10 percent and causes a 10 percent bias in the mean will increase or maintain the preguideline level of fairness. A dark gray point indicates that a guideline system

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

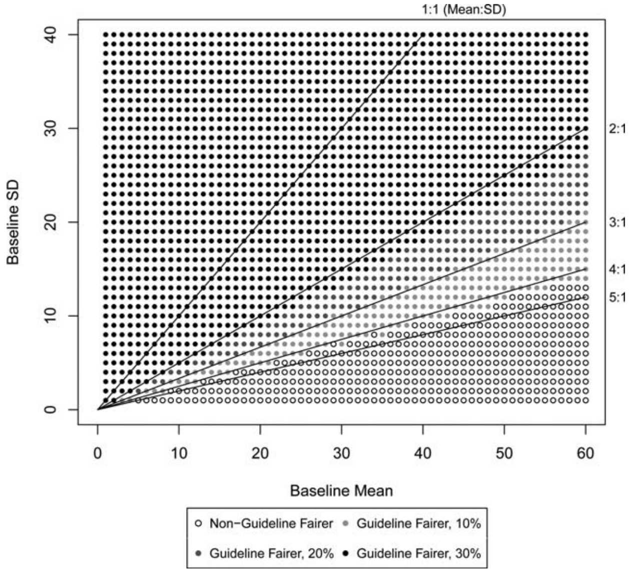

Figure 8. Unconditional Margin Graph: Fairness of Preguideline and Postguideline Sentences for Percentage Change in Mean and Standard Deviation of 10%, 20%, and 30%.

which decreases the standard deviation by 20 percent and increases the mean by 20 percent would increase or maintain fairness. Thus, the dark gray point at (8, 3) indicates that, for a preguideline distribution with a mean of eight and a standard deviation of three, a guideline system that decreases the standard deviation by 20 percent and causes a 20 percent bias in the mean will increase or maintain fairness. Of course, since reductions in standard deviation have a diminishing marginal effect on fairness, all dark gray dots also satisfy the conditions for a light gray dot. Finally, a black point indicates that a guideline system that decreases the standard deviation by 30 percent and increases the mean by 30 percent would increase or maintain the preguideline level of fairness. Point (8, 5) is one such example. Once again, because of diminishing marginal returns, a black dot satisfies the conditions for a light gray dot and a dark gray dot.

The straight lines labeled by ratio in Figure 8 demonstrate that conditional margin graphs present consistent and predictable behavior. The line labeled, '1:1,' illustrates that any conditional margin graph associated with equal preguideline parameters (e.g., mean of 5, standard deviation of 5) is associated with a black dot. In other words, wherever the preguideline distribution has a mean and standard deviation of equal values, a 30 percent decrease in standard deviation and 30 percent bias in mean will increase or maintain the fairness of the distribution. Similarly,

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

conditional margin graphs associated with preguideline values with a ratio of 3:1 (e.g., mean of 6, standard deviation of 2) are associated with a light gray dot. Wherever the preguideline distribution has a mean and standard deviation in a ratio of 3:1, a 10 percent decrease in standard deviation and 10 percent bias in the mean will increase or maintain fairness. Finally, all conditional margin graphs associated with preguideline values with a ratio of 5:1 (e.g., mean of 10, standard deviation of 2) are white. Thus, wherever the preguideline distribution has a mean and standard deviation in a ratio of 5:1, a 10 percent decrease in standard deviation and 10 percent bias in mean will decrease fairness overall.

