

REGULATING LOCAL VARIATIONS IN FEDERAL SENTENCING

Stephanos Bibas*

INTRODUCTION.....	137
I. TYPES, COSTS, AND SOURCES OF LOCAL VARIATION.....	139
A. <i>Justified and Unjustified Types of Local Variation</i>	139
B. <i>Sources of Local Variation</i>	141
II. A CASE STUDY OF UNJUSTIFIED VARIATION: FAST-TRACK PROGRAMS	145
III. A MIXED BAG: SUBSTANTIAL ASSISTANCE DEPARTURES	148
A. <i>How Substantial Assistance Works in Practice</i>	148
B. <i>How To Achieve the Right Blend of Uniformity</i>	151
CONCLUSION	154

INTRODUCTION

While federalism justifies variations among state laws, federal criminal law is supposed to be a uniform national response to crimes of national import. On paper, a single set of federal criminal statutes and Federal Sentencing Guidelines applies uniformly throughout the United States. But in practice, federal criminal charges and sentences vary greatly from state to state and from district to district. For example, some districts regularly prosecute low-level drug offenders. Others set high drug-quantity thresholds for charging and refer less significant cases to state authorities. In some districts, defendants must go to great lengths to earn cooperation discounts at sentencing. In others, much less cooperation will suffice.

Some of these variations reflect legitimate local responses to local crime patterns, needs, knowledge, and concerns. Other variations reflect local

* Associate Professor, University of Iowa College of Law; Visiting Associate Professor, University of Pennsylvania Law School, Fall 2005; Visiting Associate Professor, University of Chicago Law School, Winter and Spring 2006; Former Assistant U.S. Attorney, U.S. Attorney's Office for the Southern District of New York. B.A., Columbia University; B.A., M.A., University of Oxford; J.D., Yale Law School. E-mail: bibas@philo.org. Thanks to David Baldus, Rachel Barkow, Douglas Berman, Ronald Wright, and Tung Yin for their helpful comments and to Brian Raimondo and Robert Zink for their research assistance.

hostility to national policy choices, methods, and values. The law must accord some weight to local needs, concerns, and limitations, while still ensuring horizontal equity and consistency with national policy. This problem exemplifies the enduring tensions between *ex ante* rules and *ex post* discretion, between equality and individualization, and between a synoptic bird's-eye perspective and localized knowledge.

Equal treatment of similar offenders in different places is one important value in sentencing, but not the only one. A range of variation is necessary, and indeed healthy, to adapt national policy to localities. While some types of variation are necessary and even desirable, others are not. Variation that is too great or too blatant comes at the steep price of inequality, unfairness, and reduced deterrence.

Part I of this Article considers how regional and local sentencing patterns vary. Part I.A differentiates justified from unjustified variations. Justified variations are tactical responses to particular localized crime patterns, knowledge, and concerns. Variations are unjustified, however, when they reflect local hostility to national policy choices; arbitrariness; racial, ethnic, or class bias; or perhaps local implementation strategies at odds with national strategy. Part I.B considers how judges, head and line prosecutors, defense counsel, probation officers, and juries introduce local variations into sentencing patterns.

Part II is a case study of an unjustified variation that has sprung from macro-level crime problems. Southwestern border districts use fast-track programs, offering massive charge and sentence discounts to dispose of thousands of immigration and drug crimes swiftly. Supporters praise fast-track programs as a traditional use of prosecutorial discretion to respond to unique local caseloads and to punish the worst offenders most harshly. But these programs introduce large and blatant inequalities, undercut national policy, cloak the need to reallocate enforcement priorities, and truncate procedural protections. In the PROTECT Act, Congress authorized fast-track policies but limited their sentencing departures, a troubling compromise that sanctioned inequality while regulating it. Congress should abolish or at least further restrict these programs.

Part III moves onto a case study of justified and unjustified variations that stem from micro-level local practices: sentencing discounts for cooperating with law enforcement authorities under Federal Sentencing Guidelines § 5K1.1. Part III.A explains how this provision works and how districts implement it in practice. Districts vary greatly in how many defendants receive so-called "5K1 letters," how large the defendants' discounts are, how much cooperation it takes to win a letter, and why prosecutors offer discounts. Part III.B then considers the right blend of uniformity and local variation in this area. While line prosecutors have the best insight into their own cases and needs, procedural oversight and substantive guidelines on the acceptable forms of cooperation and the appropriate levels of sentencing discounts can improve

their decisions. In sum, national equality is a virtue, one that calls on us to minimize some but not all types of local variation. Prosecutorial discretion is a force for individualized justice, and inflexible rules can never take its place. Nevertheless, procedural and substantive regulation of charging, plea bargaining, and sentencing can check hostility to national policy while accommodating local problems and knowledge.

I. TYPES, COSTS, AND SOURCES OF LOCAL VARIATION

A. *Justified and Unjustified Types of Local Variation*

Justified local variations have principled rationales that are not at odds with national policies. First, local crime problems, caseloads, and knowledge vary and require varied responses. A local crime problem, such as a sudden rash of shootings, may require a swift and severe response, such as a crackdown on illegal gun trafficking. Usually there is no time to seek legislation or increased enforcement funding. Agents and prosecutors must use their enforcement discretion to respond ad hoc to crises, and judges may cooperate by issuing stiffer sentences.¹ Moreover, federal agents and prosecutors have local knowledge about how particular crimes are being committed. By targeting a particular money-laundering tactic, for example, agents may be able to bring down local drug rings that rely heavily on that tactic. Publicized targeting programs can also reassure the local populace that the crime du jour is under control, stemming crime waves and deterring copycat crimes. If the federal justice system responds to local enforcement needs, the result will be increased local respect and cooperation. If the system ignores pressing needs, federal law may lose local credibility and trust.

Local variations that lack these justifications carry significant costs. Two identical defendants who violate the same federal law in the same way in different places deserve the same punishment. Imposing different punishments undercuts national uniformity and equality. Moreover, consistent enforcement sends clear, unequivocal messages to prospective criminals. Conversely, variations undercut deterrence and the law's expressive message. This risk is especially great because criminal defendants tend to be overoptimistic and assume that they will receive sentences toward the lenient end of the spectrum.² Variations also make the law seem arbitrary, undercutting its perceived fairness and legitimacy. And once one locale carves out an exception to federal law, others may follow suit.

