

APPELLATE REVIEW OF SENTENCES: RECONSIDERING
DEFERENCE

MICHAEL M. O'HEAR*

ABSTRACT

American appellate courts have long resisted calls that they play a more robust role in the sentencing process, insisting that they must defer to what they characterize as the superior sentencing competence of trial judges. This position is unfortunate insofar as rigorous appellate review might advance uniformity and other rule-of-law values that are threatened by broad trial court discretion. This Article thus provides the first systematic critique of the appellate courts' standard justifications for deferring to trial court sentencing decisions. For instance, these justifications are shown to be based on premises that are inconsistent with empirical research on cognition and decision making. Despite the shortcomings of the standard justifications, this Article suggests that there is a stronger argument for deference that is based on the trial judge's background knowledge regarding the particular circumstances of the local community and courthouse. Even the potential benefits of localization, though, do not clearly outweigh the rule-of-law costs of appellate deference. Thus, this Article concludes with a proposal for a sliding-scale approach to deference that strengthens the appellate role, but also accommodates localization values in the cases in which they are most salient.

* Associate Dean for Research and Professor, Marquette University Law School. Editor, *Federal Sentencing Reporter*. B.A., J.D. Yale University. I am grateful for comments on an earlier draft from Jelani Jefferson Exum, Carissa Byrne Hessick, F. Andrew Hessick, Janie Kim, Susan Mandiberg, Dan Markel, Chad Oldfather, and John Pfaff, and for research assistance from Brian Borkowicz.

TABLE OF CONTENTS

INTRODUCTION	2125
I. THE CASE AGAINST DEFERENCE:	
A. PRELIMINARY VIEW	2130
II. THE CASE FOR DEFERENCE	2135
A. <i>What the Courts Say About Deference</i>	2135
1. <i>There Are No Clearly Correct Answers</i>	2136
2. <i>Trial Judges Have More Experience</i>	2137
3. <i>Trial Judges Have Access to</i> <i>Demeanor Evidence</i>	2141
4. <i>Sentencing Requires Fine-Grained</i> <i>Judgment</i>	2149
5. <i>Sentencing in a Blink</i>	2151
B. <i>The Localization Account</i>	2152
C. <i>Assessing the Case for Deference</i>	2155
III. THE SLIDING-SCALE APPROACH	2158
A. <i>Indicia of Low Transaction Costs</i>	2160
B. <i>Indicia of High Legality Concerns</i>	2161
C. <i>Indicia of Weak Localization Benefits</i>	2162
D. <i>Recap and Illustration</i>	2163
CONCLUSION	2165

INTRODUCTION

More than thirty-five years ago, Judge Marvin Frankel famously denounced sentencing in the United States as “law without order.”¹ In the absence of clear legal standards governing punishment or meaningful appellate review, the unchecked discretion of sentencing judges produced what Frankel characterized as “horrible disparities”²—“a wild array of sentencing judgments without any semblance of the consistency demanded by the ideal of equal justice.”³ In some states, the criticisms of Frankel and other reformers resulted in the adoption of binding sentencing guidelines.⁴ In most states today, however, sentencing occurs without guidelines or with unenforceable, advisory guidelines.⁵ In these states, appellate review might supply what guidelines do not: a safeguard against outlier sentences and a source of reassurance that terms of imprisonment are no longer than demonstrably necessary to advance the basic purposes of punishment. Yet state appellate courts have, almost without exception, shied away from these important tasks.⁶

1. This phrase was the subtitle of Frankel’s influential 1973 book on sentencing. See MARVIN E. FRANKEL, *CRIMINAL SENTENCES: LAW WITHOUT ORDER* (1973). For a discussion of Frankel’s influence, see Michael M. O’Hear, *The Original Intent of Uniformity in Federal Sentencing*, 74 U. CIN. L. REV. 749, 759, 762-63 (2006).

2. FRANKEL, *supra* note 1, at 21.

3. *Id.* at 7.

4. As of 2004, twenty-one states had presumptive or mandatory guidelines systems. Stephanos Bibas & Susan Klein, *The Sixth Amendment and Criminal Sentencing*, 30 CARDOZO L. REV. 775, 786 (2008). In 2004, however, the Supreme Court declared such systems unconstitutional to the extent they relied on judicial fact-finding to determine a presumptive sentencing range. *Blakely v. Washington*, 542 U.S. 296, 303-04 (2004). *Blakely* prompted five states to convert their guidelines from presumptive or mandatory to voluntary. Bibas & Klein, *supra*, at 785-86.

5. As of 2004, seventeen states had no guidelines, and seven (plus the District of Columbia) had advisory guidelines. Bibas & Klein, *supra* note 4, at 785. The Supreme Court’s *Blakely* decision has since caused an additional five states to move into this camp. *Id.* at 785-86.

6. See, e.g., MODEL PENAL CODE: SENTENCING § 7.ZZ cmt. k (Tentative Draft No. 1, 2007) (“No state sentencing system with advisory guidelines has yet produced effective appellate-court scrutiny of trial-court penalties.”). In a forthcoming paper, Professor John Pfaff suggests that Indiana may be an exception to the general rule of anemic appellate sentence review in jurisdictions without mandatory sentencing guidelines. John F. Pfaff, *The Future of Appellate Sentencing Review: Booker in the States*, 93 MARQ. L. REV. (forthcoming 2009) (manuscript at 23, on file with author).

The extraordinary deference given to sentencing judges doubtlessly owes much to the appellate courts' fear that more robust review would invite a sharp increase in the number of sentence appeals.⁷ But the appellate courts themselves defend deference on other grounds, chiefly that trial level judges are better positioned to make good sentencing decisions. The Pennsylvania Supreme Court's reasoning is typical:

The rationale behind such broad discretion and the concomitantly deferential standard of appellate review is that the sentencing court is "in the best position to determine the proper penalty for a particular offense based upon an evaluation of the individual circumstances before it." Simply stated, the sentencing court sentences flesh-and-blood defendants and the nuances of sentencing decisions are difficult to gauge from the cold transcript used upon appellate review. Moreover, the sentencing court enjoys an institutional advantage to appellate review, bringing to its decisions an expertise, experience, and judgment that should not be lightly disturbed.⁸

Although several scholars have called for more robust appellate review of sentences,⁹ the case for deference has not yet been critically examined on its own terms. Instead, scholars focus on rule-of-law concerns: high deference means that each sentencing judge makes his or her own law, and there is no real safeguard against unwarranted disparities.¹⁰ But this criticism does not really address

7. Kevin R. Reitz, *Sentencing Guidelines Systems and Sentencing Appeals: A Comparison of Federal and State Experiences*, 91 NW. U. L. REV. 1441, 1446 (1997).

8. *Commonwealth v. Walls*, 926 A.2d 957, 961 (Pa. 2007) (quoting *Commonwealth v. Ward*, 568 A.2d 1242, 1243 (Pa. 1990)).

9. See, e.g., Carissa Byrne Hessick & F. Andrew Hessick, *Appellate Review of Sentencing Decisions*, 60 ALA. L. REV. 1, 28 (2008) (criticizing appellate deference to policy decisions by sentencing judges); Barry L. Johnson, *Discretion and the Rule of Law in Federal Guidelines Sentencing: Developing Departure Jurisprudence in the Wake of Koon v. United States*, 58 OHIO ST. L.J. 1697, 1738 (1998) (criticizing "deferential abuse of discretion review"); Reitz, *supra* note 7, at 1471, 1480 (criticizing passive role of appellate courts in Pennsylvania's sentencing system). The most recent draft of revisions to the Model Penal Code's sentencing provisions also calls for "review that is more searching than under an abuse-of-discretion standard." MODEL PENAL CODE: SENTENCING § 7.ZZ cmt. g, illus. 1 (Tentative Draft No. 1, 2007).

10. See, e.g., Hessick & Hessick, *supra* note 9, at 31-32 ("[L]eaving substantive sentencing policy to district court discretion ... sacrifice[s] equal treatment and the rule of law.");

the basic claim of the appellate courts: that is, deference results in higher quality sentences that more fully take into account all of the nuances of each case, a value that Professor Kyron Huigens calls “fine-grainedness.”¹¹ If deference commonly results in higher quality sentences in this sense, it is not immediately clear why rule-of-law values should trump fine-grainedness values (or vice versa, for that matter). If we grant the appellate courts’ premise that trial courts are better positioned to make high quality sentencing decisions, we seem to be at an impasse.

But, in truth, the case for deference is not nearly as strong as the appellate courts seem to believe. Indeed, psychological research on cognition and decision making provides reasons to think that appellate courts may actually be *better* positioned than trial courts to make high quality, fine-grained sentencing decisions. For instance, the appellate judge’s reliance on a “cold transcript” may actually *help* by providing insulation from misleading visual cues at the in-person sentencing hearing.¹² On the other hand, moving beyond the vague and dubious clichés invoked by the appellate courts, it is possible to imagine one real advantage that trial judges do have: they are more likely to possess background information not contained in the formal court record regarding relevant local circumstances.¹³ Thus, for instance, trial judges are probably better positioned than appellate judges to adapt sentences to fit community values.¹⁴ But localization is not equally appropriate in all cases. If a defendant’s victims are dispersed across several communities,

Johnson, *supra* note 9, at 1740-45 (discussing tension between rule-of-law interests and deferential abuse-of-discretion review); Reitz, *supra* note 7, at 1471 (identifying disparity as a dominant concern in Pennsylvania’s system of weak appellate review); Ian Weinstein, *The Discontinuous Tradition of Sentencing Discretion: Koon’s Failure To Recognize the Reshaping of Judicial Discretion Under the Guidelines*, 79 B.U. L. REV. 493, 506 (1999) (criticizing adoption of abuse-of-discretion standard in the federal system as creating situations in which “outcomes continue to turn on who does the judging”); Ronald F. Wright, *Rules for Sentencing Revolutions*, 108 YALE L.J. 1355, 1379 (1999) (criticizing proposal for deferential “reasonableness review” in the federal system as creating a system in which “[e]ach trial judge would be free to apply her own chosen principles, in her own reasonable way”).

11. Kyron Huigens, *Solving the Williams Puzzle*, 105 COLUM. L. REV. 1048, 1063-64 (2005).

12. *See infra* Part II.A.3.

13. Sherrilyn A. Ifill, *Judging the Judges: Racial Diversity, Impartiality and Representation on State Trial Courts*, 39 B.C. L. REV. 95, 103-05 (1997).

14. *Id.*

for example, it might seem arbitrary to select the values of any one community to control the sentencing decision.

Once persuasive grounds for appellate deference are identified, it becomes easier to see that the justification for deference is not uniformly strong across the board. Thus, a selective, sliding-scale approach to deference may be an effective way to limit deference to those cases in which it is actually most likely to contribute to high quality decisions, while maintaining an emphasis on rule-of-law values elsewhere.¹⁵

In developing these points, this Article proceeds as follows. Part I identifies in more detail the basic values at stake and takes a first cut at making the case *against* deference. Part II critically analyzes the case that has been made *for* deference and advances a new account based on the trial judge's background knowledge of local circumstances. Finally, Part III proposes the sliding-scale approach as preferable to high deference across the board.

Three caveats are in order before proceeding. First, my concern in this Article is solely with the role of appellate courts in jurisdictions that either employ nonbinding, "advisory" sentencing guidelines, or do not have sentencing guidelines at all. A rich body of scholarly literature addresses the role of appellate courts in policing guidelines compliance in mandatory jurisdictions.¹⁶ However, the

15. A dozen years ago, Professor Cynthia K.Y. Lee proposed a thoughtful sliding-scale approach for appellate deference in connection with "departure" decisions in the then-mandatory federal sentencing guidelines system. Cynthia K.Y. Lee, *A New "Sliding-Scale of Deference" Approach to Abuse of Discretion: Appellate Review of District Court Departures Under the Federal Sentencing Guidelines*, 35 AM. CRIM. L. REV. 1, 41-47 (1997). Her approach emerged from a conclusion, similar to mine, that neither the appellate court nor the trial court has clear superiority with respect to all aspects of all sentencing decisions. *Id.* at 34. Her proposal differs from mine in that hers was directed to the then-mandatory federal system, while mine is designed for use in nonmandatory systems (which would now include the federal system). Additionally, her proposal relied on traditional views regarding the relative competence of trial and appellate courts in deciding questions of law and fact. *See, e.g., id.* at 44 ("Here, nondeferential review makes sense, because the characterization of a ground for departure seems to constitute a question of law rather than a question of fact.... [D]istrict courts are better positioned to evaluate the facts of the case."). My analysis casts doubt on these traditional views of relative institutional competence and instead advances the idea that knowledge of local circumstances supplies the most persuasive argument for superior trial court competence.