Unlike Figure 8, which depicts whether standard deviation or bias has a greater impact on fairness, Figure 9 depicts how much greater an impact standard deviation has on fairness. More specifically, Figure 9 indicates the level of bias needed to counteract the positive effect of a 10 percent decrease in standard deviation. A white point indicates that a 10 percent decrease in standard deviation has a smaller effect on fairness than a 10 percent bias in the mean. The point at (8, 1), for example, indicates that for a preguideline distribution with a mean of eight and a standard deviation of one, a guideline system that decreases the standard deviation by 10 percent and biases the mean by 10 percent decreases fairness overall. A light gray point indicates that a 10 percent decrease in standard deviation has an equal or larger effect on fairness than a 10 percent bias in the mean. The point at (8, 2), for example, indicates that for a preguideline distribution with a mean of eight and a standard deviation of two, a guideline system that decreases the standard deviation by 10 percent and biases the mean by 10 percent increases or maintains the level of fairness overall. A dark gray point indicates that a 10 percent decrease in standard deviation has an equal or larger effect on fairness than a 20 percent bias in the mean. Accordingly, dark gray points indicate that standard deviation has an effect on fairness that is over twice as powerful as bias. Finally, a black point indicates that a 10 percent decrease in standard deviation has an equal or larger effect on fairness than a 30 percent bias in the mean. Accordingly, black points indicate that standard deviation has an effect on fairness that is over three times as powerful as bias in the mean.

As in Figure 8, the findings are consistent within ratios. All preguideline distributions with equal mean and standard deviation values are associated with black dots, meaning that a 10 percent decrease in standard deviation has an effect on fairness that is equal to or greater than a shift in the mean that is three times larger. All preguideline distributions with mean and standard

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

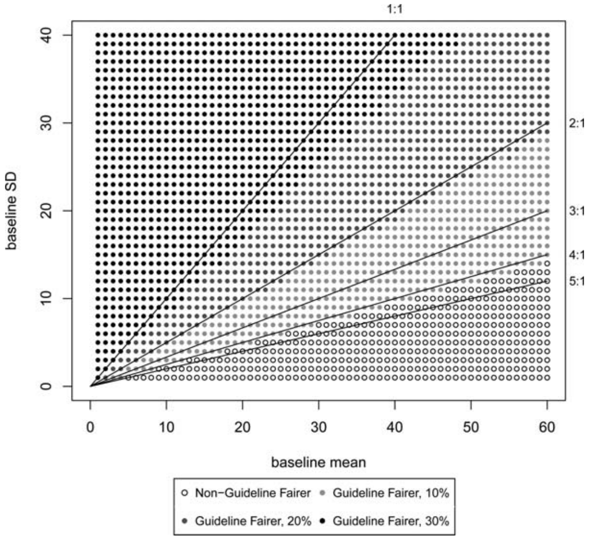

Figure 9. Unconditional Margin Graph: Fairness of Preguideline and Postguideline Distributions for 10% Decrease in Standard Deviation and 10%, 20%, and 30% Increase in the Mean.

deviation values with a ratio of 2:1 are associated with dark gray dots, meaning that a 10 percent decrease in standard deviation has an effect on fairness that is equal to or greater than a bias effect two times larger. All preguideline distributions with mean and standard deviation values with a ratio of 4:1 are associated with light gray dots, meaning that a 10 percent decrease in standard deviation has an equal or greater effect on fairness than a comparable bias effect. All preguideline distributions with mean and standard deviation values with a ratio of 5:1 are associated with white dots, meaning that a 10 percent decrease in standard deviation has a weaker effect on fairness than a comparable bias effect.

Nonlinear Fairness

I began with the assumption that unfairness is both linear and symmetric with respect to the ideal sentence. Here, I relax these assumptions by exploring several alternative functions of unfairness. I relax the linearity assumption by examining the behavior of Blackmun's Thesis if unfairness follows an exponential form of (X^2), (X^2)/3 or (X^2)/5 where X represents the difference between a sentence and the ideal sentence. I then relax the symmetry assumption by allowing for different functional forms for sentences above and below the ideal sentence.

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

Table 1 presents the results for a preguideline distribution with a mean of eight and a standard deviation of six. The top panel shows the amount of bias necessary to equal the positive effects of a 10 percent decrease in the standard deviation after the enactment of guidelines. The bottom panel shows the amount of bias necessary to equal the positive effects of a 20 percent decrease in standard deviation. The rows of the tables indicate the functional form of unfairness for sentences that are greater than the ideal sentence, and the columns indicate the functional form of unfairness for sentences that are less than the ideal sentence.