Even when a local variation carries little immediate cost, it may reflect and

1. See Vincent L. Broderick, *Local Factors in Sentencing*, 5 FED. SENT'G REP. 314, 315-16 (1993); Reena Raggi, *Local Concerns, Local Insights: Further Reasons for More Flexibility in Guideline Sentencing*, 5 FED. SENT'G REP. 306, 306 (1993).

2. See Stephanos Bibas, *Plea Bargaining Outside the Shadow of Trial*, 117 HARV. L. REV. 2463, 2498-2502 (2004).

reinforce troubling social values. Local variations may stem from or create racial, ethnic, or class disparities, as inner-city minorities may suffer heavier penalties than suburban whites who commit identical crimes. Indeed, concern about racial and ethnic disparities was one driving force behind the sentencing reform movement that culminated in the Federal Sentencing Guidelines.³

Local independence can also be troubling when it creates variation simply out of hostility to national policy rather than out of bias or prejudice. Policy and value variations are appropriate among states because federalism respects state sovereignty, but this conclusion does not justify variation within the national government. Our democratically elected representatives have decided to enact uniform national criminal laws to address national problems and enforce them with one voice through one agency—the U.S. Department of Justice. Locales that disagree cannot in effect secede from federal criminal law any more than they can secede from the Union. While some locales dislike the War on Drugs, for example, they should neither disregard federal law nor water down enforcement, but instead should agitate for change through Congress. Otherwise, these locales send equivocal messages to potential criminals, undercutting the deterrence and denunciation of crimes across the country.

The argument for uniformity is strongest for uniquely national crimes, such as immigration violations, and for crimes that in practice have to be prosecuted nationally, such as interstate drug rings. Both kinds of crimes typically have repercussions that extend far beyond a single district and do not simply displace state law and state policy choices.⁴

A more debatable class of variations occurs when a locale insists that it knows best how to implement federal values locally. A local U.S. Attorney's Office might claim, for example, that its high volume of cases requires a strategy of offering very lenient plea bargains. That office may achieve more deterrence if it plea bargains many cases swiftly for low sentences rather than holding out for average sentences in fewer cases. Because the argument for this low-price strategy rests not on local ends but on means, it is more justifiable than simple disagreement with national policy. However, this low-price approach is an implementation strategy, not a tactic, and strategy is a longer-

3. See William W. Wilkins, Jr. et al., *The Sentencing Reform Act of 1984: A Bold Approach to the Unwarranted Sentencing Disparity Problem*, 2 CRIM. L.F. 355, 362-63 (1991).

4. This argument does, however, carry less weight where federal enforcement simply duplicates state enforcement. When that happens, a prosecutor or agent's arbitrary decision to file a case federally preempts traditional state value choices. See generally *United States v. Lopez*, 514 U.S. 549, 559-64 (1995) (striking down the Federal Gun-Free School Zones Act of 1990 for exceeding Congress's Commerce Clause powers and intruding into states' traditional province of prescribing and enforcing local criminal law); Sara Sun Beale, *Too Many and Yet Too Few: New Principles To Define the Proper Limits for Federal Criminal Jurisdiction*, 46 HASTINGS L.J. 979, 1000 & n.90 (1995) (describing then-U.S. Attorney Rudolph Giuliani's Federal Day program, which prosecuted federally all drug dealers arrested by local police on one randomly chosen day each week).

term approach that is more amenable to national resolution. Moreover, because implementation strategies are visible, strategic disuniformity is likely to produce copycats in other districts and perceptions of inequity. A patchwork of varying implementation strategies also hides from Congress the longer-term issue of inadequate enforcement budgets and priorities, stifling national resolution of the underlying problem.

In short, justified variation is grounded in tactical decisions about localized crime problems—particularly, transient crime waves. Unjustified variation, in contrast, stems from value disagreement; from legally irrelevant factors such as race, ethnicity, sex, and class; or perhaps from strategic choices, especially concerning enduring crime problems.

How can Congress discourage unjustified variation while allowing justified variation to flourish? First, legislation could order the Sentencing Commission to monitor the problem. The Sentencing Commission could use its data to measure the different types of variation and how that variation affects charges, plea bargains, cooperation agreements, and sentence lengths. Variations that correlate closely with temporary, localized crime problems would be presumptively justifiable. Variations that are unrelated to local crime problems or that track legally irrelevant factors would be presumptively unwarranted. Qualitative surveys, as well as quantitative statistics, could ferret out evidence that variation stems from policy or value disagreement rather than local crime problems or knowledge. Second, congressional oversight committees could incorporate these data into their feedback loop, using them to investigate and reduce bad judicial and prosecutorial variations.

B. Sources of Local Variation

Many actors contribute to sentencing variations in the federal criminal justice system. Most obviously, judges sentence defendants differently. Local courthouse cultures, norms, and practices lead to local variations. In some districts, judges depart downward from the Sentencing Guidelines range in two to three percent of cases; in others, the rate is ten or more times higher.⁵ Judges in the Second, Ninth, and District of Columbia Circuits are twelve or more times more likely to depart downward than judges in the Fifth Circuit.⁶ Circuits vary, in part because some circuits' rulings give district judges greater freedom

5. See U.S. SENTENCING COMM'N, DOWNWARD DEPARTURES FROM THE FEDERAL SENTENCING GUIDELINES 35 fig.4, 36 fig.5 (2003) [hereinafter U.S. SENTENCING COMM'N, DOWNWARD DEPARTURES], <http://www.ussc.gov/departprt03/departprt03.pdf> (last visited Sept. 18, 2005); U.S. SENTENCING COMM'N, 2002 SOURCEBOOK OF FEDERAL SENTENCING STATISTICS 53-55 tbl.26 [hereinafter U.S. SENTENCING COMM'N, 2002 SOURCEBOOK].