16. *See, e.g.,* KATE STITH & JOSÉ A. CABRANES, *FEAR OF JUDGING: SENTENCING GUIDELINES IN THE FEDERAL COURTS* 167-68 (1998); Douglas A. Berman, *Balanced and Purposeful Departures: Fixing a Jurisprudence That Undermines the Federal Sentencing Guidelines*, 76 NOTRE DAME L. REV. 21 (2000); Daniel J. Freed, *Federal Sentencing in the Wake of the*

number of states with mandatory guidelines, which has never been a majority, has actually been declining in recent years,¹⁷ and even the federal system switched from mandatory to advisory in 2005.¹⁸ The current trend away from mandatory guidelines makes a reconsideration of the appellate role in nonmandatory jurisdictions especially timely.¹⁹

Second, my analysis concerns only subconstitutional appellate review. A considerable amount of scholarly work has been done on sentence review under the Cruel and Unusual Punishment Clause of the Eighth Amendment.²⁰ At present, however, the governing

Guidelines: Unacceptable Limits on the Discretion of Sentencers, 101 YALE L.J. 1681, 1727-40, 1747-48 (1992); Johnson, *supra* note 9; Lee, *supra* note 15; Reitz, *supra* note 7; Weinstein, *supra* note 10.

17. See *supra* notes 4-5. Excluding the five states with jury sentencing (which presents a very different set of issues for appellate review than I consider here), the seventeen states without guidelines, as of late 2008, were Florida, Georgia, Idaho, Iowa, Louisiana, Massachusetts, Mississippi, Montana, Nebraska, Nevada, New Hampshire, North Dakota, Rhode Island, South Carolina, South Dakota, West Virginia, and Wyoming. Bibas & Klein, *supra* note 4, at 785 n.44. The twelve states with advisory guidelines were Alabama, California, Delaware, Indiana, Maryland, New Jersey, Ohio, Pennsylvania, Tennessee, Utah, Virginia, and Wisconsin, as well as the District of Columbia. *Id.* at 785 n.45, 786 n.48. In June 2009, Wisconsin effectively abandoned even advisory guidelines and should now be counted among the states without guidelines. See 2009 Wis. Act 28, § 3386m (removing section 973.017(2)(a) of Wisconsin's sentencing statute that required judges to consider guidelines). The analysis of this Article is most immediately relevant to the foregoing twenty-nine states, the District of Columbia, and the federal system. Of course, given recent trends, it is possible that additional states will move into this camp in coming years.

18. *United States v. Booker*, 543 U.S. 220, 245 (2005).

19. The federal system's recent switch from mandatory to advisory guidelines has prompted some other recent scholarly work on the appellate role in advisory systems, with particular attention given to the questions of whether it is appropriate for the federal courts to give a presumption of reasonableness to sentences within the recommended guidelines range and whether district judges should be given the authority to reject policy choices embodied in the federal guidelines. See, e.g., Hessick & Hessick, *supra* note 9, at 18-28. These are important questions for the federal system, but not necessarily for other advisory systems. See, e.g., Michael M. O'Hear, *Appellate Review of Sentence Explanations: Learning from the Wisconsin and Federal Experiences*, 93 MARQ. L. REV. (forthcoming 2009) (manuscript at 37-38) [hereinafter O'Hear, *Appellate Review of Sentence Explanations*], available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1427489 (comparing federal and Wisconsin systems). In any event, the present Article addresses the appellate role at a somewhat higher level of generality.

20. See, e.g., Bradford R. Clark, *Constitutional Structure, Judicial Discretion, and the Eighth Amendment*, 81 NOTRE DAME L. REV. 1149 (2006); Youngjae Lee, *International Consensus as Persuasive Authority in the Eighth Amendment*, 156 U. PA. L. REV. 63 (2007); John F. Stinneford, *The Original Meaning of "Unusual": The Eighth Amendment as a Bar to Cruel Innovation*, 102 NW. U. L. REV. 1739 (2008).

Eighth Amendment jurisprudence does not provide a meaningful basis for the review of noncapital sentences.²¹ If appellate review is to become more robust, the more promising path would appear to be the courts' statutory or inherent authority to review trial court sentencing decisions.

Third, the Article focuses not on the formal label attached to appellate review, but on the actual degree of deference accorded sentencing decisions.²² Most jurisdictions seem to have adopted "abuse of discretion," or some slight variation on that terminology, to describe the formal standard of review for sentences.²³ But abuse of discretion is a notoriously protean standard, which sometimes involves virtually limitless deference and sometimes almost no deference.²⁴ I have no quarrel with continued use of "abuse of discretion" (or any of its close cousins) as the formal standard of review for sentences; my concern is with the content that is given to the label—high deference versus low deference.

I. THE CASE AGAINST DEFERENCE: A PRELIMINARY VIEW

Although scholars have largely neglected the question of how much deference appellate courts should give to trial courts in a jurisdiction without mandatory guidelines, a sizeable body of work addresses the parallel question of how much discretion trial court judges should have in a system with mandatory guidelines.²⁵ In grappling with this question, different scholars use different terminology and reach different bottom-line conclusions, but there is a general consensus as to the fundamental values that are at

21. See, e.g., *Ewing v. California*, 538 U.S. 11, 30-31 (2003) (plurality) (upholding sentence of twenty-five years to life for shoplifting three golf clubs after a challenge on Eighth Amendment grounds).

22. To be clear, I refer here to what is sometimes called "substantive" appellate review. I have written elsewhere regarding "procedural" appellate review, particularly review of the adequacy of the explanation given on the record by the sentencing judge. O'Hear, *Appellate Review of Sentence Explanations*, *supra* note 19; Michael M. O'Hear, *Explaining Sentences*, 36 FLA. ST. U. L. REV. 459 (2009).

23. MODEL PENAL CODE: SENTENCING § 7.ZZ cmt. g (Tentative Draft No. 1, 2007); see, e.g., *Gall v. United States*, 552 U.S. 38, 45-46 (2007) (noting that federal standard of review is abuse of discretion); *State v. Gallion*, 678 N.W.2d 197, 203 (Wis. 2004) (noting that Wisconsin standard of review is erroneous exercise of discretion).

24. Hessick & Hessick, *supra* note 9, at 16.

25. See *supra* notes 9-11 and accompanying text.

stake. These same values also play an important role in assessing the appellate role in a nonmandatory system.

The fundamental, competing values may be characterized as “legality” and “fine-grainedness.”²⁶ *Legality* refers to those traits that we commonly associate with the ideal of rule of law.²⁷ These traits include fair notice of legal norms²⁸ and uniformity in their application—that is, similar treatment of similar cases.²⁹ From a defendant’s perspective, legality thus has both *ex ante* and *ex post* dimensions. On the *ex ante* side, legality values indicate that prospective offenders should be able to discern their real sentencing exposure before committing a crime, defendants engaged in plea bargaining should be able to determine the penal consequences of accepting a particular plea deal, and convicted defendants (and their lawyers) preparing for sentencing should be able to figure out how to tailor their evidence and arguments so as to address what really matters. *Ex post* (that is, after sentencing), defendants should be able to see that their sentences were in line with what other similarly situated defendants received.

Fine-grainedness refers to a high degree of correlation between the judgments of a legal system and community views as to just punishment, taking into account the full range of morally relevant offense and offender characteristics.³⁰ As Professor Huigens argues, fine-grainedness seems no less important than legality to the legitimacy of the criminal justice system: “Given the high stakes involved, morally credible outcomes are a necessity if we are to maintain public support for the legal system. No legal system that routinely punishes morally innocent conduct or that chronically fails to punish morally guilty conduct can expect to survive in the long run.”³¹

26. I borrow these terms from Professor Huigens. Huigens, *supra* note 11, at 1063-64.

27. *Id.* at 1063.

28. *Id.*

29. Although Huigens does not discuss uniformity *per se* in his definition of legality, other scholars have associated uniformity with rule-of-law values, *see, e.g.*, Johnson, *supra* note 9, at 1745, and that connection seems intuitively correct. Indeed, uniformity seems at least implicit in the notion of fair notice of legal norms; if norms are not applied in a reasonably consistent fashion, then actors cannot have fair notice of what is demanded of them.

30. Huigens, *supra* note 11, at 1064.

31. *Id.* The invocation of moral right and wrong seemingly connects fine-grainedness to retributive theories of punishment. Professor Paul Robinson, in particular, has advanced the argument that the perceived legitimacy of the criminal justice system rests on the correlation

The tension between legality and fine-grainedness arises from the basic legalist impulse to abstract away from the particularities of specific cases and make decisions based on generic categories—an analytical process that is apt to cause important case-specific nuances to become lost in the shuffle. This is not to say that legality and fine-grainedness are wholly incompatible, but it is to suggest that pushing hard in one direction likely requires trade-offs in the other. For instance, an emphasis on legality might lead to the adoption of inflexible mandatory guidelines, but because guideline drafters cannot possibly hope to account fully for all of the potentially relevant offense and offender characteristics, fine-grainedness is likely to suffer. Not surprisingly, then, in the guidelines-versus-judges debate, legality is invoked to support guidelines,³² while fine-grainedness is invoked to support judicial discretion.³³

But how do these values play out in the debate over appellate deference? The remainder of this Part will develop the basic argument against high deference based on legality. As for fine-grainedness, the next Part will show that it does not clearly support either side in the debate.

In order to see how legality supports the low deference position, it is helpful to disaggregate the different types of decisions that may go into the selection of a single sentence. Professors Paul Robinson and Barbara Spellman have proposed just such a disaggregated model, which includes six different types of sentencing decisions³⁴: (1) *policymaking*—prioritizing among the fundamental, often

between its judgments and public views of desert. *See, e.g.*, Paul H. Robinson & John M. Darley, *The Utility of Desert*, 91 NW. U. L. REV. 453, 457 (1997). Robinson may or may not be correct, but nothing in my analysis here depends on a public preference for pure desert over other approaches to punishment, such as the traditional utilitarian approach. I assume only that there is some desired range of sentences in any given case that embody a well-informed application of public values, whatever those values may be, to the particular facts of the case, and that a fine-grained sentence is one within that range.

32. *See, e.g.*, FRANKEL, *supra* note 1, at 112-13 (advocating creation of sentencing guidelines based on rule-of-law concerns).

33. *See, e.g.*, Kate Stith & José A. Cabranes, *Judging Under the Federal Sentencing Guidelines*, 91 NW. U. L. REV. 1247, 1282-83 (1997) (criticizing guidelines on the ground that justice requires “judgment that takes account of the complexities of the individual case”).

34. Paul H. Robinson & Barbara A. Spellman, *Sentencing Decisions: Matching the Decisionmaker to the Decision Nature*, 105 COLUM. L. REV. 1124, 1128-29 (2005). I modify their terminology a bit here in order to avoid confusion later on.

competing values of the sentencing system;³⁵ (2) *policy articulation*—translating policy preferences into more specific, categorical principles about what factors are relevant at sentencing and how much weight they should be given;³⁶ (3) *fact finding*—determining the facts regarding offense and offender that help to establish the existence (or nonexistence) of particular sentencing factors;³⁷ (4) *judgment-making*—making the normative judgments regarding the facts that help to establish the existence (or nonexistence) of particular sentencing factors;³⁸ (5) *determining punishment amount*—deciding how severe the sentence ought to be;³⁹ and (6) *determining punishment method*—deciding how to achieve the desired level of severity through some appropriate combination of punishment types, such as incarceration, fine, and supervised release.⁴⁰

With these different aspects of the sentencing decision in mind, it becomes easier to see how robust appellate review of sentences might contribute to legality. Most obviously, the policymaking and policy articulation decisions present opportunities for appellate courts to ensure that sentences are imposed according to clear, uniform principles. Highly deferential appellate review leaves it to each individual sentencing judge to make his or her own policy choices. Uniform results seem much less likely if disparate principles are applied. Fair notice concerns may also be implicated, for policy articulation decisions establish a shadow criminal code of sorts.

By way of illustration, imagine that Judge A decides that drug dealing in a particular neighborhood beset with drug-related violence is a highly aggravating sentencing factor: defendants convicted of dealing in the neighborhood always receive the maximum

35. These values include the traditionally recognized purposes of punishment (retribution, deterrence, incapacitation, and rehabilitation), as well as other systemic values such as uniform treatment of similarly situated offenders and efficient case processing. *See id.* at 1129.

36. *Id.* at 1129-30.

37. *Id.* at 1130.

