Solving the Apprendi Puzzle

Kyron Huigens

Benjamin N. Cardozo School of Law, khuigens@yu.edu

Follow this and additional works at: https://larc.cardozo.yu.edu/faculty-articles

Part of the Law Commons

Recommended Citation


Available at: https://larc.cardozo.yu.edu/faculty-articles/357
Solving the *Apprendi* Puzzle

KYRON HUIGENS*

INTRODUCTION

The *Apprendi* puzzle is presented by two conflicting lines of Supreme Court authority, each of which has strong intuitive appeal and substantial constitutional support. On one hand, the Court has held, with only two qualifications, that any fact that the state invokes to justify punishment must be proved beyond a reasonable doubt to a jury,¹ with attendant constitutional trial rights such as notice in charging.² If the Constitution protects the traditional normative architecture of criminal law, then these must be substantial limitations that a state cannot evade with mere changes in form.³ Specifically, the state cannot simply declare that some elements of an offense are now "sentencing factors" that need be proved only to the sentencing judge by a preponderance of the evidence.⁴ On the other hand, the Court has recognized the validity of lodging sentencing discretion in judges, as well as the prerogative of the sentencing judge to exercise discretion on any factual basis or none at all.⁵ Furthermore, the Court has recognized the validity, indeed the desirability, of the state’s making explicit the facts, standards, and procedures on which judges exercise their sentencing discretion.⁶ From this perspective, if a sentencing judge can impose punishment for no apparent reason, then there can be no objection if the state requires that judge to impose punishment more explicitly and deliberately.⁷ Thus, Congress or the states should be able to designate "sentencing factors" and to provide for their proof by a preponderance to the sentencing judge without traditional trial protections and with little, if any, constitutional constraint. The Court has so

---

* Associate Professor of Law, Benjamin N. Cardozo School of Law. I wish to thank Joseph Kennedy, Eric Muller, Arnold Loewy, Ronald Miller, Andrew Taslitz, and Sara Sun Beale for their helpful comments. Thanks also are due to the editors of *The Georgetown Law Journal* for their insightful editorial assistance.

1. See *Apprendi* v. New Jersey, 530 U.S. 466, 490 (2000) ("Other than the fact of a prior conviction, any fact that increases the penalty for a crime beyond the prescribed statutory maximum must be submitted to a jury, and proved beyond a reasonable doubt.").
3. See *In re Winship*, 397 U.S. 358, 365-66 (1970) ("[C]ivil labels and good intentions do not themselves obviate the need for criminal due process safeguards . . . ."); *Mullaney v. Wilbur*, 421 U.S. 684, 698 (1975) ("Moreover, if *Winship* were limited to those facts that constitute a crime as defined by state law, a State could undermine many of the interests that decision sought to protect without effecting any substantive change in its law.").
4. See *Apprendi*, 530 U.S. at 476.
5. *Williams v. New York*, 337 U.S. 241, 251-52 (1949) ("And it is conceded that no federal constitutional objection would have been possible if the judge here had sentenced appellant to death because appellant’s trial manner impressed the judge that appellant was a bad risk for society, or if the judge had sentenced him to death giving no reason at all.").
7. Id. at 92 ("We have some difficulty fathoming why the due process calculus would change simply because the legislature has seen fit to provide sentencing courts with additional guidance.").
held on a number of occasions, with an appropriate emphasis on the principles of legislative primacy and federalism.  

*Apprendi v. New Jersey* is only the latest round in a battle of two Court factions defined by these positions, but it is appropriate for several reasons to call this problem the *Apprendi* puzzle. *Apprendi* is by any measure a landmark case—the point at which the battle turned dramatically in favor of the long-time losers. *Apprendi* holds that any fact, except criminal history, that increases the available punishment for an offense beyond the applicable statutory maximum penalty must be determined by a jury on proof beyond a reasonable doubt. During the two decades before *Apprendi* in which the question seemed to be settled to the contrary, courts and legislatures supposed that a valid sentence could be imposed on the basis of some facts, not limited to criminal history, that had been proved to the sentencing judge only by a preponderance. The *Apprendi* decision immediately affected thousands of sentences in cases pending appeal. Furthermore, because the *Apprendi* principle is almost certainly retroactive, thousands of additional convictions may soon be vulnerable to collateral attack on habeas corpus. But this is only the beginning of the problem.

---

8. *See id.* at 86 ("[W]e should hesitate to conclude that due process bars the State from pursuing its chosen course in the area of defining crimes and prescribing penalties." (citing *Patterson v. New York*, 432 U.S. 197, 201–02 (1977))).


10. *See id.* at 476 (Stevens, J.) ("At stake in this case are constitutional protections of surpassing importance . . ."); *id.* at 524 (O'Connor, J., dissenting) (describing *Apprendi* as "a watershed change in constitutional law").


12. *Id.* at 564–65 (Breyer, J., dissenting) ("For another thing, this Court's case law . . . led legislatures to believe that they were permitted to increase a statutory maximum sentence based on a sentencing factor.").