6. U.S. SENTENCING COMM'N, FIFTEEN YEARS OF GUIDELINES SENTENCING: AN ASSESSMENT OF HOW WELL THE FEDERAL CRIMINAL JUSTICE SYSTEM IS ACHIEVING THE GOALS OF SENTENCING REFORM 112 (2004) [hereinafter U.S. SENTENCING COMM'N, FIFTEEN-YEAR REPORT] (measuring judicial downward departures for mitigating circumstances, excluding fast-track and substantial assistance departures).

to depart than others circuits' rulings and in part because of local traditions.⁷ But even within the same circuit, district departure rates vary widely.⁸ In some districts, judges allow stipulated-sentence plea agreements to trump the Guidelines, while judges in other districts do not accept such agreements.⁹ Districts also vary in how far their judges depart from the Guidelines.¹⁰ In addition, judges use different parts of the sentencing range to sentence defendants. Most routinely use the bottom of the range, but some use the entire range. This last phenomenon, however, explains only a tiny fraction of sentence variation.¹¹

Prosecutors also contribute significantly to local variations. Different U.S. Attorney's Offices take different approaches to charging and declining cases. One district may charge defendants with possessing a few pounds of marijuana. In contrast, the Southern District of California routinely referred to state authorities or deferred federal prosecution of couriers caught with less than 125 pounds of marijuana.¹² Different offices also vary in dropping charges as part

7. Compare Lisa M. Farabee, *Disparate Departures Under the Federal Sentencing Guidelines: A Tale of Two Districts*, 30 CONN. L. REV. 569, 610-14 (1998) (describing the "tradition of independence" among Connecticut federal judges), with *id.* at 623-26 (stating that Massachusetts federal judges are more willing to follow the letter of the Guidelines, in part because "the First Circuit stringently interpreted and applied the Guidelines").

8. U.S. SENTENCING COMM'N, FIFTEEN-YEAR REPORT, *supra* note 6, at 112 (noting, among other examples, that the District of Montana's downward-departure rate for mitigating circumstances is about one quarter of the Northern District of California's rate, even though both are in the Ninth Circuit).

9. FED. R. CRIM. P. 11(c)(1)(C). Compare, e.g., *United States v. Aguilar*, 884 F. Supp. 88, 91-92 (E.D.N.Y. 1995) (Weinstein, J.) (allowing the parties to use a stipulated-sentence plea agreement to trump the Guidelines because the bargain led to a fairer sentence than the Guidelines would have), with Probation Officers Advisory Group, *Probation Officers Advisory Group Survey*, 8 FED. SENT'G REP. 303, 305 (1996) (reporting that in 73% of districts, stipulated-sentence agreements are rare (occurring in 5% or fewer of cases), while in 7% of districts they occur in more than half of cases and in 11% of districts they occur in between one-quarter and one-half of cases), and *id.* at 307 app. A (reporting that in many circuits stipulated-sentence agreements are rare and that in the Fifth Circuit "[j]udges are hesitant to take such pleas").

10. See, e.g., Jeffrey T. Ulmer, *The Localized Uses of Federal Sentencing Guidelines in Four U.S. District Courts: Evidence of Processual Order*, 28 SYMBOLIC INTERACTION 255, 263-64 & tbl.1 (2005) (reporting that § 5K1.1 departures in some districts are typically much more generous than in others).

11. U.S. SENTENCING COMM'N, FIFTEEN-YEAR REPORT, *supra* note 6, at 102-03 (noting that placement within the Guidelines range explains less than one-sixtieth as much variation as substantial assistance departures and less than one-thirtieth as much as other downward departures); *id.* at 109-10 (noting that 880 of 911 federal judges typically sentence at the bottom of the range); see also 28 U.S.C. § 994(b)(2) (2005) (specifying that the top of each sentencing range shall be no more than six months or twenty-five percent above the bottom of the range, whichever is greater, except that if the minimum is thirty or more years, the maximum may be life).

12. Alan D. Bersin & Judith S. Feigin, *The Rule of Law at the Margin: Reinventing Prosecution Policy in the Southern District of California*, 12 GEO. IMMIGR. L.J. 285, 293-95 (1998).

of plea agreements, in agreeing to or not opposing downward departures, and in recommending sentences. For example, in some districts, prosecutors agree as part of plea bargains to recommend sentences at the top or bottom of ranges, while in other districts prosecutors do not.¹³

The main U.S. Department of Justice in Washington, D.C. (Main Justice), periodically tries to regulate federal prosecutors by promulgating centralized policies. For example, a memorandum from Attorney General John Ashcroft to federal prosecutors reiterated that prosecutors must usually charge and not bargain away the most serious, readily provable criminal charges.¹⁴ There is more centralized bureaucracy and oversight at the federal than the state level, which helps to keep far-flung federal offices in line. Each new President appoints the U.S. Attorney for each district, who is accountable to the Attorney General and is supposed to implement these Main Justice policies. While U.S. Attorneys typically come from the local bar, they are unelected and so cannot claim a mandate to implement idiosyncratic local priorities. Nevertheless, federal prosecutors are human beings, and their values, priorities, and tactics doubtless vary. Some have worked recently in Main Justice and are wedded to its mission,¹⁵ while others are far removed from Main Justice and its priorities. Indeed, one of the largest urban districts is known for its autonomy and resistance to Main Justice oversight.¹⁶

Head prosecutors inevitably vary in their decisions about which crime problems to target. Some priorities come from Main Justice and Congress, but

13. When I was a federal prosecutor, the practice of the U.S. Attorney's Office for the Southern District of New York was not to plea bargain over the point within the range at which the judge should sentence the defendant. Federal prosecutors from other districts, such as the District of Columbia, have told me that in their districts plea bargains commonly stipulate to particular sentences within the ranges. *See also* Frank O. Bowman, III & Michael Heise, *Quiet Rebellion? Explaining Nearly a Decade of Declining Federal Drug Sentences*, 86 IOWA L. REV. 1043, 1122 (2001) (finding evidence that some drug-trafficking plea bargains stipulated to sentences at the tops of the applicable ranges).

14. Memorandum from Attorney General John Ashcroft on Department Policy Concerning Charging Criminal Offenses, Disposition of Charges, and Sentencing, to All Federal Prosecutors (Sept. 22, 2003), *reprinted in* 16 FED. SENT'G REP. 129 (2003).