38. *Id.* For instance, if a court determines that the leaders of a criminal enterprise should receive a longer sentence than those with lesser roles (a policy articulation decision), finding the historic facts about a defendant's role will still leave a separate, normative decision to be made regarding whether these facts establish that the defendant had a sufficiently important position in the enterprise to warrant increased punishment.

39. *Id.* at 1130-31.

40. *Id.* at 1131-32.

sentence from Judge A. Thus, there is a mandatory minimum sentencing law in effect in Judge A's courtroom. But how are drug dealers to obtain notice of this "law"? Trial level sentencing decisions are not commonly published, and Judge A might or might not make any effort to publicize the policy decision.⁴¹ Moreover, Judges B, C, and D in the same courthouse may have made a contrary decision that neighborhood is irrelevant in sentencing drug offenders. Even if somehow aware of Judge A's proclivities, a defendant might still feel that notice of the "mandatory minimum" was unfair because the odds were low that he would be sentenced in Judge A's courtroom. And imagine the defendant's sense of resentment if a fellow dealer from the same neighborhood, fortunate enough to be assigned to Judge B's courtroom, receives a sentence that is only a fraction the length of his own.

By contrast, in a system with robust appellate review, an appellate court could ensure that the same policy decisions govern in every courtroom in a given courthouse. And appellate court decisions, especially those articulating general rules, are commonly published, which would also diminish notice concerns.

It is admittedly less obvious that legality has anything to do with the types of sentencing decisions that are not in the nature of policymaking or policy articulation. The other sorts of decisions in the Robinson-Spellman model tend to be case specific and fact intensive; their resolution seemingly does not lend itself to the development and articulation of principles of general applicability.

But legality concerns are nonetheless implicated. Although the specific information considered for sentencing purposes will never be precisely the same from case to case, similar issues will recur, and the legality principle indicates that they should be resolved in similar ways. For instance, whether a member of a drug-trafficking organization has enough of a leadership role to warrant an enhanced sentence will require the weighing of many case-specific facts, but certain patterns will emerge across cases. Particular defendants will not have exactly the same sets of responsibilities, but there will nonetheless seem no good, principled way of distin-

41. In general, judges need provide little by way of explanation of their sentences and have an incentive to avoid stating their policy decisions because such decisions are most likely to attract critical attention from appellate courts. Reitz, *supra* note 7, at 1445-46.

guishing their blameworthiness. Whether or not a clear, sensible rule can be articulated to draw the line between “leaders” and “followers”—I think probably not⁴²—the legality principle nonetheless indicates that similar cases ought to be treated alike. Appellate courts, by virtue of their smaller numbers of judges and their use of collegial decision-making processes, seem more likely to ensure such uniformity than are the trial courts they supervise. Simply put, in appellate courts, outlier judges are less likely to control outcomes. Moreover, the decisions of appellate courts become readily available benchmarks for use by all judges in future cases, thus advancing the cause of fair notice and diminishing the likelihood of subsequent disparities in similar cases. I suggest, in other words, that appellate courts may advance legality values even when they are not acting in a law-declaration mode. But these values cannot be advanced effectively if appellate review merely rubber-stamps lower court decisions.

II. THE CASE FOR DEFERENCE

The previous Part suggested a legality-based argument *against* deference. This Part now considers the argument *for* deference. First, I summarize and identify weaknesses in the justifications for deference that have been offered by the appellate courts themselves. Next, weaving together some of the most promising themes in the case law and drawing on the concept of localization, I suggest a new justification for deference. Finally, I show that even this case for deference does not clearly justify the legality costs.

A. *What the Courts Say About Deference*

A survey of the sentence review jurisprudence from across all fifty states and the federal system reveals little systematic explanation of the deference principle but the repeated invocation of a handful of loosely related themes. This Section assesses each of these themes in turn.

42. See Stith & Cabranes, *supra* note 33, at 1266-67 (discussing complexity of case law interpreting federal guidelines provision regarding role in the offense).

1. *There Are No Clearly Correct Answers*

One common justification for deference is the indeterminacy of the legal standards governing sentencing. As the Indiana Supreme Court stated,

[W]hether we regard a sentence as appropriate at the end of the day turns on our sense of the culpability of the defendant, the severity of the crime, the damage done to others, and myriad other factors that come to light in a given case. Individual judgments as to the proper balance to be struck among these considerations will necessarily vary from person to person, and judges, whether they sit on trial or appellate benches, are no exception. There is thus no right answer as to the proper sentence in any given case. As a result, the role of an appellate court in reviewing a sentence is unlike its role in reviewing an appeal for legal error or sufficiency of evidence.⁴³

Other state supreme courts have echoed this point.⁴⁴

As a justification for giving high deference to sentencing decisions, though, the court's analysis proves too much. As the saying goes, we are all legal realists these days,⁴⁵ and it is now widely accepted that many "legal" questions reviewed de novo by appellate

43. *Cardwell v. State*, 895 N.E.2d 1219, 1224 (Ind. 2008).

44. *See, e.g., State v. Hodari*, 996 P.2d 1230, 1232 (Alaska 2000) ("[R]easonable judges, confronted with identical facts, can and will differ on what constitutes an appropriate sentence."); *People v. Duran*, 533 P.2d 1116, 1119 (Colo. 1975) ("Sentencing is by its very nature a discretionary decision and is an art, not a science." (citation omitted)); *State v. Broadhead*, 814 P.2d 401, 405 (Idaho 1991) ("Such determinations [of the length of a term of confinement] cannot be made with precision." (quoting *State v. Toohill*, 650 P.2d 707, 710-11 (Idaho Ct. App. 1983))); *State v. Formaro*, 638 N.W.2d 720, 725 (Iowa 2002) ("The application of these [sentencing] goals and factors to an individual case, of course, will not always lead to the same sentence. Yet, this does not mean the choice of one particular sentencing option over another constitutes error. Instead, it explains the discretionary nature of judging and the source of the respect afforded by the appellate process."); *State v. Gallion*, 678 N.W.2d 197, 209 (Wis. 2004) ("We are mindful that the exercise of [sentencing] discretion does not lend itself to mathematical precision."). For a federal case making a similar observation in support of deference, see *United States v. Jones*, 531 F.3d 163, 174 (2d Cir. 2008) ("Sentencing is not, after all, a precise science. Rarely, if ever, do the pertinent facts dictate one and only one appropriate sentence. Rather, the facts may frequently point in different directions so that even experienced district judges may reasonably differ." (citation omitted)).

45. *See, e.g., Michael Steven Green, Legal Realism as Theory of Law*, 46 WM. & MARY L. REV. 1915, 1917 (2005) ("[I]t is often said—indeed so often said that it has become a cliché to call it a 'cliché'—that we are all realists now." (footnote omitted)).

courts lack objectively correct answers.⁴⁶ If indeterminacy alone were the criterion, it is not clear whether *any* decisions would be reviewed nondeferentially.⁴⁷ Observing the indeterminacy of sentencing standards cannot end the analysis; rather, it begs the question of whether sentencing questions are the sorts of indeterminate questions as to which legality values indicate that appellate courts should seek to ensure uniform answers, or whether there are other competing values that outweigh legality concerns.

Indeed, the indeterminacy point is question-begging at an even more fundamental level, for indeterminacy could be seen as less an *explanation* for deference than a *result* of deference. Although statutory sentencing standards may be highly indeterminate, appellate courts could, in interpreting and applying those standards, develop rules that are much less indeterminate. The courts could create rules similar to the vague constitutional standard prohibiting unreasonable searches and seizures, which has spawned a host of far more specific legal rules through judicial processes akin to the development of the common law.⁴⁸ But such processes seem unlikely in the sentencing area as long as appellate courts remain so deferential to sentencing judges.⁴⁹ Again, we are left to wonder what countervailing values warrant the cost in legality.

2. Trial Judges Have More Experience

If indeterminacy alone does not provide a satisfactory basis for distinguishing sentencing from constitutional interpretation and other sorts of “legal” decisions traditionally reviewed *de novo*, then some additional justification must be offered for deference. Typically, the additional justification sounds in competence: appellate judges should defer because trial judges will tend to arrive

46. See Chad M. Oldfather, *Universal De Novo Review*, 77 GEO. WASH. L. REV. 308, 321 n.57 (2009) (“[T]he fact that law declaration is now regarded as one of the twin functions of appellate courts stands as evidence of broad acceptance of the proposition that law is ‘underdeterminate.’”).

47. See *supra* notes 43-44 and accompanying text.

48. See, e.g., *Maryland v. Buie*, 494 U.S. 325, 334 (1990) (holding that police may conduct search incident to arrest without probable cause); *Terry v. Ohio*, 392 U.S. 1, 27 (1967) (holding that police may stop and frisk without probable cause).

49. See *supra* note 8 and accompanying text.

at better answers.⁵⁰ This Section considers the first of three grounds that are offered for this conclusion—experience. The other two grounds, relating to demeanor evidence and fine-grainedness, are considered in the next two Sections.

Experience played an important role, for instance, in the United States Supreme Court's decision in *Gall v. United States*, in which the Court overturned the Eighth Circuit's appellate review standard for sentences as insufficiently deferential.⁵¹ The Court observed that “[d]istrict courts have an institutional advantage over appellate courts in making these sorts of determinations [about whether to go outside the range recommended by the sentencing guidelines], especially as they see so many more Guidelines sentences than appellate courts do.”⁵² The Court noted that district judges sentence an average of 117 defendants per year, with only a small fraction of the sentences being appealed.⁵³ The imbalance between sentences imposed and sentences appealed likely means that the average trial judge sees a greater number and diversity of sentencing fact patterns in any given year than the average appellate judge.

But does this really translate into superior decision-making competence? Although the United States Supreme Court is not alone in suggesting as much, there are good reasons to doubt this conclusion.⁵⁴ For one thing, it is misleading to simply compare the

50. See *supra* note 8 and accompanying text.

51. 552 U.S. 38, 46 (2007) (rejecting the “Court of Appeals’ rule requiring ‘proportional’ justifications for departures from the Guidelines range”).

52. *Id.* at 52 (alteration in original) (internal quotation marks omitted) (quoting *Koon v. United States*, 518 U.S. 81, 98 (1996)).

53. *Id.* at 52 n.7 (citing ADMIN. OFFICE OF U.S. COURTS, 2006 FEDERAL COURT MANAGEMENT STATISTICS 167).

54. See, e.g., *United States v. Martin*, 520 F.3d 87, 92 (1st Cir. 2008) (“[T]he sentencing court possesses a number of institutional advantages, including ... the cumulative experience garnered through the sheer number of district court sentencing proceedings that take place day by day.” (citing *Gall*, 552 U.S. at 51-52)); *United States v. Jones*, 531 F.3d 163, 170-71 (2d Cir. 2008) (identifying, as institutional advantage of sentencing courts, that they “impose scores of sentences each year” (citing *Gall*, 552 U.S. at 51-52)); *United States v. Poynter*, 495 F.3d 349, 351 (6th Cir. 2007) (“And unlike the trial court, most appellate judges have little experience sentencing individuals. While trial judges sentence individuals face to face for a living, we review transcripts for a living.”); *People v. Babcock*, 666 N.W.2d 231, 243 (Mich. 2003) (“Because of the trial court’s ... experience in sentencing, the trial court is better situated than the appellate court to determine whether a [sentencing] departure is warranted in a particular case.”); *Commonwealth v. Walls*, 926 A.2d 957, 961 (Pa. 2007) (“[T]he sentencing court enjoys an institutional advantage to appellate review, bringing to its decisions an expertise, experience, and judgment that should not be lightly disturbed.”).

annual number of sentencing cases seen by trial and appellate judges, at least if the idea is to suggest that trial judges have available a greater storehouse of knowledge regarding other criminal cases. Many appellate judges have prior experience as trial judges or criminal lawyers. For instance, among the eleven active judges of the Eighth Circuit, from which *Gall* emerged, seven have served either as trial court judges or as prosecutors.⁵⁵ Moreover, appellate judges, unlike trial judges, are not limited to their own personal knowledge in deciding cases, but can also draw on the experience of two co-panelists or even more than that on a supreme court.⁵⁶ The collective experience with criminal cases on any given appellate panel may match or exceed the individual experience of many trial judges.⁵⁷ Additionally, it seems likely that there are diminishing marginal returns from experience. Once a judge has seen her first fifty drug cases, it is not at all clear whether the second fifty will add anything helpful to the judge's storehouse of knowledge.⁵⁸

But perhaps there is something about the experience point that goes beyond the availability of knowledge. Perhaps sentencing is a skill that must be practiced regularly in order to achieve the best results. It is commonly said, for instance, that professional athletes get "rusty" when they do not play regularly; perhaps appellate judges suffer a similar disability when they do not have a steady flow of sentences to review.