The main diagonal presents the results when the assumption of linearity is relaxed while maintaining symmetry. Perhaps unsurprisingly, relaxing linearity while maintaining symmetry has little impact on the results of the analysis. For example, Table 1 shows that a 10 percent decrease in standard deviation for a preguideline distribution with a mean of eight and a standard deviation of six is equivalent to a 28 percent increase in bias if fairness is linear. The results are relatively stable if fairness is modeled using a different functional form (24 percent, 25 percent, and 25 percent).

The results change more substantially when the symmetry assumption is also relaxed. The first column of Table 1 provides the results when the fairness of sentences below the ideal is modeled linearly, but the functional form of unfairness above the ideal is varied. Here we can observe the relative magnitude of the variance and bias effects when the unfairness of sentences increases more rapidly for sentences above the ideal sentence.

Table 1. Percent Increase in Bias Equivalent to 10% and 20% Decrease in Standard Deviation—Mean = 8, SD = 6

		10 Percent Reduction in SD			
		Under Ideal			
		Linear	$X^2/5$	$X^2/3$	X^2
Over Ideal	Linear	28%	43%	77%	190%
	$X^2/5$	18%	24%	39%	92%
	$X^2/3$	13%	16%	25%	63%
	X^2	11%	11%	13%	25%

		20 Percent Reduction in SD			
		Under Ideal			
		Linear	$X^2/5$	$X^2/3$	X^2
Over Ideal	Linear	40%	60%	84%	210%
	$X^2/5$	31%	37%	56%	99%
	$X^2/3$	25%	30%	37%	75%
	X^2	21%	21%	24%	38%

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

Table 2. Percent Increase in Bias Equivalent to 10% and 20% Decrease in Standard Deviation—Mean = 8, SD = 4

10 Percent Reduction in SD					
		Under Ideal			
		Linear	$X^2/5$	$X^2/3$	X^2
Over Ideal	Linear	20%	23%	42%	113%
	$X^2/5$	18%	19%	32%	69%
	$X^2/3$	12%	13%	20%	49%
	X^2	8%	8%	10%	20%

20 Percent Reduction in SD					
		Under Ideal			
		Linear	$X^2/5$	$X^2/3$	X^2
Over Ideal	Linear	28%	31%	46%	115%
	Linear	28%	31%	46%	115%
	$X^2/5$	26%	30%	38%	72%
	$X^2/3$	21%	21%	28%	56%
	X^2	13%	15%	18%	28%

When the fairness of sentences over the ideal sentence follows the exponential function $(X^2)/5$ rather than a linear function, the impact of the standard deviation is smaller. As Table 1 reveals, for a preguideline distribution with a mean of eight and a standard deviation of six, a 10 percent reduction in standard deviation has an equivalent effect on fairness of an 18 percent increase in bias. And a 20 percent reduction in standard deviation has an equivalent effect on fairness of a 31 percent increase in bias. The variance effect is smaller when the fairness of sentences above the ideal is modeled as $(X^2)/3$. A 10 percent reduction in standard

Table 3. Percent Increase in Bias Equivalent to 10% and 20% Decrease in Standard Deviation—Mean = 8, SD = 2

	10 Percent Reduction in SD				
		Under Ideal			
		Linear	$X^2/5$	$X^2/3$	X^2
Over all	Linear	10%	7%	10%	33%
	$X^2/5$	18%	10%	17%	36%
	$X^2/3$	10%	7%	11%	26%
	X^2	5%	5%	5%	11%

	20 Percent Reduction in SD				
		Under Ideal			
		Linear	$X^2/5$	$X^2/3$	X^2
Over Ideal	Linear	15%	9%	13%	34%
	$X^2/5$	21%	14%	20%	37%
	$X^2/3$	16%	11%	14%	28%
	X^2	10%	8%	9%	14%

deviation is equivalent to a 13 percent increase in bias, and a 20 percent reduction in standard deviation is equivalent to a 25 percent increase in bias. Again, the effect is smaller when the fairness of sentences above the ideal sentence is modeled as (X^2). Here, a 10 percent decrease in standard deviation is equivalent to an 11 percent decrease in bias, and a 20 percent decrease in standard deviation is equivalent to a 21 percent increase in bias.