15. Indeed, a few of the offices closest to Washington, D.C., (such as the U.S. Attorney's Office for the District of Columbia and the Alexandria branch of the U.S. Attorney's Office for the Eastern District of Virginia) routinely host Main Justice personnel who rotate in as Special Assistant U.S. Attorneys to gain trial experience. *See* Okianer Christian Dark, *Transitioning from Law Teaching to Practice and Back Again: Proposals for Developing Lawyers Within the Law School Program*, 28 J. LEGAL PROF. 17, 19 (2003-2004) (explaining how the author served as a Special Assistant U.S. Attorney to gain trial experience after working in the Main Justice); Mark D. Greenberg & Harry Litman, *The Meaning of Original Meaning*, 86 GEO. L.J. 569, 569 nn.*, ** (1998) (noting that both authors worked in the Main Justice and then as Special Assistant U.S. Attorneys in the Eastern District of Virginia).

16. Chitra Ragavan, *The Pardon Buck Stops in New York: U.S. Attorney Mary Jo White Takes the Lead*, U.S. NEWS & WORLD REP., Mar. 26, 2001, at 24 (stating that "because of its legendary independence and tenacity," the U.S. Attorney's Office for the Southern District of New York "is known as the 'sovereign district' of New York").

others bubble up in response to local crime problems. Head prosecutors inevitably allocate scarce resources in deciding which crimes to investigate and prosecute.¹⁷

Line prosecutors vary greatly as well. Even though they are hired by a presidential appointee, Assistant U.S. Attorneys inevitably reflect the communities from which they come. U.S. Attorney's Offices, moreover, vary greatly in their sizes, urbanness, and rates of turnover. To regulate larger and more transient workforces and ensure uniformity, large urban offices have more centralized policies and supervisory oversight.¹⁸ Each office has its own culture, charging and sentencing practices, and customary sentences for particular crimes.¹⁹

Defense counsel are a third possible source of disparity. Some districts have aggressive, knowledgeable federal defenders' offices that exploit and stretch every possible loophole in the Guidelines for their clients.²⁰ Other districts have overburdened, less-aggressive, or less-experienced defense lawyers who know less about how to exploit the Guidelines.²¹ Because repeat players who pool information are best able to exploit these complexities, federal public defenders probably achieve lower sentences than private lawyers

17. One could argue that prosecutor-initiated local variations are thus more justifiable and more inevitable than judge-initiated variations. Congress has the power to override these allocation decisions by funding prosecutor or agent positions dedicated to pursuing particular kinds of crime. It should, however, be reluctant to supersede superior local knowledge and so should refrain from earmarking particular positions for particular crimes. Also, in studying variations, Congress needs to separate out prosecutor-initiated downward departures from judge-initiated downward departures. To facilitate congressional oversight, prosecutors' statistical reports should distinguish departures on prosecutors' motions or with their agreement, departures that prosecutors neither support nor oppose, and departures that prosecutors actively oppose. Now that prosecutors initiate downward departures in more than one-fifth of all cases, Congress should scrutinize them carefully, instead of simply blaming the problem on judges.

18. See JAMES EISENSTEIN, COUNSEL FOR THE UNITED STATES: U.S. ATTORNEYS IN THE POLITICAL AND LEGAL SYSTEMS 117 (1978); JOAN E. JACOBY, THE AMERICAN PROSECUTOR: A SEARCH FOR IDENTITY 70-71 (1980).

19. See EISENSTEIN, *supra* note 18, at 5-7.

20. See Ilene H. Nagel & Stephen J. Schulhofer, *A Tale of Three Cities: An Empirical Study of Charging and Bargaining Practices Under the Federal Sentencing Guidelines*, 66 S. CAL. L. REV. 501, 529-30 (1992) (quoting probation officers' observations that in one district prosecutors are "outgunned" because of public defenders' "commitment to the idea that knowledge [of the Guidelines] is power").

21. See Douglas A. Berman, *From Lawlessness to Too Much Law? Exploring the Risk of Disparity from Differences in Defense Counsel Under Guidelines Sentencing*, 87 IOWA L. REV. 435, 444-57 (2002) (explaining how the increasing complexity of federal sentencing law risks creating disparate outcomes based on the quality, experience, and funding of one's defense lawyer); Bibas, *supra* note 2, at 2484 n.83 (providing other sources that substantiate and explore the disparity in the quality of defense counsel); Farabee, *supra* note 7, at 625 (reporting judges' comments that defense counsel in Massachusetts argue far less for departures than do Connecticut defense counsel, in part because Massachusetts lawyers are not well trained in Guidelines advocacy).

who take occasional ad hoc court appointments.²² Thus, districts that rely heavily on federal defenders' offices may well have systematically lower sentences as a result.

A fourth source of disparity is probation officers. Probation officers prepare presentence reports, which are supposed to allow judges to base sentences on defendants' actual conduct. In some districts, probation officers simply summarize and digest whatever information prosecutors and agents give to them. In others, they conduct truly independent investigations.²³ In some districts, probation officers apply the Guidelines strictly, while in others they are more willing to recommend departures.²⁴ In some districts, judges rubber-stamp the parties' plea bargains and sentencing recommendations, even when they conflict with probation officers' reports. In others, judges defer to probation officers' versions of the facts and Guidelines computations, thwarting plea bargains that try to circumvent the Guidelines.²⁵ Indeed, at least one district employs lawyers as probation officers; this arrangement may lead judges to give their recommendations more weight.²⁶

One might think that juries introduce local sentencing disparities. Today, however, plea bargains resolve ninety-five percent of adjudicated cases. And in the few jury trials, juries do not set sentences, do not know the sentencing consequences of their verdicts, and are instructed not to consider sentences.²⁷ Thus, juries play a negligible role in sentencing variation.

II. A CASE STUDY OF UNJUSTIFIED VARIATION: FAST-TRACK PROGRAMS

Each year, more than 900,000 people cross the Mexican-American border illegally into California, Arizona, New Mexico, and Texas.²⁸ Federal agents in these districts arrest far more illegal immigrants, alien smugglers, and drug traffickers than prosecutors could possibly bring to trial. Until the 1990s, federal prosecutors pursued a tiny fraction of the most serious offenders and

22. See Bibas, *supra* note 2, at 2534.

23. See Gerald W. Heaney, *The Reality of Guidelines Sentencing: No End to Disparity*, 28 AM. CRIM. L. REV. 161, 200-02 (1991); see also *Assessing the Probation Officers' Survey: Does Fact Bargaining Undermine the Sentencing Guidelines?*, 8 FED. SENT'G REP. 299, 305 (1996).

24. Compare Farabee, *supra* note 7, at 604-07 (describing flexibility among probation officers in Connecticut), with *id.* at 618-21 (describing Massachusetts probation officers as "true guardians of the Guidelines").