On the other hand, it is not clear why one would become a better sentencer with regular practice or become a worse sentencer by doing it less frequently. After all, sentencing is not an activity for which there is effective feedback.⁵⁹ Athletes get immediate feedback on the quality of their performance—they get a hit or they strike

55. According to biographical information available through the Federal Judicial Center's website, three Eighth Circuit judges (Murphy, Melloy, and Shepherd) have trial-level judicial experience, while four (Wollman, Bye, Colloton, and Gruender) have prosecutorial experience. Federal Judicial Center, Biographical Directory of Federal Judges, <http://www.fjc.gov/public/home.nsf/hisj> (last visited Mar. 4, 2010) (type in judge's name in the white search box name for a description of the judge's professional experience).

56. *Cf.* Oldfather, *supra* note 46, at 329-30 (explaining role of panels).

57. *Id.*

58. See Barbara A. Spellman, *On the Supposed Expertise of Judges in Evaluating Evidence*, 156 U. PA. L. REV. PENNUMBRA 1, 7 (2007), <http://www.pennumbra.com/responses/03-2007/Spellman.pdf>.

59. *Id.*

out, they make the shot or not, they complete the pass or throw an interception—and they are able to make adjustments based on that feedback. But how is a sentencing judge to know whether he or she is performing well? The basic feedback mechanism within the judicial branch is supplied by appellate review, but high deference undermines the effectiveness of that feedback. Elected judges may also receive occasional feedback from the political system, but voters generally do not have access to anything more than anecdotal information about the sentencing practices of individual judges.⁶⁰ In short, there is no apparent reason to think that sentencing is the sort of activity that lends itself to the development of expertise through regular practice.⁶¹

In the end, the experience point seems no less question-begging than the indeterminacy point. If appellate judges are less competent because they see fewer sentencing cases, they could surely increase their number of sentencing cases by adopting standards of review that are more friendly to appellants. And, even as matters now stand, it is not clear that appellate judges see a materially smaller proportion of sentencing issues than they do any other type of issue. Given high settlement rates across the board in civil and criminal cases, trial judges likely address all types of issues, and not just sentencing issues, with greater frequency than appellate judges—including those types of “legal” issues that are traditionally reviewed *de novo*. In short, we still lack a convincing basis for distinguishing sentencing decisions from those that are accorded much less appellate deference.

60. Judges may also get feedback of a sort when a light sentence is imposed on a defendant and that defendant turns around and commits another crime; this may signal that the judge should have imposed a more severe sentence in the first place for incapacitation or specific deterrence purposes. This is an unsatisfactory feedback mechanism, though, because it is unsystematic. New crimes may or may not come to the judge’s attention, particularly in a multimember trial court in which the new case may be assigned to a different judge. This mechanism is also asymmetric because the judge will get negative feedback on unduly lenient sentences, but not on unduly harsh sentences.

61. *Cf.* Spellman, *supra* note 58, at 7 (“[E]xpertise develops out of ‘many thousands of hours of specific types of practice and training’ — a process called ‘deliberate practice.’ Deliberate practice requires focused programmatic study. It includes appropriate feedback about performance. It includes identifying errors and working on procedures to eliminate them.... And trial judges can sit through hundreds of cases and never do the focused study or have the fast reliable feedback necessary for developing expertise.” (citation omitted)).

3. Trial Judges Have Access to Demeanor Evidence

It is said that “the role of demeanor in assessing witness credibility provides one of the standard, and oldest, justifications for appellate deference to lower court fact finding.”⁶² This general observation holds true in the sentencing context, with appellate courts frequently citing the face-to-face contact between sentencing judge and defendant as a basis for deference.⁶³ Although the courts tend to focus on the importance of the *defendant’s* demeanor in particular, it has been suggested that the demeanor of other witnesses and parties may also be useful in selecting a sentence.⁶⁴

Recalling the different types of decisions that go into the selection of a sentence, demeanor evidence seems most clearly related to fact-

62. Max Minzner, *Detecting Lies Using Demeanor, Bias, and Context*, 29 CARDOZO L. REV. 2557, 2559 (2008).

63. See, e.g., *Gall v. United States*, 552 U.S. 38, 51-52 (2007) (“The sentencing judge has access to, and greater familiarity with, ... the individual defendant before him than the ... appeals court.” (citation omitted)); *United States v. Jones*, 531 F.3d 163, 170-71 (2d Cir. 2008) (identifying that they “hear all the evidence relevant to sentencing, make credibility determinations, and interact directly with the defendant” as institutional advantages of sentencing courts (citing *Gall*, 552 U.S. at 51)); *People v. Duran*, 533 P.2d 1116, 1119 (Colo. 1975) (“In our view, the appellate court is always at a disadvantage because of the fact that it does not have the benefit of the trial court’s contact with the defendant, which is material and relevant in every case.”); *People v. Stacey*, 737 N.E.2d 626, 629 (Ill. 2000) (“The trial judge has the opportunity to weigh such factors as the defendant’s credibility [and] demeanor.”); *State v. Favela*, 911 P.2d 792, 810 (Kan. 1996) (“Thus, deference should be granted to the sentencing judge, who has had first-hand experience with the defendant.”); *People v. Babcock*, 666 N.W.2d 231, 268 (Mich. 2003) (“It is clear that the Legislature has imposed on the trial court the responsibility of making difficult decisions concerning criminal sentencing, *largely on the basis of what has taken place in its direct observation.*” (emphasis added)); *State v. Riley*, 497 N.W.2d 23, 25 (Neb. 1993) (“The appropriateness of a sentence is necessarily a subjective judgment and includes the sentencing judge’s observation of the defendant’s demeanor and attitude.”); *State v. Gallion*, 678 N.W.2d 197, 203 (Wis. 2004) (“[S]entencing decisions of the circuit court are generally afforded a strong presumption of reasonability because the circuit court is best suited to consider the ... demeanor of the convicted defendant.” (citation omitted)).

64. See, e.g., *United States v. Martin*, 520 F.3d 87, 92 (1st Cir. 2008) (“[T]he sentencing court possesses a number of institutional advantages, including ... the opportunity to see and hear the principals and the testimony at first hand.”); *United States v. Poynter*, 495 F.3d 349, 351 (6th Cir. 2007) (“Unlike the trial court, we did not see the defendant, the victims or family members testify at the sentencing hearing.”); *State v. Ahearn*, 300 S.E.2d 689, 697 (N.C. 1983) (“The Fair Sentencing Act was not intended, however, to remove all discretion from our able trial judges. The trial judge should be permitted wide latitude in arriving at the truth as the existence of aggravating and mitigating circumstances, for it is only he who observes the demeanor of the witnesses and hears the testimony.”).

finding decisions, which often require credibility determinations. There seems no good reason, however, why the sentencing judge's access to demeanor evidence should result in deference to other sorts of decisions such as those of the policymaking or policy articulation variety. If it is thought important to defer to demeanor-based decisions, then sentencing judges could be asked to identify their findings of fact separately, as is required, for instance, by Federal Rule of Civil Procedure 52(a) in connection with civil bench trials. Appellate courts might be required to accept these findings as true (unless clearly erroneous), but still have a free hand to reevaluate the significance of the findings. Indeed, deference might be even more narrowly granted to findings expressly based on demeanor evaluation. Because hearsay is permitted for sentencing purposes,⁶⁵ important sentencing facts are often found based on written submissions alone; the availability of demeanor evidence provides no apparent justification for appellate deference to decisions of that type.

More importantly, though, the longstanding view that demeanor evidence results in better decision making flies in the face of an emerging consensus in the legal and social science literature that people generally do a poor job in evaluating demeanor evidence.⁶⁶ Studies suggest, for instance, that lies are accurately detected by observers only about half the time.⁶⁷ Indeed, when confronted with demeanor evidence, test subjects consistently focus on the wrong cues: "People generally believe[] that a reduction in smiling, an increase in furtive glances, fidgeting, and gaze avoidance indicate[] that [a] witness [is] lying. In fact, none of these beliefs [has] support in the social science literature."⁶⁸ Moreover, because of their tendency to focus on these unreliable visual cues, subjects who *watch* witnesses speaking are less accurate in detecting lies than those who read a transcript after the fact.⁶⁹ This suggests that, if anything, appellate judges might do a *better* job of fact-finding than

65. *Williams v. New York*, 337 U.S. 241, 250-51 (1949).

66. Minzner, *supra* note 62, at 2557.

67. *Id.* at 2561.

68. *Id.* at 2562 (citing Jeremy A. Blumenthal, *A Wipe of the Hands, A Lick of the Lips: The Validity of Demeanor Evidence in Assessing Witness Credibility*, 72 NEB. L. REV. 1157, 1194 (1993)).

69. *Id.* (citing Olin Guy Wellborn III, *Demeanor*, 76 CORNELL L. REV. 1075, 1086-87 (1991)).

trial judges *precisely because they do not have access to demeanor evidence*.⁷⁰

To be sure, much of the research on demeanor evidence has involved lay test subjects in experimental settings, and trial judges operating in a courtroom might do better. But the research to date has found little evidence that those who might be thought lie detection “experts,” such as law enforcement officers, forensic psychiatrists, lawyers, and judges, perform much better than laypeople in experimental settings.⁷¹ Moreover, as Professor Chad Oldfather observes, the commentators who have assessed the transferability of the experimental results to the courtroom setting have “uniformly concluded” that the courtroom setting is unlikely to produce substantially better results, and may actually worsen the problem.⁷²

Although lie detection is the task most commonly associated with demeanor evidence, it is possible that such evidence is also used for a slightly different fact-finding task at sentencing: not just deciding between competing versions of historical fact, but also determining the defendant’s current emotional state with respect to the offense. And, indeed, some of the cases do suggest that demeanor evidence may be relevant at sentencing for purposes of remorse detection.⁷³

70. In making a similar point regarding appellate deference generally, Professor Chad Oldfather has helpfully elaborated the advantages of fact-finding based on a transcript. Chad M. Oldfather, *Appellate Courts, Historical Facts, and the Civil-Criminal Distinction*, 57 VAND. L. REV. 437, 451-57 (2004). First, courtroom testimony “is present only for an instant, then disappears,” *id.* at 451, while testimony on the printed page can be reread and reconsidered in light of subsequent testimony, *id.* at 455. Second, “oral language encourages an intuitive and emotional thought process in its hearers.” *Id.* at 453. Relatedly, “limitations on human ability to hold orally communicated information in memory impair the ability to process the information in an intellectually complex fashion.” *Id.* at 454. Finally, use of a transcript helps the appellate court to avoid the pitfall of focusing on unreliable visual cues. *Id.* at 454, 457-58.

71. *Id.* at 458. It is not surprising that experience in viewing demeanor evidence does not translate into expertise in evaluating it, for true expertise emerges from what psychologists call “deliberate practice,” which includes focused study and feedback on performance. Spellman, *supra* note 58, at 7. It is not clear how sentencing judges would get the sort of feedback necessary to make them better at evaluating demeanor evidence.

72. Oldfather, *supra* note 70, at 458-59. Oldfather highlights cross-examination as a potentially important problem in the courtroom setting. *See id.* at 459 (“Cross-examination may have a further distorting effect both because of the typically suspicious tone of the questioning, which has been shown to create a suspicious mindset in observers, and the stress caused in the witness, which may lead her to give off the sorts of behavioral cues that observers improperly associate with deception.”).

73. *See, e.g.*, Commonwealth v. Jones, 613 A.2d 587, 591 (Pa. Super. Ct. 1992) (en banc) (noting superior position of sentencing judge to “view the defendant’s ... displays of remorse,

It is commonly said that genuine contrition is, to quote the United States Sentencing Commission, “a sound indicator of rehabilitative [sic] potential.”⁷⁴ If so, and if the trial judge’s access to demeanor evidence offered remorse detection advantages, then trial judges might be better positioned than appellate judges to determine recidivism risk.⁷⁵

But those are both big “ifs.” Despite the conventional wisdom that present feelings of remorse predict future desistance from crime, very little research has been done to substantiate the remorse-recidivism connection.⁷⁶ To be sure, psychologists recognize that a person’s feelings of guilt regarding something he or she has done tend to induce pro-social behavior (such as efforts to repair the harm the person has caused⁷⁷) and that the discomfort of such guilty feelings may inhibit future transgressions.⁷⁸ There is also evidence that a person’s “proneness ... to guilt” is associated with more constructive responses to anger and interpersonal conflict.⁷⁹ But it is far from clear that the emotion that sentencing judges characterize as “remorse” or “contrition” is the same (or at least has the same behavioral consequences) as the emotion that psychologists label

defiance or indifference”), *quoted in* Commonwealth v. Walls, 926 A.2d 957, 961 (Pa. 2007).