A similar pattern is observed in Tables 2 and 3, which present the results for a preguideline distribution with a mean of 8 and a standard deviation of 4 and 2, respectively. Consistent with earlier results, the smallest variance effect is present in Table 3 for a preguideline distribution with a standard deviation of 2. Under the assumptions most favorable to Blackmun's Thesis—where unfairness for sentences above the ideal is modeled as (X^2) and the fairness of sentences below the ideal is modeled as linear—a 10 percent and 20 percent reduction in the standard deviation is equivalent to a 5 percent and 10 percent increase in the bias. To summarize, while the variance effect decreases under these assumptions of asymmetry and nonlinearity, it is remarkable that the effects remain substantial in size given how much more rapidly unfairness is assumed to increase for sentences above the ideal (X^2) than for sentences below (linear).[16]

The reverse pattern is observed for the relative magnitude of the variance and bias effects if the functional form for the fairness of sentences *above* the ideal sentence is held constant, while the form for the fairness of sentences *below* is varied. The first row in each table provides the results when the fairness of sentences above the ideal sentence is modeled linearly, but the functional form of unfairness below the ideal is varied. Across most rows in Tables 1 throught 3, the magnitude of the variance effect increases substantially.[17]

In summary, fairness may not follow a linear or symmetric pattern above and below the ideal sentence. It is impossible, however, to examine all plausible nonlinear functional forms of fairness. This section has attempted to illustrate how the primary results of the analysis may be affected when the assumptions of linearity and symmetry are relaxed. The results provide little evidence that relaxing the linearity assumption has a substantial

[16] For example, a sentence of 13 years and a sentence of 3 years are both 5 years away from the ideal sentence of 8 years. If we assume that the fairness of sentences below the ideal follow a linear function, and the fairness of sentences above the ideal follow the exponential function X^2 then a 3-year sentence has an unfairness of 5 while a 13-year sentence has an unfairness of 25.

[17] The few exceptions are in Table 3 due to the substantial proportion of sentences that have a numeric value of unfairness that is between -1 and 1. In this small range of values, the exponential function is less steep than the linear function.

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

effect on the results unless the symmetry assumption is also relaxed. Under the assumption of asymmetry, the magnitude of the variance effect relative to the bias effect may shrink but it remains substantial, and substantively important. This is true even if the functional definition of fairness strongly favors Blackmun's Thesis, as it does when the fairness of sentences above the ideal is modeled as X^2, and the fairness of sentences below the ideal is modeled as linear.

Discussion

This article set out to critically examine Blackmun's Thesis, a widely expressed view among legal scholars and practitioners that increasing uniformity in sentencing through guidelines decreases the fairness of sentences in individual cases on average. To avoid the need to draw thick normative assumptions about morally equivalent cases and ideal sentences, and to overcome limitations in the availability of baseline data collected through the individual case approach, the article did not use data on a specific guideline enactment. Instead, it took a more general approach by estimating plausible bounds of the relevant parameters, and by exploring the effects for all possible values between these bounds. The article also made some simplifying assumptions about the distribution of sentences from morally equivalent cases. It assumed, for example, that these distributions are normally distributed.

The quantitative analysis revealed that increasing sentencing uniformity (i.e., decreasing in standard deviation) through guidelines can have a dramatic positive effect on fairness, and that this effect may often outweigh the negative effects of bias. The analysis began with the assumption that the fairness of sentences is linear and symmetric with respect to the ideal sentence. For morally equivalent cases with larger standard deviations (mean to standard deviation ratio of 4:3), increases in uniformity under these assumptions can have a two or three times greater positive effect on fairness than the negative effect of bias. Similar results were observed for cases with more moderate standard deviations (2:1). Reductions in standard deviation have a less impressive impact on fairness for preguideline distributions in which the standard deviation is low relative to the mean (i.e., ratio of 4:1 or 5:1). However, where the mean to standard deviation ratio is 4:1, each percentage change in the mean represents a dramatically larger shift than each percentage shift in the standard deviation: where the preguideline distribution has a mean of eight and a standard deviation of two, a 10 percent change in the mean is 0.8, while a

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

10 percent decrease in the standard deviation is 0.2. It is remarkable that even in this circumstance, where the bias in the mean is four times larger than the change in standard deviation, we still observe that a 10 percent decrease in standard deviation has a greater or equal effect on fairness than a 10 percent bias in the mean.