25. See Ulmer, *supra* note 10, at 264 tbl.1, 266-67.

26. See *id.* at 266.

27. Bowman & Heise, *supra* note 13, at 1182-83 & n.475.

28. BUREAU OF IMMIGRATION STATISTICS, U.S. DEP'T OF HOMELAND SECURITY, 2003 YEARBOOK OF IMMIGRATION STATISTICS 155 tbl.37 (2004) (reporting that Southwest Border Patrol offices located 905,065 deportable aliens in 2003), <http://uscis.gov/graphics/shared/statistics/yearbook/2003/Table37.xls> (last visited Sept. 18, 2005). Presumably there are even more illegal immigrants whom the Border Patrol does not succeed in locating.

declined to prosecute the rest. Most immigration-related cases in San Diego, for example, went unpunished because the San Diego District Attorney refused to prosecute cases related to the border.²⁹

In the last decade, federal prosecutors along the southwestern border established fast-track programs. While the details vary, these programs typically ask defendants to waive indictment, discovery, and presentence reports; plead guilty at the initial appearance; and consent to immediate sentencing. In return, prosecutors agree to recommend downward departures or let defendants plead to lesser charges. Because these cases move much more quickly, prosecutors can process many more of them.³⁰

Supporters defend these programs as necessary expedients to process huge numbers of cases. They view these charging, plea bargaining, and sentencing decisions as exercises of traditional prosecutorial discretion.³¹ The enormous caseloads along the border, they argue, are unique. By offering large sentence reductions, these districts can process many more cases, saving time and money and reducing court congestion.³² Processing more cases through truncated procedures also generates more deterrence and leaves fewer crimes unpunished. The effect is to reduce the hidden inequality of not punishing many crimes at all. Because border districts are deluged with illegal immigration and drug trafficking, they need drastic measures to stem the unique local harms that they suffer.³³ Besides, by fully prosecuting only the worst offenders and offering fast-track deals to the rest, prosecutors tailor punishments to defendants' culpability, softening unduly harsh sentences.³⁴ In short, supporters claim that the Southwest's large crime problem requires a different enforcement strategy.

29. See William Braniff, *Local Discretion, Prosecutorial Choices and the Sentencing Guidelines*, 5 FED. SENT'G REP. 309, 309-10 (1993).

30. *Implementing the Requirements of the PROTECT Act: Hearing Before the U.S. Sentencing Comm'n*, Sept. 23, 2003, at 11-12 (testimony of the Hon. Marilyn L. Huff, judge, U.S. District Court for the Southern District of California) [hereinafter *Implementing the Requirements of the PROTECT Act*]; *id.* at 74-79 (testimony of Paul K. Charlton, U.S. Attorney for the District of Arizona); see also Bersin & Feigin, *supra* note 12, at 301; Braniff, *supra* note 29, at 310-11 (noting that by charge-bargaining felonies down to misdemeanors, the U.S. Attorney's Office for the Central District of California was able to handle 6000 more cases).

31. Braniff, *supra* note 29, at 311; see also *Implementing the Requirements of the PROTECT Act*, *supra* note 30, at 4 (testimony of the Hon. Marilyn L. Huff, judge, U.S. District Court for the Southern District of California); *id.* at 48 (testimony of the Hon. Lourdes G. Baird, judge, U.S. District Court for the Central District of California).

32. See *Implementing the Requirements of the PROTECT Act*, *supra* note 30, at 5-14 (testimony of the Hon. Marilyn L. Huff, judge, U.S. District Court for the Southern District of California).

33. See Alan D. Bersin, *Reinventing Immigration Law Enforcement in the Southern District of California*, 8 FED. SENT'G REP. 254, 254 (1996); Bersin & Feigin, *supra* note 12, at 286-90; Braniff, *supra* note 29, at 309-11, 313.

34. Bersin, *supra* note 33, at 256; see also *Implementing the Requirements of the PROTECT Act*, *supra* note 30, at 19-22 (testimony of Steven Hubachek, Assistant Federal Public Defender for the Southern District of California).

True, southwestern districts do face an enormous volume of cases—a macro-level structural force that may require special treatment. But fast-track programs nonetheless qualify as unjustified variations, as they create significant, visible inequities. Institutionalized fast-track programs are far different from traditional, low-visibility, case-by-case prosecutorial discretion. They are different from rules of thumb that prosecutors use in secret triage to decide which cases to pursue with their limited resources. In essence, the border districts began engaging in wholesale legislative policymaking, supplanting Congress's dominant role.

Nor are these approaches inevitable. The Southern District of Florida processes its many immigration and drug-trafficking cases without a fast-track program and tries more cases than all five border districts combined.³⁵ Prosecutors can always claim that their huge workloads justify larger discounts, as there are always more cases than they can try. Though here the claim is strong, it sets a bad precedent,³⁶ and there is precious little oversight or check on these claims.

The result is inequality under a supposedly nationwide law that serves nationwide interests. The whole country, not just the Southwest, needs federal protection from drug smugglers, alien smugglers, and aggravated felons who reenter via the Southwest. Blatant inequality serves as a precedent for nonborder districts and tempts other courts to lower their sentences to match.³⁷ The result is a patchwork of inequality that subverts nominal sentences set by Congress and the Sentencing Commission. And in the process, fast-track programs bypass most procedural safeguards, truncating plea bargaining even more and increasing the risk of convicting the innocent.³⁸ These districts

35. See *Implementing the Requirements of the PROTECT Act*, *supra* note 30, at 108-09, 118-19 (testimony of Frank O. Bowman, law professor and former Assistant U.S. Attorney, Southern District of Florida) (“This gaping chasm between what might be done [to process cases along the border] and what is now done cries out, I think, for some explanation. Why aren’t the border districts trying hundreds of cases every year and using the threat of trials to force guidelines compliance?”).

36. Indeed, many districts that are not along the border have followed the Southwest’s lead. Now, up to half of judicial districts have some form of fast-track program. U.S. SENTENCING COMM’N, DOWNWARD DEPARTURES, *supra* note 5, at 64; see also *United States v. Perez-Chavez*, No. 2:05-CR-00003PGC, 2005 U.S. Dist. LEXIS 9252, at *44 app. A (D. Utah May 16, 2005) (reporting that many districts that do not adjoin the Mexican-American border have fast-track programs, including the Northern District of California; the Northern District of Georgia; the Districts of Idaho, Nebraska, North Dakota, and Oregon; and the Western District of Washington).