74. U.S. SENTENCING GUIDELINES § B321 cmt. (Preliminary Draft 1986) (1986), *reprinted in* 51 Fed. Reg. 35,080, 35,116 (Oct. 1, 1986).

75. In addition to advancing the utilitarian goal of addressing recidivism risks effectively, remorse detection might also have a role to play in retributive approaches to punishment. More specifically, the emotional suffering that a defendant experiences as a result of remorse may diminish the need for externally imposed punishment in order to achieve the retributive goal of just deserts. Gregg J. Gold & Bernard Weiner, *Remorse, Confession, Group Identity, and Expectancies About Repeating a Transgression*, 22 BASIC & APPLIED SOC. PSYCHOL. 291, 291 (2000).

76. Shadd Maruna & Heith Copes, *What Have We Learned from Five Decades of Neutralization Research*, 32 CRIME & JUST. 221, 256 n.15 (2005).

77. Rob M.A. Nelissen & Marcel Zeelenberg, *When Guilt Evokes Self-Punishment: Evidence for the Existence of a Dobby Effect*, 9 EMOTION 118, 118 (2009).

78. Grazyna Kochanska et al., *Guilt and Effortful Control: Two Mechanisms That Prevent Disruptive Developmental Trajectories*, 97 J. PERSONALITY & SOC. PSYCHOL. 322, 322 (2009). To be more precise, guilt may be an important check on antisocial conduct for individuals who have diminished capacity to exercise deliberate, “effortful control” over their conduct. *Id.* at 330.

79. June Price Tangney et al., *Relation of Shame and Guilt to Constructive Versus Disruptive Responses to Anger Across the Lifespan*, 70 J. PERSONALITY & SOC. PSYCHOL. 797, 806 (1996).

“guilt,” as opposed, for instance, to such emotions as shame and embarrassment, with which guilt is often confused.⁸⁰

The potential conflation of shame and guilt is particularly problematic: “When shamed, a person’s focal concern is with the entire self. Some negative behavior or failure is taken as a reflection of a more global and enduring defect of the self.”⁸¹ When experiencing guilt, by contrast, “a person’s focal concern is with a specific behavior or failure, somewhat apart from the global self.”⁸² Much research indicates that shame and guilt have quite different implications for psychological functioning, whether considered as states at a particular moment in time or as traits (proneness to guilt versus proneness to shame).⁸³ For instance, a recent review of the literature found a “strong, empirically grounded consensus” that “shame-prone individuals are more likely to feel anger and to manage their anger in an unconstructive fashion.”⁸⁴ Whatever may be plausibly hypothesized about a link between guilt and low recidivism risk, it seems unlikely that shame would have the same relationship to recidivism. Yet, the guilt-shame distinction is frequently overlooked even in clinical settings,⁸⁵ and the two emotions are often confused or treated interchangeably.⁸⁶ Confounding matters, guilt and shame are often experienced at the same time, or are felt sequentially with respect to the same incident.⁸⁷ The positive psychological associations with guilt and guilt-proneness seem to manifest themselves only when guilt is not accompanied by shame; thus, the particular emotional state that seems of greatest relevance

80. See June Price Tangney et al., *Are Shame, Guilt, and Embarrassment Distinct Emotions?*, 70 J. PERSONALITY & SOC. PSYCHOL. 1256, 1256 (1996) (“[B]oth psychologists and laypeople may find it difficult to differentiate these three types of affective experiences.”). Another similar, but distinguishable, emotional state is regret. Marcel Zeelenberg & Seger M. Breugelmanns, *The Role of Interpersonal Harm in Distinguishing Regret from Guilt*, 8 EMOTION 589, 589 (2008).

81. Tangney et al., *supra* note 79, at 797.

82. *Id.* at 798.

83. *Id.*

84. Raffaele Rodogno, *Shame and Guilt in Restorative Justice*, 14 PSYCHOL. PUB. POL’Y & L. 142, 155 (2008).

85. Tangney et al., *supra* note 79, at 807.

86. Jeffery S. Ashby et al., *Psychologists’ Labeling of the Affective States of Shame and of Guilt*, 34 PSYCHOTHERAPY 58, 59 (1997).

87. Tangney et al., *supra* note 79, at 801.

to sentencing would be what psychologists call “shame-free guilt.”⁸⁸ There is no reason, however, to think that sentencing judges are consistently seeking out this specific and hard-to-identify emotional state when they are looking for remorse.

Moreover, even assuming that remorse reliably predicts recidivism risk, there are good reasons to doubt that demeanor evidence actually contributes much to successful remorse detection. Critics have long expressed doubt about the capacity of trial judges to distinguish between genuinely felt and purely self-serving expressions of remorse at sentencing.⁸⁹ Although some research suggests that emotional states can be accurately determined by “reading” the expressions on a person’s face, this is a skill that may require considerable training and practice.⁹⁰ Moreover, it seems likely that

88. *See id.* (discussing distinction between “shame-free guilt and guilt-free shame”). To be sure, under the retributive justification for taking remorse into account, *supra* note 75, it might be argued that any painful emotional state, including shame, would suffice to accomplish the punitive objectives. On the other hand, shame may be *too* powerful a negative emotion to advance retributive ends effectively; shame typically triggers strong avoidance responses, including blame-externalization and other-directed anger. Tangney et al., *supra* note 79, at 806. The instability of shame as an emotional state would seem to make it difficult to fit into the just deserts scheme. A possibly related problem with the “retributive remorse” theory is the difficulty of accurate affective forecasting: a considerable body of research now shows that negative emotions tend to be considerably less durable than is commonly supposed, even in response to seemingly major traumas like the death of a spouse. Jeremy A. Blumenthal, *Law and the Emotions: The Problems of Affective Forecasting*, 80 IND. L.J. 155, 166-72 (2005) (summarizing studies). As Professor Jeremy Blumenthal concludes, “On average, we are poor at predicting important elements of future emotional experiences, whether our own or another’s, even minutes into the future.” *Id.* at 172. He explains these results, in part, by reference to the “psychological immune system—a system of cognitive mechanisms that transforms our mental representations of negative events so that they give rise to more positive emotions.” *Id.* at 175-76 (quoting Daniel T. Gilbert et al., *The Trouble with Vronsky: Impact Bias in the Forecasting of Future Affective States*, in *THE WISDOM IN FEELING* 114, 124 (Lisa Feldman Barrett & Peter Salovey eds., 2002)). In any event, for present purposes, the point is that even the most keenly felt sense of shame (or any other negative emotion) at sentencing might moderate with surprising swiftness afterwards, perhaps greatly diminishing its usefulness from a retributive standpoint.

89. *See, e.g.*, *United States v. Hendrix*, 505 F.2d 1233, 1236 (2d Cir. 1974); Michael M. O’Hear, *Remorse, Cooperation, and “Acceptance of Responsibility”: The Structure, Implementation, and Reform of Section 3E1.1 of the Federal Sentencing Guidelines*, 91 NW. U. L. REV. 1507, 1554-56 (1997) [hereinafter O’Hear, *Remorse*].

90. MALCOLM GLADWELL, *BLINK: THE POWER OF THINKING WITHOUT THINKING* 238-39 (2005). Compare, for instance, the picture of a single, untrained judge attempting to discern emotions from a defendant’s in-court performance with the way emotional states are studied by academic researchers: subjects are videotaped (which permits repeat viewing), trained teams of coders seek to identify specific responses from the videotapes, and reliability is

the reliance on misleading visual cues that impair lie detection would also to some extent impair remorse detection; indeed, the remorse detection task may be conceived of as determining whether the defendant is lying when he says “I’m sorry” or “I accept responsibility.” Finally, the practice of rewarding seemingly “genuine” remorse may create unwarranted disparities by benefitting the savvy repeat player (who knows what to say and how to act in the courtroom), while disadvantaging first-timers, the poorly counseled, the mentally ill, the young, the poor, and members of racial and ethnic minority groups.⁹¹

The latter consideration points to a more general problem with the use of demeanor evidence at sentencing for any purpose: the risk that demeanor assessments will be infected by racial bias and will exacerbate the racial disparities that already plague the criminal justice system.⁹² Mock jury experiments indicate that the same prosecution evidence is found more convincing and longer sentences are recommended when the defendant is black than when the defendant is white,⁹³ which is consistent with well documented racial disparities in sentence lengths that do not seem attributable to differences in offense severity or criminal history.⁹⁴ Although the psychological research has not focused on cross-racial lie detection or remorse detection per se, there is some evidence that cross-cultural lie detection is less accurate,⁹⁵ and there are reasons to think that the same phenomenon would be observed in cross-racial settings.⁹⁶ Professor Joseph Rand has thus suggested the existence of a “Demeanor Gap” between the races.⁹⁷

quantitatively assessed across teams of coders. *See, e.g.*, Kochanska et al., *supra* note 78, at 323-24 (discussing coding processes and reliability data from one series of studies on guilt).

91. O’Hear, *Remorse*, *supra* note 89, at 1548-53, 1555.

92. *See* Michael M. O’Hear, *Rethinking Drug Courts: Restorative Justice as a Response to Racial Injustice*, 20 STAN. L. & POL’Y REV. 463, 466-67, 473-77 (2009) [hereinafter O’Hear, *Rethinking Drug Courts*] (discussing magnitude and harmfulness of racial disparities).

93. *See* Chet K.W. Pager, *Blind Justice, Colored Truths and the Veil of Ignorance*, 41 WILLAMETTE L. REV. 373, 393-94 (2005) (summarizing studies).

94. For a discussion of the relevant research, *see* O’Hear, *Rethinking Drug Courts*, *supra* note 92, at 466-72.

95. Pager, *supra* note 93, at 397 (citing Charles F. Bond, Jr. et al., *Lie Detection Across Cultures*, 14 NONVERBAL BEHAV. 189 (1990)).

96. *Id.* at 397-400.

97. Joseph W. Rand, *The Demeanor Gap: Race, Lie Detection, and the Jury*, 33 CONN. L. REV. 1, 4 (2000).

Unconscious group bias likely contributes to these disparities.⁹⁸ However, negative racial stereotypes are less likely to influence decisions if race is unknown, less salient, or discovered only after the processing of other, more relevant information.⁹⁹ Dr. Chet Pager has thus proposed that witnesses testify in court behind screens that would disguise their race.¹⁰⁰ But note that appellate courts already have the functional equivalent of Pager's screen: they do not see defendants, but only read about them through the printed sentencing record. It is possible, of course, that appellate courts will discover the defendant's race through the record—it might be noted, for instance, in reports prepared by police or probation officers, or otherwise be evident from the defendant's transcribed testimony—but, as Pager suggests, it may matter not only *whether* a decision maker knows the defendant's race, but also *when* and *how* race is discovered. Thus, there may be a real difference between being confronted face-to-face by a defendant of another race at the start of the sentencing hearing and learning about the defendant's race as a more abstract datum after other, more relevant information has already been reviewed.¹⁰¹

In sum, whether the fact-finding task is lie detection or remorse detection, access to in-court visual cues (including demeanor and race) seems about as likely to lead the trial judge astray as to facilitate good decision making. Thus, the appellate judge's neces-

98. Pager, *supra* note 93, at 401-11.

99. *See id.* at 432.

100. *Id.* at 375. Although this may sound like a dramatic and unlikely departure from tradition, Malcolm Gladwell reports that a similar revolution has occurred in orchestra auditions in the past thirty years, with screens now commonly being used to shield the identities of auditioning musicians. GLADWELL, *supra* note 90, at 250. During that time period, the number of women in major U.S. orchestras has increased fivefold, *id.*, which suggests that screening may be a powerful way of reducing the effect of bias in decision making.

101. *See* Pager, *supra* note 93, at 431 (noting that “[w]hen actually *faced with* a black witness, the effect on stereotype activation is stronger than merely knowing the witness' race as a separate fact”). Appellate judges are also shielded from the full emotional effect of victim impact evidence. Although the use of victim impact evidence at sentencing has generated concerns that the emotions elicited by such evidence may distort decision making, the psychological research is equivocal and still at a preliminary stage. For a recent summary of the relevant literature, see Bryan Myers et al., *Psychology Weighs in on the Debate Surrounding Victim Impact Statements and Capital Sentencing: Are Emotional Jurors Really Irrational?*, 19 FED. SENT'G REP. 13 (2006).

sary reliance on a transcript may be less a limitation than a source of institutional advantage.