In a subsequent analysis, I relaxed the assumptions that the fairness of sentences follows a linear and symmetric function. Given the difficulty of examining the unlimited number of possible nonlinear and asymmetric functions, I explored several illustrative examples. First, I relaxed nonlinearity by modeling fairness according to several different exponential functions. This had only minor effects on the results. Second, I also relaxed the assumption of symmetry by applying a different functional form for sentences below the ideal sentence than for sentences above. The magnitude of the variance effect relative to the bias effect decreased when the slope of unfairness steepened more rapidly for sentences above the ideal sentence, but its size remained substantial and of continued substantive importance. This finding was observed even when functional forms of fairness were applied that strongly favored Blackmun's Thesis (e.g., X^2 for sentences above the ideal sentence, and a linear function for sentences below). The reverse pattern was observed when the slope of unfairness steepened more rapidly for sentences below the ideal sentence. Under these assumptions, the variance effect grew substantially relative to the bias effect.

Ultimately, the simulation models illustrate that decreasing the standard deviation of sentences from morally equivalent cases through sentencing guidelines may have large positive impacts on fairness that outweigh the negative effects of bias even when model assumptions favor Blackmun's Thesis. This is an important insight because it suggests that even where guideline systems increase (or decrease) the severity of sentences for an entire distribution of cases, modest increases in uniformity can yield a net increase in fairness. This bolsters the view that carefully developed guidelines likely increase rather than decrease fairness on average in individual cases. In turn, this provides significant evidence against Blackmun's Thesis.

The results of the study have two general policy implications. First, the large effect on fairness resulting from decreases in the standard deviation suggest that legislators and sentencing commissions can be less concerned about the potential of producing unfairness by constraining judicial discretion through robust sentencing guidelines. Second, the common legislative practice of enacting comprehensive guidelines that cover all criminal

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

offenses, such as the Federal Sentencing Guidelines, may be misguided. Some crime categories are likely characterized by low levels of sentencing disparity. Attempts to decrease sentencing variation among those cases will have only small positive effects on fairness that will more likely be outweighed by the bias effect. Sentencing commissions should collect data through the identical case approach to identify offense types or case features associated with high levels of disparity. They should then develop guidelines that focus on those cases.

References

Abrams, David S., Marianne Bertrand, & Sendhil Mullainathan (2012) "Do Judges Vary in Their Treatment of Race?" 41 *J. of Legal Studies* 347–83.

Albonetti, Celesta A. (1997) "Sentencing under the Federal Sentencing Guidelines: Effects of Defendant Characteristics, Guilty Pleas, and Departures on Sentence Outcomes for Drug Offenses, 1991-1992," 31 *Law & Society Rev.* 789–822.

Alschuler, Albert (1991) "The Failure of the Sentencing Guidelines: A Plea for Less Aggregation," 58 *Univ. of Chicago Law Rev.* 901–51.

Anderson, James M., Jeffrey R. Kling, & Kate Stith (1999) "Interjudge Sentencing Disparity: Before and After the Federal Sentencing Guidelines," 42 *J. of Law & Economics* 271–307.

Austin, William, & Thomas A. Williams (1977) "A Survey of Judges' Responses to Simulated Legal Cases," 68 *J. of Criminal Law Criminology* 306–10.

Baldus, David C., Charles Pulaski, & George Woodworth (1983) "Comparative Review of Death Sentences: An Empirical Study of the Georgia Experience," 74 *J. of Criminal Law & Criminology* 661–753.

Baumer, Eric P. (2013) "Reassessing and Redirecting Research on Race and Sentencing," 30 *Justice Q.* 231–61.

Bushway, Shawn D., & Anne Morrison Piehl (2001) "Judging Judicial Discretion: Legal Factors and Racial Discrimination," 35 *Law & Society Rev.* 733–64.

Bushway, Shawn D., Emily G. Owens, & Anne Morrison Piehl (2012) "Sentencing Guidelines and Judicial Discretion: Quasi-Experimental Evidence from Human Calculation Errors," 9 *J. Empirical Legal Studies* 291–319.