37. *United States v. Ramirez-Ramirez*, 365 F. Supp. 2d 728, 732 (E.D. Va. 2005); Sentencing Memorandum at 8-10, *United States v. Galvez-Barrios*, 355 F. Supp. 2d 958 (E.D. Wis. 2005); U.S. SENTENCING COMM’N, DOWNWARD DEPARTURES, *supra* note 5, at 67; Erin T. Middleton, Note, *Fast-Track to Disparity: How Federal Sentencing Policies Along the Southwest Border Are Undermining the Sentencing Guidelines and Violating Equal Protection*, 2004 UTAH L. REV. 827, 848-50 (advocating downward departures to equalize sentences and avoid equal protection violations).

38. See *Implementing the Requirements of the PROTECT Act*, *supra* note 30, at 52-56

should bring their chronic enforcement problems out into the open and let Congress decide whether to allocate more money, rather than concealing the problems with idiosyncratic strategies.³⁹

Congress regulated ad hoc fast-track programs in § 401(m)(2)(B) of the PROTECT Act. That provision directed the Sentencing Commission to issue a policy statement authorizing fast-track downward departures of up to four levels. For a defendant to receive the departure, the Attorney General and U.S. Attorney must authorize an early disposition program, and the prosecutor must move for the departure.⁴⁰

In other words, Congress struck a compromise. On the one hand, it legitimated the various ad hoc programs that had been proceeding without formal approval. On the other hand, it regulated them procedurally (by requiring high-level approval) and substantively (by capping the departure at four levels). The regulation adds democratic legitimacy to prosecutors' unauthorized policymaking and strategic choices. It also limits the size of the disparity in each case. And yet, by bringing these reductions out into the open and sanctioning them, the compromise openly enshrines unwarranted local disparities.⁴¹ This blatant, pragmatic, ad hoc acceptance of local disparities may pave the way for further local and regional variations. By openly allowing some disparities and attacking others, Congress risks appearing hypocritical.

Instead, Congress should abolish or clamp down on this unjustified variation to reduce inequality as far as possible. If Congress simply cannot abolish these programs, at the very least it should limit the regions, crimes, and numbers of persons covered by them. There is no excuse for letting these ad hoc programs proliferate beyond their *raison d'être*—southwestern illegal immigrants and low-level drug couriers.

III. A MIXED BAG: SUBSTANTIAL ASSISTANCE DEPARTURES

A. *How Substantial Assistance Works in Practice*

A second case study involves variation that is attributable not to local crime problems, but to local customs and habits. Prosecutors have long struck deals with informants and cooperating witnesses. Federal Sentencing Guidelines § 5K1.1 formalizes and cabins this venerable practice by

(testimony of Maria E. Stratton, Federal Public Defender for the Central District of California).

39. See *id.* at 106-11 (testimony of Frank O. Bowman, III, law professor and former Assistant U.S. Attorney, Southern District of Florida).

40. Pub. L. No. 108-21, § 401(m)(2)(B), 117 Stat. 650 (2003); see U.S. SENTENCING GUIDELINES MANUAL § 5K3.1 (2004) (implementing this statutory provision).

41. See *Implementing the Requirements of the PROTECT Act*, *supra* note 30, at 99-101 (testimony of Frank O. Bowman, III, law professor and former Assistant U.S. Attorney, Southern District of Florida).

authorizing downward departures for cooperating defendants. To earn the departure, a defendant must provide assistance that is substantial in the investigation or prosecution of others. In addition, the prosecution must file a motion, commonly called a “5K1 letter,” certifying that the cooperator provided substantial assistance. Finally, the judge must find that, in fact, the defendant provided substantial assistance in the investigation or prosecution of others. The judge’s findings unlock the Guidelines, allowing the judge to depart below the otherwise applicable sentencing range. Section 5K1.1 does not specify what percentage or how many months’ reduction the judge should award.⁴² Section 5K1.1 is controversial to begin with. It treats some defendants much more leniently than others based not on their just deserts, but on their ability to help catch and convict others. The resulting inequality of sentences is a necessary evil because it gives some culpable defendants less punishment than they deserve in order to catch other defendants.

Compounding this inherent inequity is another cluster of inequities: prosecutors and judges vary widely in how they implement the provision. First, districts vary greatly in the raw percentages of defendants who earn substantial assistance departures. In some districts, fewer than four percent receive them, while in others the rate is forty percent or more.⁴³

Second, prosecutors’ offices vary in defining what conduct qualifies for a substantial assistance departure. For example, in some districts cooperators must take part in undercover activity, while in others they can earn departures simply by providing marginally useful information.⁴⁴ Offices are almost evenly divided on whether a defendant can qualify simply by providing information about his own crimes.⁴⁵

Third, districts vary greatly in the size or extent of departures that they award. In some districts, cooperators routinely earn discounts of one-third to one-half off the otherwise applicable Sentencing Guidelines range. In others, most cooperators earn discounts of two or three levels, which reduce the applicable range by roughly twenty-five to thirty-five percent.⁴⁶ In some

42. U.S. SENTENCING GUIDELINES MANUAL § 5K1.1 (2004).

43. U.S. SENTENCING COMM’N, 2002 SOURCEBOOK, *supra* note 5, at 53-55 & tbl.26.

44. *See* Ulmer, *supra* note 10, at 263-65 & tbl.1 (reporting that in the pseudonymous Northeast District defendants received departures even if they “provided information of questionable value,” while in the Western District “you have to put yourself in jeopardy. Take a risk. Wear a wire, buy drugs, something like that” (quoting an experienced prosecutor)).

45. *See* LINDA DRAZGA MAXFIELD & JOHN H. KRAMER, SUBSTANTIAL ASSISTANCE: AN EMPIRICAL YARDSTICK GAUGING EQUITY IN CURRENT FEDERAL POLICY AND PRACTICE 9, 26 exhb.4 (1998), <http://www.ussc.gov/publicat/5kreport.pdf> (last visited Sept. 18, 2005); Stanley Marcus, *Substantial Assistance Motions: What Is Really Happening?*, 6 FED. SENT’G REP. 6, 6-8 (1993). This disagreement is surprising, as the text of § 5K1.1 appears to require substantial assistance against others, not just against oneself.