4. Sentencing Requires Fine-Grained Judgment

The complexity of the sentencing decision provides a final theme that is prominent in the appellate courts' discussion of deference. The Maryland high court, for instance, observed, "A criminal sentencing decision is never one easily made, and involves a plethora of considerations, both obvious and subtle. Thus it would be illogical to conduct any review of a sentence using stringent and rigid standards."¹⁰² Other appellate courts around the country have made similar statements.¹⁰³ Such observations echo the "no clear right answer" theme discussed above, but may be advancing an analytically distinct point: it is not so much that sentencing standards are indeterminate per se, but that applying them is difficult work (an "arduous task," as the Iowa Supreme Court put it¹⁰⁴) and involves the weighing of "case-specific detailed factual circumstances."¹⁰⁵ In other words, the courts seem to be connecting the "fine-grainedness" of good sentencing decisions to appellate deference.

It does seem clear enough that the appellate courts reduce the arduousness of their own work by reviewing fine-grained decisions deferentially. What is less clear is why this justifies deference. After all, if reducing workload was always an overriding consideration, appellate courts would review everything deferentially; but they do not do so.

102. *Thomas v. State*, 634 A.2d 1, 7 (Md. 1993).

103. *See, e.g., People v. Duran*, 533 P.2d 1116, 1119 (Colo. 1975) (en banc) ("The trial judge must balance the many facets which enter into a sentencing decision to achieve a result which protects the rights of society and the defendant."); *State v. Formaro*, 638 N.W.2d 720, 725 (Iowa 2002) (referring to the "host of factors that weigh in on the often arduous task of sentencing a criminal offender, including the nature of the offense, the attending circumstances, the age, character, and propensity of the offender, and the chances of reform" (citation omitted)); *People v. Babcock*, 666 N.W.2d 231, 243 (Mich. 2003) ("It is clear that the Legislature has imposed on the trial court the responsibility of making difficult decisions concerning criminal sentencing [T]he question at issue grows out of, and is bounded by, case-specific detailed factual circumstances.").

104. *Formaro*, 638 N.W.2d at 725.

105. *Babcock*, 666 N.W.2d at 243 (quoting *Buford v. United States*, 532 U.S. 59, 65 (2001)).

Lurking in the background may be a vision of specialization and comparative expertise: trial courts concentrate on fine-grained decisions that involve weighing many case-specific facts, while appellate courts specialize in what the Michigan Supreme Court terms “recurring, purely legal matter[s]” and “question[s] readily resolved by reference to general legal principles and standards alone.”¹⁰⁶ Such an argument from comparative expertise features prominently in the assessment of appellate standards of review more generally.¹⁰⁷

But the argument, although familiar, rings hollow in a number of respects. I have already questioned, for instance, whether the sort of feedback mechanisms exist that would allow for the development of true sentencing expertise among trial judges.¹⁰⁸ Moreover, to the extent there is any comparative advantage in making fine-grained decisions, the advantage may well lie with *appellate* courts because they employ group decision-making procedures. The psychological research on decision making indicates that groups are better than individuals at taking multiple factors into account.¹⁰⁹

Even granting the dubious assumption that trial court judges have an expertise that permits them to make higher quality fine-grained decisions than appellate courts, deference would most clearly be justified only as to decisions that are, in fact, fine-grained. Yet, as Robinson and Spellman suggest in their model of sentencing decisions, the selection of a sentence within a discretionary regime may involve policymaking and policy articulation decisions, to which deference would not be justified.¹¹⁰ Moreover, even as to the more case-specific aspects of the sentencing decision, it is not clear that trial courts actually routinely engage in a nuanced balancing of the plethora of potential considerations, as opposed to employing quick, “rough justice” calculations based on instinctive responses to a small number of variables. Indeed, experience in busy urban courtrooms

106. *Id.* (quoting *Buford*, 532 U.S. at 65).

107. Oldfather, *supra* note 46, at 327.

108. *See supra* text accompanying notes 59-61.

109. Robinson & Spellman, *supra* note 34, at 1146 (citing Daniel Gigone & Reid Hastie, *The Impact of Information on Small Group Choice*, 72 J. PERSONALITY & SOC. PSYCHOL. 132, 139 (1997)). As Oldfather has observed, however, the premise that multiple judges participate meaningfully in each appellate decision may have less truth today than in an era of smaller dockets. Oldfather, *supra* note 46, at 329.

110. Robinson & Spellman, *supra* note 34, at 1146.

suggests that the rough justice approach may, if anything, be more typical in practice.¹¹¹ An ideal of fine-grainedness, no matter how attractive in theory, seems a poor justification for across-the-board appellate deference if the ideal is not commonly achieved in practice.

5. *Sentencing in a Blink*

In his 2005 best-seller *Blink: The Power of Thinking Without Thinking*, journalist Malcolm Gladwell popularized the notion that quick, intuitive judgments are often a better basis for decision making than lengthier, more deliberate judgments.¹¹² Although no appellate courts have yet invoked *Blink* as a basis for deference to sentencing decisions, the courts' characterization of sentencing as indeterminate¹¹³ and complex¹¹⁴ might suggest to some an inchoate view that deliberation can do little to improve the quality of sentencing decisions relative to intuition. Indeed, perhaps good sentencing decisions are so dependent on inarticulable factors that the more deliberative processes associated with appellate decision making, especially the generation of formal written opinions through a collegial drafting and editing process, actually put appellate courts at a disadvantage relative to trial courts. One could thus imagine a justification for appellate deference grounded on the view that trial courts are better able to harness what Gladwell calls the "power of thinking without thinking."¹¹⁵

Even Gladwell recognizes, however, that intuitive judgments are often led astray by superficial appearances,¹¹⁶ including race,¹¹⁷ and that the ability to make consistently good snap judgments requires training. As he puts it, we must "manage and educate [our]

111. See, e.g., STEVE BOGIRA, COURTROOM 302: A YEAR BEHIND THE SCENES IN AN AMERICAN CRIMINAL COURTHOUSE 39 (2005) (noting the speedy disposition of cases in a Chicago criminal courtroom and describing one case in which a total of fifteen minutes was spent on accepting a plea and determining a sentence).

112. GLADWELL, *supra* note 90, at 16-17; see also Chris Guthrie et al., *Blinking on the Bench: How Judges Decide Cases*, 93 CORNELL L. REV. 1, 5 (2007).

113. See *supra* text accompanying notes 43-44.

114. See *supra* text accompanying note 8.

115. GLADWELL, *supra* note 90.

116. *Id.* at 72-98.

117. *Id.* at 77-86.

unconscious reactions.”¹¹⁸ In light of these limitations, Professor Chris Guthrie and his coauthors have argued persuasively that *judging* should not be left to purely intuitive approaches.¹¹⁹ Their general arguments seem to carry over to sentencing;¹²⁰ indeed, the arguments closely echo criticisms made above regarding experience-¹²¹ and demeanor-based¹²² justifications for appellate deference. First, Guthrie and his colleagues argue, “intuition is the likely pathway by which undesirable influences, like race, gender, or attractiveness of parties, affect the legal system.”¹²³ Concerns regarding racial bias, of course, have a particular salience in *criminal* cases.¹²⁴ Second, “judges are unlikely to obtain [the] accurate and reliable feedback” on their decisions that is necessary to improve the quality of their intuitive judgments.¹²⁵

Moreover, in addition to such concerns regarding the *substantive* quality of intuitive decisions, there may be other reasons to reject the “blink” model for sentencing. For instance, I have argued elsewhere that *procedural* justice requires that sentences be explained with a degree of rigor that is probably incompatible with quick, intuitive decision making.¹²⁶

In sum, notwithstanding the surprising power of unconscious thinking in some contexts, the “blink” phenomenon seems no more likely to supply a persuasive justification for appellate deference than do the themes that are expressly invoked by the appellate courts.

B. The Localization Account

Let us now consider another, more promising type of advantage enjoyed by the trial judge: his or her background knowledge regarding local circumstances in the courthouse and the community. In explaining the relevance of local circumstances, it may be helpful

118. *Id.* at 16.

119. Guthrie et al., *supra* note 112, at 29-33.

120. *See id.*

121. *See supra* Part II.A.2.

122. *See supra* Part II.A.3.

123. Guthrie et al., *supra* note 112, at 31.

124. O’Hear, *Rethinking Drug Courts*, *supra* note 92, at 466-67.

125. Guthrie et al., *supra* note 112, at 32.

126. O’Hear, *Appellate Review of Sentence Explanations*, *supra* note 19, at 5-7.

to distinguish among the different types of sentencing decisions identified by Robinson and Spellman.¹²⁷ First, consider policymaking and policy articulation.¹²⁸ Trial judges generally work and reside in the communities in which they sentence defendants and so probably tend to be better informed about local values and needs regarding crime and punishment than the appellate judges above them in the court system who serve much larger geographical areas. If the trial judges are elected, of course, there is that much greater reason to presume that the trial judges are in touch with, and responsive to, community values and needs. At the policy level, then, deference to trial judges may serve to adapt sentencing to suit particular communities, in lieu of imposing a top-down, one-size-fits-all approach.¹²⁹ Thus, for instance, if one region in a state favors rehabilitative approaches to punishment, and another favors retributive, both can have their preferences satisfied. Or, if an urban community thinks that unlicensed gun possession is an extraordinarily dangerous and blameworthy crime, and a rural community in the same state thinks that gun ownership is a fundamental right and not appropriate for severe punishment, appellate deference can help to ensure that sentencing practices in both communities conform to local values.¹³⁰

Knowledge of local circumstances also may be helpful with respect to fact-finding decisions.¹³¹ The research on lie detection shows that accuracy improves considerably when test subjects have contextual knowledge from a source other than the witness whose credibility is at issue.¹³² The trial judge's independent knowledge of the community may provide just such useful contextual information. Such knowledge might help one to determine, for instance, whether it is plausible that a street corner drug dealer in a particular

127. See *supra* text accompanying notes 34-40.

128. See *supra* text accompanying notes 35-36.

129. See Vincent L. Broderick, *Local Factors in Sentencing*, 5 FED. SENT'G REP. 314, 314 (1993); see also Michael M. O'Hear, *Localization and Transparency in Sentencing: Reflections on the New Early Disposition Departure*, 27 HAMLINE L. REV. 357, 360-62 (2004).

130. Drawing on public choice theory, I have argued elsewhere that decentralization of sentencing policy in areas where there is intense public disagreement, such as with drug crimes, maximizes overall welfare over the long run (assuming that certain conditions, such as spillover effects, are not present). Michael M. O'Hear, *Federalism and Drug Control*, 57 VAND. L. REV. 783, 856-58 (2004) [hereinafter O'Hear, *Federalism and Drug Control*].

131. See *supra* text accompanying note 37.

132. Minzner, *supra* note 62, at 2568.

neighborhood has no gang affiliation, or whether it is believable that poor road conditions, and not just the defendant's recklessness, contributed to a fatal car accident.

As we move to the final three types of decisions—judgment-making, determining punishment amount, and determining punishment method¹³³—knowledge of values in the local community may still play a role, but knowledge of local courthouse circumstances also becomes important. For instance, in jurisdictions in which it is customary for the prosecutor and/or defense lawyer to offer sentencing recommendations, it may be helpful to know something about the experience and reputation of the attorneys. The attorneys may have private information of their own regarding the circumstances of the defendant, victim, or community, and their recommendations may reflect a thoughtful, well-informed, fine-grained assessment of the case. But, absent prior experience with the attorneys or knowledge of their reputations, it would be difficult for an appellate judge to know which recommendations warrant deference and which are better dismissed as empty posturing. Sentencing recommendations from probation officers present a similar set of issues. Once again, the private knowledge of trial judges regarding the experience and reputation of particular probation officers may give trial judges an important advantage over appellate judges.¹³⁴

Another potentially useful form of local knowledge is the pattern of sentencing practices in a particular courthouse.¹³⁵ The individual judge will certainly have a feel for her own sentencing practices and is likely also to know something about the sentencing practices of

133. See *supra* text accompanying notes 38-40.

134. Another potential advantage lies in the trial judge's greater familiarity with the local media. A trial judge is likely to have a much better sense of media interest in a case and how different sentencing possibilities might be reported to the public. In this way, the trial judge may be better positioned to ensure that sentences in high-profile cases satisfy public preferences and do not result in a loss of public confidence in the criminal justice system. On the other hand, although it seems desirable for sentences to conform generally to community values, it is not clear that media interest in a particular case should affect the outcome. Judicial catering to the media raises concerns about unwarranted sentencing disparities depending on the presence of a reporter in the courtroom and the possibility of pandering to a temporarily inflamed public opinion. See Broderick, *supra* note 129, at 315.