Chanenson, Steven L. (2004) "The Next Era of Sentencing Reform," 54 *Emory Law J.* 379–460.

Diamond, Shari Seidman, & Hans Zeisel (1975) "Sentencing Councils: A Study of Sentence Disparity and Its Reduction," 43 *Univ. of Chicago L. Rev.* 109–49.

Dixon, Jo (1995) "The Organizational Context of Criminal Sentencing," 100 *American J. of Sociology* 1157–98.

Fischman, Joshua, & Max M. Schanzenbach (2012) "Racial Disparities Under the Federal Sentencing Guidelines: The Role of Judicial Discretion and Mandatory Minimums," 9 *J. of Empirical Legal Studies* 729–64.

Forst, Brian, & Charles Wellford (1981) "Punishment and Sentencing: Developing Sentencing Guidelines Empirically from Principles of Punishment," 33 *Rutgers Law Rev.* 799–837.

Frase, Richard S. (1995) "State sentencing guidelines," 78 *Judicature* 173–79.

Freed, Daniel J. (1992) "Federal Sentencing in the Wake of Guidelines: Unacceptable Limits on the Discretion of Sentences," 101 *Yale Law J.* 1681–754.

Gaudet, Frederick J., George S. Harris, & Charles W. St. John (1934) "Individual Differences in Penitentiary Sentences Given by Different Judges," 18 *J. of Applied Psychology* 675–80.

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

Hofer, Paul J., Kevin R. Blackwell, & R. Barry Ruback (1999) "The Effect of the Federal Sentencing Guidelines on Inter-judge Sentencing Disparity," 90 *J. of Criminal Law & Criminology* 239–322.

Johnson, Brian (2003) "Racial and Ethnic Disparities in Sentencing Departures Across Modes of Conviction," 41 *Criminology* 449–89.

Karle, Theresa Walker, & Thomas Sager (1991) "Are Federal Sentencing Guidelines Meeting Congressional Goals?: An Empirical and Case Law Analysis," 40 *Emory Law J*. 393–444.

Kaut, Paula M. (2002) "Location, Location, Location: Interdistrict and Intercircuit Variation in Sentencing Outcomes for Federal Drug-Trafficking Offenses," 19 *Justice Q*. 633–71.

Kim, Pauline T. (2004) "Lower Court Discretion," 82 *New York Univ. Law Rev.* 383–442.

Lynch, Scott M., & Bruce Western (2004) "Bayesian Posterior Predictive Checks for Complex Models," 32 *Sociological Methods & Research* 301–35.

Miethe, Terance D. (1987) "Charging and Plea Bargaining Practices under Determinate Sentencing," 78 *J. Criminal Law & Criminology* 155–76.

O'Hear, Michael M. (2006) "The Original Intent of Uniformity in Federal Sentencing," 74 *Univ. Cincinnati. Law Rev.* 749–817.

Ogletree, Charles (1987) "The Death of Discretion? Reflections on the Federal Sentencing Guidelines," 101 *Harvard Law Rev.* 1938–60.

Osler, Mark (2003) "Must Have Got Lost: Traditional Sentencing Goals the False Trail of Uniformity and Process. And the Way Back Home," 54 *South Carolina Law Rev.* 649–89.

Partridge, Anthony, & William B. Eldridge (1974) "Second Circuit Sentencing Study—A Report to the Judges of the Second Circuit."

Payne, A. Abigail (1997) "Does Inter-Judge Disparity Really Matter? An Analysis of the Effects of Sentencing Reform in Three Federal District Courts," 17 *International Rev. of Law & Economics* 337–66.

Pfaff, John F. (2006) "The Continued Vitality of Structured Sentencing Following *Blakely*: The Effectiveness of Voluntary Guidelines," 54 *U.C.L.A. Law Rev.* 236–307.

Rhodes, Williams (1991) "Federal Criminal Sentencing: Measurement Issues With Application to Pre-Guideline Sentencing Disparity," 81 *J. Criminal Law and Criminology* 1002–33.