46. *See* Ronald S. Safer & Matthew C. Crowl, *Substantial Assistance Departures: Valuable Tool or Dangerous Weapon?*, 12 FED. SENT’G REP. 41, 43-44 (1999); Ulmer, *supra* note 10, at 263-65 & tbl.1; *see also* Bibas, *supra* note 2, at 2488 & n.100 (citing a source for

districts, prosecutors recommend specific sentences for cooperators, while in about one-third of federal districts they do not.⁴⁷ These recommendations serve as mental anchors or benchmarks and so greatly influence actual sentences.⁴⁸

Districts also vary because judges in some districts resent what they view as unduly severe Guidelines sentences. These judges are happy to exploit substantial assistance departures to escape the Guidelines' strictures.⁴⁹ Other judges, however, seek to peg their departures to the Guidelines, offering proportionate and more modest discounts. Likewise, districts vary in how they write and approve 5K1 letters. In some, an individual supervisor or a committee reviews and approves letters before line prosecutors may file them. In others, line prosecutors have freer rein.⁵⁰ The latter approach allows more room for individual prosecutors' senses of justice to influence sentences. In particular, line prosecutors often come to sympathize with their cooperators and thus seek or acquiesce in lower sentences, instead of minimizing the sentence discount paid for information.⁵¹

Fourth, districts vary in reviewing line assistants' substantial assistance recommendations. Many districts have no written review policies, and even those that do are often inconsistent in enforcing these policies.⁵²

Fifth, districts vary in their reasons for offering cooperation discounts. Most obviously, districts use the discounts to gain assistance in successfully prosecuting cases that they otherwise could not pursue. Large cooperation discounts can induce gang members, who would otherwise remain silent out of fear and loyalty, to flip and cooperate against gang leaders.⁵³ Thus, large urban districts that prosecute many complex, multi-defendant cases may need to use cooperation discounts more often than districts that handle smaller cases. Less obviously, prosecutors who are risk averse or too lazy to find evidence in other ways may overbuy cooperation. By doing so, they shore up cases that were probably winnable with fewer or no cooperators.⁵⁴ In some districts,

the proposition that a three-level discount is on average a thirty-five percent reduction in the severity of the sentence).

47. MAXFIELD & KRAMER, *supra* note 45, at 33 exhb.11, 34 exhb.12.

48. *See* Bibas, *supra* note 2, at 2518-19.

49. *See* Daniel C. Richman, *Cooperating Defendants: The Costs and Benefits of Purchasing Information from Scoundrels*, 8 FED. SENT'G REP. 292, 294 (1996); Jack B. Weinstein, *A Trial Judge's Reflections on Departures from the Federal Sentencing Guidelines*, 5 FED. SENT'G REP. 6, 12 (1992).

50. *See* MAXFIELD & KRAMER, *supra* note 45, at 7-8, 25 exhb.3; Richman, *supra* note 49, at 292.

51. *See* Nagel & Schulhofer, *supra* note 20, at 531-32; Richman, *supra* note 49, at 292; *see also* Ellen Yaroshesky, *Cooperation with Federal Prosecutors: Experiences of Truth Telling and Embellishment*, 68 FORDHAM L. REV. 917, 944 (1999) (quoting a former federal prosecutor about the danger of "falling in love with your rat").

52. MAXFIELD & KRAMER, *supra* note 45, at 24 exhb.2, 25 exhb.3.

53. *See* Safer & Crowl, *supra* note 46, at 42, 44.

54. *See* Richman, *supra* note 49, at 293.

prosecutors go even further, using cooperation discounts simply as plea bargaining tools to mitigate seemingly harsh sentences. This desire explains why, in at least one district, the fig leaf of almost any information suffices to earn a 5K1 letter.⁵⁵ There is even evidence that conscious or unconscious bias may be entering the system. On average, men, blacks, Hispanics, noncitizens, and high-school dropouts receive fewer and smaller 5K1 departures than women, nonminorities, citizens, and high school graduates.⁵⁶

B. *How To Achieve the Right Blend of Uniformity*

What is the right blend of uniformity and local variation in awarding substantial assistance departures? As a rule, line prosecutors have the best insight into their cases, their own need for cooperation, and the likely fruits of that cooperation. They have lived with their cases the longest and have a feel for the factual nuances and cooperators' truthfulness and value.⁵⁷ The criminal justice system trusts prosecutorial discretion and judgment, in part because no rule can substitute for it.

Yet, even though prosecutorial discretion is necessary and useful, it should not be unfettered or arbitrary. Procedural and substantive regulations can help to channel, guide, and equalize cooperation-related decisions. By doing so, these regulations can counteract prosecutors' self-interest in disposing of cases quickly. The remainder of this section sketches out a few procedural and substantive suggestions for increasing national uniformity while preserving appropriate local discretion.

Procedural mechanisms: Supervisory prosecutors are well placed to regulate line prosecutors' behavior. U.S. Attorneys are accountable to Main Justice and to the President who appointed them, and they and their deputies can supervise substantial assistance decisions. Main Justice could promulgate written policies detailing the types of supervision and approval that U.S. Attorneys and division and unit chiefs should exercise over line prosecutors. Line prosecutors should be required to write up reports summarizing and justifying each case, its evidentiary holes, the need for cooperating witnesses, each cooperator's background, and the anticipated payoff from cooperation. These reports would go to office-wide committees, which would read each report, ask questions, and compare each case to others.⁵⁸ The committee would

55. See Ulmer, *supra* note 10, at 263-65 & tbl.1 (reporting that the pseudonymous Northeast District routinely awards "soft 5K1s" in exchange for "information of questionable value" because, according to defense attorneys, all the parties want to soften Guidelines sentences that seem too harsh); see also Ian Weinstein, *Regulating the Market for Snitches*, 47 BUFF. L. REV. 563, 613-15 (1999).

56. MAXFIELD & KRAMER, *supra* note 45, at 31 exhb.9, 34 exhb.12.

57. See Richman, *supra* note 49, at 294.

58. See Bibas, *supra* note 2, at 2541-42; see also Ronald Wright & Marc Miller, *The Screening/Bargaining Tradeoff*, 55 STAN. L. REV. 29, 62-66 (2002) (describing a system of

apply standardized criteria set forth in the Main Justice policy to ensure greater uniformity across districts. This process would vet both cooperation agreements early on in a case and 5K1 letters shortly before sentencing.