135. See David F. Levi & Mitu Gulati, *Judging Measures*, 77 UMKC L. REV. 381, 409 (2008) (recommending that sentencing data should be collected and provided to judges).

other judges in the same courthouse.¹³⁶ Such knowledge can help to bring about a measure of uniformity within the courthouse.

Finally, when it comes to determining *method* of punishment, local knowledge regarding community-based sanctions may be of great value. For purposes of designing a community-based sanction or deciding that such a sanction is inappropriate, it is very helpful to know the answers to such questions as: Does the local probation office provide effective supervision? Are there good addiction treatment options available in the community? What are the offender's prospects for obtaining good employment or useful vocational training? What is the character of the neighborhood in which the offender resides?

In sum, the trial judge's private knowledge of local circumstances helps to fill in the gaps in the standard account of appellate deference. By giving room for knowledge of local circumstances to be brought to bear in the sentencing process, deference can lead to better sentencing decisions, particularly in the sense that sentences better reflect community values, the practice norms of particular courthouses, and the availability of local resources to support community-based sanctions. But recognizing this potential advantage of deference does not necessarily mean that the sum total of advantages and disadvantages ultimately favors deference.

C. Assessing the Case for Deference

As suggested above, the appellate courts' characterization of sentencing as "difficult"¹³⁷ and "arduous"¹³⁸ points to one advantage of deference: it reduces the appellate courts' transaction costs. A highly deferential approach discourages appeals, demands a less rigorous review of the record when there is an appeal, and provides a ready basis for summary rejection on the merits with little written analysis. But, of course, transaction costs alone cannot justify deference, for transaction costs could be minimized by treating *all*

136. Even if that information is not systematically collected and shared, the judges who share a courthouse routinely talk amongst themselves and with the lawyers who practice in front of them; such informal exchanges supply information about courthouse norms that is less likely available to appellate judges.

137. *People v. Babcock*, 666 N.W.2d 231, 243 (Mich. 2003).

138. *State v. Formaro*, 638 N.W.2d 720, 725 (Iowa 2002).

lower court decisions deferentially. It may be that sentencing decisions tend to be *especially* costly because of the number and variety of factors that ought to be considered; but not all sentencing records are thick. In busy courtrooms, judges often have little more to go on than a charging document, a rap sheet, and a few brief in-court statements from the defendant. In contrast, some “legal” decisions themselves require review of many diverse texts—consider, for instance, the complex weighing of arcane historical sources that can go into originalist constitutional interpretation.¹³⁹ Moreover, although every sentencing case is unique in some respects, there are regularly recurring issues and fact patterns in sentencing,¹⁴⁰ such that an appellate court’s investment of effort in clear, well-reasoned sentencing opinions could save transaction costs in resolving future cases. Thus, although the transaction costs argument may provide some support for deference, it hardly seems a showstopper.¹⁴¹

On the other side, as described in Part I, legality values support nondeferential review.¹⁴² To be sure, as suggested in the previous Section, trial judges may deploy their knowledge of local sentencing practices so as to mitigate the legality costs of deference.¹⁴³ More specifically, there are good reasons to suspect that interjudge disparities within a given courthouse are less significant than intercourthouse disparities. Empirical research, though limited, provides some support for this proposition.¹⁴⁴ But defendants from

139. See, e.g., *Apprendi v. New Jersey*, 530 U.S. 466, 476-82, 501-18, 525-29 (2000) (discussing at length the significance of various eighteenth- and nineteenth-century treatises and cases for the contemporary scope of the right to jury fact-finding beyond a reasonable doubt in criminal cases in the majority, concurring, and dissenting opinions).

140. See, e.g., *United States v. Williams*, 78 F. Supp. 2d 189, 190 (S.D.N.Y. 1999) (“Marvin Williams, the defendant, presents a picture that is all too familiar to any District Judge sitting in an urban court.”).

141. Indeed, despite the redundant transaction costs, the courts of other nations sometimes provide much broader de novo review than American courts. See, e.g., Susan F. Mandiberg, *Why Sentencing by a Judge Satisfies the Right to Jury Trial: A Comparative Look at Blakely and Booker*, 40 MCGEORGE L. REV. 107, 129 (2009) (noting that French appellate courts provide de novo review of both factual and legal determinations).

142. See *supra* text accompanying notes 26-29.

143. See *supra* text accompanying notes 135-36.

144. Cf. U.S. SENTENCING COMM’N, FIFTEEN YEARS OF GUIDELINES SENTENCING 97-98 (2004) (reporting that 12.7 percent of variation in sentence length in federal drug trafficking cases is attributable to the city of prosecution, whereas the primary judge effect, reflecting random assignment of a case to a judge within a district, explains only 1.64 percent of variation).

many different courthouses are typically housed together in prison, and those from harsher courthouses will still know and feel the sting of the broader disparities. Moreover, the fact that the judges within a courthouse may generally adhere to unwritten norms does not address the fair notice concerns that also are encompassed by legality.

With transaction costs pointing in favor of deference and legality concerns pointing against, decisional quality seems the most likely tiebreaker. This is, in fact, the main point relied on by the appellate courts, and when localization values are incorporated into the quality argument, it is at least a plausible position. The trial judges have private knowledge regarding local circumstances, and appellate deference gives room for that knowledge to be brought to bear. If we assume that the quality of a sentencing decision is in large measure a function of its fine-grainedness—a function, that is, of its sensitivity to a diverse range of facts and considerations that are relevant to the blameworthiness of the defendant’s conduct and the appropriate societal response—then local knowledge does indeed seem likely to contribute to decisional quality.

Yet, plausible though the localization theory may be, decisional quality does not provide decisive support for deference either. Indeed, a plausible theory can be advanced that appellate courts are still better positioned than trial courts to reach high quality, fine-grained sentencing decisions, notwithstanding their lesser knowledge of local circumstances. Deference gives trial judges an opportunity to bring helpful private information to bear, but does not assure that they will do so. Instead, trial judges might place too much reliance, consciously or unconsciously, on race and other irrelevant visual cues, as the research on lie detection suggests is commonly done.¹⁴⁵ Likewise, trial judges tend to see a small number of lawyers practicing in front of them repeatedly in criminal cases and might consciously or unconsciously give too much weight to the recommendations of repeat players with whom they wish to retain a good working relationship. Moreover, as I have described in more detail elsewhere, the order in which information is presented in the trial court may tend to produce cognitive biases that lead sentencing

145. *See supra* Part II.A.3.

judges to place too much emphasis on aggravating offense circumstances relative to mitigating offender circumstances.¹⁴⁶

While appellate judges are doubtless subject to various biases of their own, it is not implausible that, on the whole, they are better positioned to render fine-grained judgments based on a full consideration of relevant offense and offender characteristics. I already have noted that transcript-based decisions may be less prone to distortion based on irrelevant visual cues than confrontation-based decisions.¹⁴⁷ More generally, group decision making, as is traditionally employed at the appellate level, is better suited to decisions that require taking multiple factors into account.¹⁴⁸ Each member of the group is able to act as a check on the tendency of the others to overlook or downplay relevant information. With the aid of a written record that can be read and reread in a reflective fashion, appellate courts thus seem to have greater capacity than trial courts to ensure that important information is not overwhelmed by strong emotional or intuitive responses to just a small number of cues. This cognitive advantage, though, may or may not be enough to outweigh the trial court's private information advantage: is it better to have a decision maker with a greater quantity of relevant information, or a decision maker better able to process its lesser amount of information effectively?

In sum, we seem to be at an impasse. Transaction cost concerns support deference, legality concerns point in the opposite direction, and it is not clear which level of court has the decisional-quality advantage.

III. THE SLIDING-SCALE APPROACH

The case for across-the-board appellate deference to trial court sentencing decisions seems less than compelling. At the same time, it is difficult to see a case for across-the-board de novo review that is any stronger. To break the impasse, I propose in this Part a sliding-scale approach to appellate review. Rather than a single,

146. The underlying psychological mechanisms are those that result in first-received information being given more weight than later-received information. O'Hear, *Appellate Review of Sentence Explanations*, *supra* note 19, at 10-12.

147. See *supra* Part II.A.3.

148. Robinson & Spellman, *supra* note 34, at 1146.

across-the-board approach, appellate courts might review sentences with greater or lesser rigor depending on the presence of certain indicia that suggest the case for deference is especially strong or weak. In this way, appellate courts might produce some of the legality and decisional-quality benefits of more rigorous review, but still avoid the added transaction costs of closer scrutiny in those cases where the costs are least likely outweighed by the benefits.¹⁴⁹

To be more specific, the “abuse of discretion” standard commonly used for reviewing sentences has a flexibility that seems quite capable of supporting different levels of deference in different circumstances.¹⁵⁰ I suggest here a default level of high deference that is diminished to the extent that the transaction costs of closer scrutiny are unusually low, the legality concerns of high deference review are unusually compelling, or the localization benefits of

149. The approach I propose—strengthening appellate review to deal with legality concerns but preserving considerable discretion at the trial court level—may appropriately be placed in the context of longstanding calls for appellate courts to develop a common law of sentencing. *See, e.g.,* Reitz, *supra* note 7, at 1500 (“I adhere provisionally to the core of the reformist vision that appellate judges should be encouraged to contribute in thoughtful, precedential, and policy-informed ways to the growth of a substantive common law of sentencing.”). For instance, in their much discussed 1998 proposal to reform federal sentencing, Professor Kate Stith and Judge José Cabranes wrote,

[I]n determining how to constrain judges’ discretion in sentencing we should look to the mechanism that has been used for centuries to impose such constraints in most other judicial matters: requiring trial judges to give reasons for their decisions, and then permitting litigants to seek review of those decisions in appellate courts.... Slowly but surely, a federal common law of sentencing would be created.

STITH & CABRANES, *supra* note 16, at 170. In criticizing their proposal, however, Professor Frank Bowman has questioned whether the judiciary is really willing to do the work required to develop a common law of sentencing. Frank O. Bowman, III, *Fear of Law: Thoughts on Fear of Judging and the State of the Federal Sentencing Guidelines*, 44 ST. LOUIS U. L.J. 299, 323 (2000). Given “too few federal judges and too many federal cases,” Bowman suggests that it may be unrealistic to expect a “persuasive common law of federal sentencing” to arise. *Id.* at 356. My proposal may be thought of as a response to Bowman’s suggestion, and, by extension, to state and federal appellate judges who share the view that they cannot realistically be expected to devote a substantially greater share of their time to sentencing issues. It is possible to identify a limited subset of cases that can be targeted on principled grounds for closer review, without necessarily opening the floodgates to routine *de novo* review of sentences.

150. *See* Lee, *supra* note 15, at 15, 19 (“Some appellate courts use a functional approach to mixed question review, deciding on a case-by-case basis which judicial actor (district judge or appellate judge) is more competent to decide the matter.... Given the wide divergence of opinion as to the meaning of abuse of discretion, use of a functional approach similar to that applied to mixed questions of fact and law has great appeal.”).

deference are unusually weak. As detailed below, various indicia can be identified that tend to signal that at least one of these conditions is present.¹⁵¹

A. Indicia of Low Transaction Costs

When an appellate court reviews a sentence, the court may be confronted by a voluminous record from the lower court, including, for instance, a trial transcript, a comprehensive presentence investigation report, written submissions from the parties, and a lengthy sentencing hearing transcript.¹⁵² But the record need not be nearly so extensive. A case might, for instance, move directly from guilty plea to sentencing, without a presentence investigation report or substantial presentations from the parties.¹⁵³ Where the record is relatively thin, of course, the appellate court absorbs minimal transaction costs in fully reconsidering all of the information presented to the sentencing judge, and deference is accordingly less likely to be justified.

In addition to cases in which there is only a meager record to review, at least one other type of case would seem to present attenuated transaction cost concerns. There are cases that involve a commonly recurring issue that can be resolved through the adoption of a general principle. In such cases, the appellate court's investment of effort to resolve the commonly recurring issue in a clear, definitive fashion may save transaction costs over the long run, as the issue will no longer need to be litigated at the trial or appellate levels. On the other hand, if the issue is a matter as to which different communities in the jurisdiction hold divergent and strongly held beliefs—as might occur, for instance, in drug or gun cases—then localization values may still warrant a high level of deference.