Rossi, Peter H., Richard A. Berk, & Alec Campbell (1997) "Just Punishments: Guideline Sentences and Normative Consensus," 13 *J. of Quantitative Criminology* 267.

Savelsberg, Joachim J. (1992) "Law that Does Not Fit Society: Sentencing Guidelines as a Neoclassical Reaction to the Dilemmas of Substantivized Law," 97 *American J. of Sociology* 1346–81.

Schwarzer, William W. (1991) "Judicial Discretion in Sentencing," 3 *Federal Sentencing Reporter* 339–41.

Scott, Ryan W. (2010) "Inter-Judge Sentencing Disparity After Booker: A First Look," 63 *Stanford Law Rev.* 1–66.

Seminar and Institute on Disparity of Sentences for Sixth, Seventh and Eighth Judicial Circuits (1962) 30 *Federal Rules of Decision* 401–505.

Sentencing Institute & Joint Council (1962) 30 *Federal Rules of Decision* 185–328.

Starr, Sonja B., & Marit Rehavi (2013) "Mandatory Sentencing and Racial Disparity: Assessing the Role of Prosecutors and the Effects of *Booker*," 123 *Yale Law J*. 2.

Stith, Kate, & Jose A. Cabranes (1998) *Fear of Judging: Sentencing Guidelines in the Federal Courts*. Chicago: University of Chicago Press.

Stith, Kate, & Steve Koh (1993) "The Politics of Sentencing Reform: The Legislative History of the Federal Sentencing Guidelines," 28 *Wake Forest Law Rev.* 223–90.

Stolzenberg, Lisa, & Stewart J. D'Alessio (1994) "Sentencing and Unwarranted Disparity," 32 *Criminology* 301–10.

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press

Stolzenberg, Ross M., & Daniel A. Relles (1997) "Tools for Intuition about Sample Selection Bias and its Correction," 62 *American Sociological Rev.* 494.

Tonry, Michael (2006) "Purposes and Functions of Sentencing," 34 *Crime and Justice* 1–52.

——— (1993) "The Failure of the U.S. Sentencing Commission's Guidelines," 39 *Crime & Delinquency* 131–49.

Ulmer, Jeffrey T. (1996) "Court Communities Under Sentencing Guidelines: Dilemmas of Formal Rationality and Sentencing Disparity," 34 *Criminology* 383–408.

Ulmer, Jeffrey T., & Brian Johnson (2004) "Sentencing in Context: A Multi-level Analysis," 42 *Criminology* 137–77.

Ulmer, Jeffrey T., Michael T. Light, & John H. Kramer (2011) "Racial Disparity in the Wake of the Booker/Fanfan Decision: An Alternative Analysis to the USSC's 2010 Report," 10 *Criminology & Public Policy* 1077–18.

U.S. Sentencing Commission (1987) *Federal Sentencing Manual*.

——— (1991) "A Report on the Operation of the Guidelines System and Short-Term Impacts on Disparity in Sentencing, Use of Incarceration, and Prosecutorial Discretion and Plea Bargaining."

——— (2010) Demographic Differences in Federal Sentencing Practices: An Update of the Booker Report's Multivariate Regression Analysis.

Waldfogel, Joel (1991) "Aggregate Inter-Judge Disparity in Federal Sentencing: Evidence from Three Districts," 4 *Federal Sentencing Reporter* 151–54.

Yang, Crystal S. (2014) "Have Inter-judge Sentencing Disparities Increased in an Advisory Guidelines Regime? Evidence from *Booker*." 89 *New York U. L. Rev.* 1268–1342.

Cases Cited

Blakely v. Washington, 542 U.S. 296 (2004).
Booker v. United States, 543 U.S. 220 (2005).
Callins v. Collins, 510 U.S. 1141 (1994).
Gall v. United States, 552 U.S. 38 (2007).
Kimborough v. United States, 552 U.S. 82 (2007).

Ben Grunwald *is a J.D. and Ph.D. student at the University of Pennsylvania studying criminal law and criminology. His research examines the effects of criminal procedure on the exercise of discretion in the criminal justice system.*

https://doi.org/10.1111/lasr.12141 Published online by Cambridge University Press