Procedural review could check excessive cooperation in several ways. First, the documentation and approval requirements would restrict the frequency of cooperation agreements and ensure some threshold showing of prosecutorial need. Likewise, procedural review might veto some 5K1 letters or restrict recommended discounts for cooperation that is only marginally useful. Second, Main Justice and congressional subcommittees could conduct periodic oversight of prosecutors' written cooperation reports. They might particularly scrutinize districts whose 5K1 rates or discount sizes well exceed the norm. Some large urban districts could justify their high rates by showing that their many complex, multi-defendant cases require cooperators; their particular local crime problems may justify a measure of variation. Third, to combat exorbitant cooperation rates in excess of forty percent in some districts, Congress or Main Justice could consider rationing. One possible way to ration would be to mandate that no more than fifteen percent of defendants in each district receive cooperation agreements; the cap for large urban districts might be higher. The office-wide review committees would then decide which proposed agreements are most meritorious.⁵⁹ It would not be easy to set the right amount of rationing, and Congress and Main Justice would probably err on the side of allowing generosity. Nevertheless, some form of rationing could force prosecutors to rank priorities and buy cooperation where they need it most.

Substantive review: Congress, Main Justice, and the Sentencing Commission could also adopt substantive policies to ensure more consistent outcomes across districts. First, they could categorize the types of assistance that qualify as substantial, much as death penalty analysis groups cases by aggravating and mitigating factors.⁶⁰ Typologies would build on descriptive evidence of existing consensuses among prosecutors' offices. They would also harmonize outliers and propose prescriptive criteria tied to the importance of particular forms of assistance.

For example, the Sentencing Commission's own data indicate a consensus that participating in an undercover investigation or testifying truthfully is

supervisory control and review that successfully restricted plea bargaining in the New Orleans District Attorney's Office). This process could be modeled on (though more streamlined than) the Department of Justice's Capital Case Review Committee process. See generally Rory K. Little, *The Federal Death Penalty: History and Some Thoughts About the Department of Justice's Role*, 26 *FORDHAM URB. L.J.* 347, 406-29 (1999). Committees would need some flexibility to expedite review of cooperation agreements in cases where time is of the essence.

59. Weinstein, *supra* note 55, at 630-31 (advancing this proposal).

60. See, e.g., David C. Baldus et al., *Arbitrariness and Discrimination in the Administration of the Death Penalty: A Legal and Empirical Analysis of the Nebraska Experience (1973-1999)*, 81 *NEB. L. REV.* 486, 557-60 & figs.5-6 (2002).

substantial assistance.⁶¹ In contrast, a slight majority of offices treat providing self-incriminating information as not even relevant to, let alone sufficient for, a 5K1 motion.⁶² The use of self-incrimination also seems to conflict with the text of § 5K1.1. Accordingly, new guidelines could forbid 5K1 reductions for providing self-incriminating evidence, but encourage reductions in exchange for truthful testimony and undercover work. Standing ready to testify or providing investigative leads earns a 5K1 letter between 27.9% and 47.6% of the time.⁶³ Congress could direct the Sentencing Commission to establish criteria for deciding which of these cases merit 5K1 letters. For example, 5K1 letters could be available only where the other available evidence is inadequate to convict a codefendant and the offense level for other crimes prosecuted exceeds some threshold.

Second, Main Justice should adopt national policies on whether prosecutors should recommend particular discounts and, if so, the appropriate range for those discounts. Congress could direct the U.S. Sentencing Commission to adopt policy statements to guide judges' substantial assistance departures and facilitate appellate review. Once again, the policies could build upon Sentencing Commission data describing existing patterns and forms of assistance but also harmonize outliers and build in prescriptive criteria. For example, courts generally reduce cooperators' sentences by a certain percentage instead of tying discounts to the offenses of conviction.⁶⁴ Courts calibrate the percentage discounts to the kind of assistance offered: testimony at trial or before a grand jury typically earns the largest discounts, followed by undercover work, providing tangible evidence, providing verbal information, and lastly merely agreeing to testify.⁶⁵ This ranking makes sense; the greater risks and payoffs of undercover operations and public testimony justify greater rewards. New policy statements could instruct prosecutors to weigh these and other factors, such as the number of other defendants investigated, prosecuted, and convicted, and the offense levels of those defendants.

These policies can only set benchmarks or norms, because cases vary too much for hard rules to fit. Discounts need to fit the degree of danger, degree of effort, and value of the cooperation, and one cannot precisely codify these factors. However, the U.S. Sentencing Commission can harmonize districts' practices by codifying the lessons that it has learned over the last two decades.⁶⁶ Congress should not simply write these criteria itself, short-circuiting the Commission's expertise. Congress should instead direct the Commission to reduce, regulate, and harmonize substantial assistance

61. MAXFIELD & KRAMER, *supra* note 45, at 27 exhb.5.

62. *Id.* at 25 exhb.3.

63. *Id.* at 27 exhb.5 (analyzing a sample of 264 drug-conspiracy members).

64. *See id.* at 18.

65. *See id.* at 32 exhb.10.

66. *See* U.S. SENTENCING GUIDELINES MANUAL ch. 1, pt. A.4(b) (2004) (envisioning eventual codification of the accretion of common law wisdom).

departures, much as Congress has already done with other downward departures.

CONCLUSION

Persistent, blatant regional and local variations require regulation lest they undercut equality, deterrence, and respect for national law. The key is to sort out justified and unjustified types of variation. Policies, values, and probably strategies should be uniform at the national level, and local hostility to the Guidelines should not be an excuse for variation. Nevertheless, locales need freedom to experiment with tactics and respond to crime waves. The law should also promote equality by regulating all actors and sources of sentencing disparity, not just judges. Main Justice should, for example, try to harmonize prosecutorial charging and plea bargaining. Congress should direct the Sentencing Commission to promulgate uniform procedures and substantive criteria in order to harmonize substantial assistance requirements among far-flung districts. Only by supervising and regulating both prosecutorial and judicial discretion can we reduce undesirable local variation while allowing desirable variation to flourish.