151. In formulating such indicia, of course, it is important to try to avoid numerous or highly indeterminate criteria that would require substantial transaction costs to sort out even before any analysis of the merits takes place.

152. See HARRY I. SUBIN ET AL., *THE PRACTICE OF FEDERAL CRIMINAL LAW: PROSECUTION AND DEFENSE* 485-526 (2006) (describing federal sentencing process and providing samples of documents created in the process).

153. See *supra* note 111.

B. Indicia of High Legality Concerns

Legality concerns focus on uniformity and fair notice.¹⁵⁴ With respect to uniformity, one indicator of concern is a sentence near the very top or bottom of a wide sentencing range. If one assumes that the most extreme sentences are generally reserved for the most extreme cases, then the imposition of an extreme sentence might be taken as a sign that the appellate court should take a careful look to ensure that the case truly is an extreme one, and that the sentence does not result instead from one judge's highly idiosyncratic response to the case. Similar logic would justify closer scrutiny of sentences far above or below a sentencing range recommended by advisory guidelines.¹⁵⁵

Likewise, a sentence to prison, as opposed to a shorter term of incarceration in a local jail or another community-based sanction, might trigger heightened legality concerns. More severe sentences appropriately demand greater reassurance of their conformity to rule-of-law values.¹⁵⁶ Additionally, prisons generally aggregate offenders from across regions within a jurisdiction; thus, prisoners tend to be acutely aware of intercourthouse disparities, which can undermine their perceptions of the legitimacy of the criminal justice system.¹⁵⁷

Fair notice concerns arise when individual sentencing judges follow idiosyncratic principles that are not generally known or readily knowable. These concerns may be diminished to the extent

154. Huigens, *supra* note 11, at 1063.

155. To be sure, appellate courts may run afoul of defendants' Sixth Amendment jury-trial rights if close scrutiny of sentences outside recommended guidelines effectively transforms advisory guidelines into mandatory guidelines. *See supra* note 4. However, the Supreme Court has indicated that sentences far outside the range may permissibly be treated differently than sentences closer to the range. *See Gall v. United States*, 552 U.S. 38, 47 (2007) ("In reviewing the reasonableness of a sentence outside the Guidelines range, appellate courts may therefore take the degree of variance into account and consider the extent of a deviation from the Guidelines.").

156. *Cf. O'Hear, Appellate Review of Sentence Explanations, supra* note 19, at 57-58 (arguing that the relative importance of liberty interests at stake supports a robust explanation requirement for sentences, especially when sentence length exceeds one year).

157. *See Jonathan D. Casper et al., Procedural Justice in Felony Cases*, 22 LAW & SOC'Y REV. 483, 486, 494 (1988) (discussing empirical research indicating that defendants' views of the fairness of the system are influenced by how their outcomes compared with those of other defendants).

that judges clearly articulate the principles on which their sentences rest. The principles are then exposed for public discussion and debate and stand a greater chance of dissemination beyond the courthouse insiders who regularly practice in front of the judge. Conversely, a Delphic sentence, imposed with a few cursory references to the circumstances of the offense or the offender, provides little reassurance that the defendant has had fair notice of what really mattered in the selection of his sentence.¹⁵⁸ For that reason, a sentence that has been explained only in perfunctory or generic terms might appropriately be subject to less deferential review.

C. Indicia of Weak Localization Benefits

If the case for higher decisional quality in the trial court rests on localization values, then cases with weak localization benefits seem relatively ill suited for highly deferential treatment. Thus, for instance, the disposition of cases that are of significant concern beyond the local community, such as the prosecution of the leaders of a statewide drug trafficking enterprise or Ponzi scheme, seems less appropriately subject to community-specific values than, say, routine street crime.¹⁵⁹ Likewise, a sentence that has been explained only by reference to generic considerations and does not expressly invoke local considerations provides a relatively weak case for appellate deference. Indeed, given the tendency of public explanation of decisions to attenuate the effects of some cognitive biases, poorly explained sentences are more likely than others to be influenced by the sorts of biases outlined in the previous Part.¹⁶⁰ Reconsideration by a multijudge panel might make a significant contribution to quality in such cases. Finally, complementing the earlier suggestion that prison sentences are more likely to raise important legality concerns, the rejection of community-based

158. The ancient oracle at Delphi offered famously inscrutable pronouncements. See STRINGFELLOW BARR, *THE WILL OF ZEUS: A HISTORY OF GREECE FROM THE ORIGINS OF HELLENIC CULTURE TO THE DEATH OF ALEXANDER* 81 (1961) (“Specific questions were likely to elicit general answers or answers which ... wore an air of being specific yet irrelevant.”).

159. Cf. O’Hear, *Federalism and Drug Control*, *supra* note 130, at 858-59 (discussing the ability of spillover effects to overcome the general presumption in favor of decentralized decision making indicated by the public-choice model of federalism).

160. O’Hear, *Appellate Review of Sentence Explanations*, *supra* note 19, at 13-14.

sanctions means that searching appellate review is less likely to sacrifice the benefits of the sentencing judge's private knowledge of community supervision and rehabilitation resources.

D. Recap and Illustration

To summarize, the following indicia suggest that the normally high levels of deference shown to sentencing decisions might be appropriately adjusted to provide for more searching appellate review:

- Thin record of sentencing related information developed in trial court
- Presence of commonly recurring sentencing question not previously resolved by appellate court
- Sentence near very top or bottom of wide statutory sentencing range
- Prison term imposed
- Perfunctory or generic explanation of sentence, especially if no reference made to local particularities
- Offense directly affected larger geographical area than judicial district in which sentencing occurred.

To be clear, I do not mean to suggest that any precise formula should govern the consideration of these indicia. Rather, the idea is that the presence of multiple indicia, or perhaps even just one in some cases, should lead to a less deferential stance than normal by the appellate court. Where the indicia are present in an unusually strong way, then the review might appropriately be in the nature of *de novo* reconsideration of the sentence.

Although the sliding-scale approach I propose differs from the across-the-board, high deference typically accorded sentencing decisions, aspects of my approach are occasionally reflected in the existing case law. A good example is the Wisconsin Supreme Court's opinion in *McCleary v. State*.¹⁶¹

In *McCleary*, the sentencing judge imposed an indeterminate prison term of up to nine years on a check forger with no prior criminal history.¹⁶² The Wisconsin Supreme Court found this to be

161. 182 N.W.2d 512 (Wis. 1971).

162. *Id.* at 516.

an abuse of discretion¹⁶³ and reduced the sentence to five years.¹⁶⁴ Although the court's decision was not systematically theorized, the court emphasized various aspects of the case that resonate with the indicia described above: the record was thin,¹⁶⁵ the sentence was nearly at the top of the ten-year statutory range,¹⁶⁶ a prison term was imposed,¹⁶⁷ and only a perfunctory explanation was given for the sentence.¹⁶⁸

Another aspect of the court's analysis also resonates with the sliding-scale approach. As discussed in the previous Part, one good reason for appellate judges to defer to trial judges is that trial judges often have extensive experience with the different probation officers who work in their courthouses, and will thus better be able to assess the weight that should be given to individual probation officers' observations and recommendations.¹⁶⁹ The sentencing judge in *McCleary* did indeed rely on a probation officer's report and recommendation.¹⁷⁰ But this particular probation officer was a "new and inexperienced caseworker, who had, according to the record, no prior experience or training in probation work."¹⁷¹ The record thus happened to make clear that the sentencing judge was not drawing on private knowledge that the probation officer had a good track record, but was instead either following a general policy of deferring

163. *Id.* at 522.

164. *Id.* at 526.

165. *See id.* at 515 (noting that sentencing hearing "afforded the defendant no opportunity to offer any proof in contravention of facts set forth in the presentence report and [did not] afford] the defendant the right of allocution"); *id.* at 524 ("In the instant case, only six pages of the record in the first sentencing were devoted to the testimony of McCleary. It may well be doubted that, on the basis of such brief appearance before the court, any observations could be made that would justify [the] sentence [initially imposed].").

166. *See id.* at 514 (noting the ten-year maximum); *id.* at 524 ("We are satisfied that in the instant case there is no evidence that judicial discretion was exercised in a manner that justifies a near-maximum sentence.").

167. *See id.* at 516.

168. *See id.* ("[T]he judge's reasons for imposing the nine-year sentence were extremely brief."); *id.* at 522 ("The problem in the instant case arises because the trial judge failed to give his reasons why a lengthy, near-maximum sentence was appropriate.").

169. *See supra* Part II.B.

170. *See McCleary*, 182 N.W.2d at 515 (noting that at the sentencing hearing "the trial judge read without comment portions of the presentence report," and quoting trial judge as saying "I intend to follow the recommendation of the Probation Department"); *id.* at 523 ("From this report the trial judge apparently drew the conclusion that the defendant had a [Nietzschean] attitude and considered himself above the law.").

171. *Id.* at 522.

to probation officers or deciding that the content of the probation officer's report was sufficiently compelling to warrant deference in the case at hand. Either way, the sentencing judge was in no better position than the appellate judges to determine how much weight ought to be given to the probation officer's report. The appellate judges could reach their own, de novo conclusion as to that question without fear of losing the benefit of any private local knowledge. And the appellate judges did indeed determine that too much weight had been given to the probation officer's report.¹⁷²

None of this is to say that *McCleary* expressly adopted the proposed sliding-scale approach to deference. Indeed, the opinion is frustratingly vague in a number of respects, such as how the court's critique of the probation officer's work fits into the broader legal analysis. But *McCleary* does suggest that some of the basic concerns underlying the sliding-scale proposal have been shared by the justices of at least one state supreme court.

CONCLUSION

Since at least the time of Judge Frankel's pathbreaking work in the early 1970s,¹⁷³ the legal scholarship on sentencing has been dominated by the guidelines-versus-discretion debate. Ultimately, the debate can have no resolution because each side's position rests on incommensurate values and uncertain empirical assumptions. Guidelines proponents invoke rule-of-law values, while opponents point out that no system of sentencing rules can ever hope to give a proper, morally defensible weight to every sentencing factor in every case. Proponents think that guidelines can get enough of the cases right that discretion can be kept low, while opponents argue that every defendant is a unique human being who is entitled to be considered as such. It seems reasonably clear that mandatory guidelines inevitably result in some cases being handled in a more clumsy fashion than would be done by a judge given wide discretion to pursue fine-grained sentencing. But the frequency of such cases is unknown and perhaps unknowable. Nor is it clear how to weigh

172. See *id.* at 523 ("This writer ... is of the opinion that the philosopher-probation agent was overzealous in attempting to make his past educational experience pertinent to his new job.")

173. See FRANKEL, *supra* note 1.

the fine-grainedness costs of such cases against the legality and other costs of a highly discretionary system. In the end, the guidelines debate breaks down into a question of where the burden of proof is placed, with some preferring to err on the side of rule-of-law values and others preferring to err on the side of ensuring the possibility of fine-grained sentences tailored to the full range of offense and offender characteristics.

With so much attention focused on the ultimately inconclusive guidelines debate, scholars and policymakers have neglected the role of appellate courts in the sentencing process—except insofar as the appellate courts police guidelines compliance. This neglect is unfortunate because robust appellate review of sentences may present the best hope for accommodating both legality and fine-grainedness values. Appellate courts are able to identify and correct outlier sentences, with their decisions then becoming shared benchmarks for lower courts. Appellate courts are also able to develop general rules over time in the areas of policymaking and policy articulation. But, unlike sentencing commissions, appellate courts are not limited to *ex ante* rulemaking and can adapt general rules and distinguish precedents on a case-by-case basis to ensure that sentences take into account all relevant offense and offender characteristics in a fine-grained way.

But stronger appellate review must overcome longstanding traditions of deference, at least outside the minority of states with binding sentencing guidelines. Fortunately, legal scholars and lawyers are becoming increasingly knowledgeable about the psychological research on cognition, which casts much doubt on the familiar grounds given by appellate courts for deference. On the other hand, a stronger argument for deference can be made by reference to transaction costs and sentencing judges' background knowledge of local circumstances. This argument seemingly leads to another impasse.

In order to get around the impasse, I suggest a sliding-scale approach to deference, with appellate review growing more robust to the extent that the transaction costs are low, the legality benefits high, or at least the likelihood of losing the benefits of local knowledge low. Various indicia can be used to help identify when low deference conditions are present.

It is true that the sliding-scale proposal contemplates more work for appellate courts. But the uniquely important interests at stake in criminal punishment surely warrant more thoughtful appellate treatment than is indicated by the repeated invocation of generic clichés about the value of demeanor evidence and the experience of trial judges.