13. The proportion of pending appeals that were affected by *Apprendi* is difficult to estimate, but even ten percent would be a substantial number. About 70,000 criminal appeals are pending at any given time. The most recent figures available list the total number of mandatory appeals pending in state intermediate appellate courts (most of which are criminal cases) as 47,336 in 1996; 51,109 in 1997; and 48,015 in 1998. *Nat'l Council of State Courts, Examining the Work of State Courts, 1998*, at 96 (1999). The total number of discretionary appeals pending in state courts during the same periods was 17,174 in 1996; 17,847 in 1997; and 18,647 in 1998. *Id.* at 97. The number of felony convictions in federal courts is four percent of the number of state convictions. *Bureau of Justice Statistics, U.S. Dep't of Justice, Felony Sentences in State Courts, 1996*, at 2 (1999). So, in addition to the approximately 67,000 criminal appeals pending in state courts, there are about another 3000 pending in federal court.

14. *See infra* Part III.F.

Neither of the two limitations set by *Apprendi*, pertaining to criminal history\(^\text{16}\) and statutory maxima,\(^\text{17}\) respectively, can stand much scrutiny. Each of the limitations is vulnerable to elimination once an appropriate vehicle for overturning its supporting precedent arrives at the Court, and some Justices already have signaled their readiness to do this.\(^\text{18}\) If this were to happen, both the Federal Sentencing Guidelines and many state sentencing schemes would be rendered unconstitutional.\(^\text{19}\) All told, hundreds of thousands of sentences would have to be reviewed and possibly vacated.\(^\text{20}\) Among the beneficiaries would be some of the most dangerous people in the country.

The solution offered here is a solution to the conceptual puzzle at the heart of *Apprendi*. There is a principled basis on which to distinguish a jury’s determination of facts in support of punishment from a judge’s determination of facts in support of punishment. The key is to understand what each institution is doing when it makes its factual determinations and the constitutional implications of fact-finding in these different functional contexts. The solution offered below pursues the *Apprendi* Court’s principled concern for fundamental fairness and the preservation of the traditional normative architecture of the criminal law. The solution rebuts the argument that the *Apprendi* dissenters and their predecessors on the Court have relied upon for half a century.

\(\text{16. } *Apprendi*, 530 U.S. at 490.\)
\(\text{17. } *Id.* at 486–87 & n.13.\)
\(\text{18. See } *id.* at 487 n.13 (Stevens, J.) (expressly leaving open the possibility that } *McMillan* \text{ may be overruled); } *id.* at 489 (stating that “it is arguable that } *Almendarez-Torres* \text{ was incorrectly decided, and that a logical application of our reasoning today should apply if the recidivist issue were contested”); *id.* at 499 (Scalia, J., concurring) (stating that the Constitution requires the broad rule “that all the facts which must exist in order to subject the defendant to a legally prescribed punishment must be found by the jury’’); *id.* at 520–22 (Thomas, J., concurring) (stating his belief that the } *Almendarez-Torres* \text{ and } *McMillan* \text{ cases were incorrectly decided); see also infra Part I.A (describing the relationship between } *McMillan* \text{ and } *Almendarez-Torres*, \text{ and the two limitations contained in the rule of } *Apprendi*.\)
\(\text{19. See } *id.* at 550–51 (O’Connor, J., dissenting) (noting that “the Court does not say whether [federal and state determinate sentencing] schemes are constitutional, but its reasoning strongly suggests that they are not’’).\)
\(\text{20. The total number of federal prisoners at the end of 1999 was 135,246. } *Bureau of Justice Statistics*, \text{U.S. Dep’t of Justice, Prisoners in 1999, at 1 (2000). The vast majority of these prisoners were sentenced under the Federal Sentencing Guidelines, and all of these would be affected by the extension of } *Apprendi*. \text{The total number of state prisoners at the end of 1999 was 1,231,475. } *Id.* \text{Not all of these prisoners were sentenced under a mandatory sentencing system. However, the number of these prisoners that would be affected by an extension of } *Apprendi* \text{ is likely to be substantial. Fourteen states, including California, Florida, and Illinois, have determinate sentencing systems similar to the Federal Sentencing Guidelines. *Bureau of Justice Assistance*, \text{U.S. Dep’t of Justice, 1996 National Survey of State Sentencing Structures 4–5 (1998). All fifty states and the District of Columbia have mandatory-minimum sentencing for at least some offenses. } *Id.* \text{In 1994, Congress passed the Truth in Sentencing Act, which encouraged states to adopt mandatory sentencing systems. See Violent Crime Control and Law Enforcement Act of 1994, Pub. L. No. 103-322, §§ 20102–20103, 108 Stat. 1796, 1816–17 (1994) (creating the Violent Offender Incarceration and Truth-in-Sentencing Incentive Grants Program). The Act has had a substantial impact on prison populations in those states. *Bureau of Justice Statistics*, \text{U.S. Dep’t of Justice, Truth in Sentencing in State Prisons 3–5 (1999); see also Bureau of Justice Assistance, U.S. Dep’t of Justice, National Assessment of Structured Sentencing 100–13 (1996) (describing impact of state structured sentencing systems on incarceration rates and prison crowding).} \)
The practical effect of my solution to the *Apprendi* puzzle, were the Court to adopt it, would be to bring on the massive difficulties that *Apprendi*’s detractors fear. This prospect is no reason not to solve the puzzle. It seems likely that the Court will take an incremental approach to solving its difficulties in this area, instead of the global approach that my argument suggests.\(^{21}\) Even so, the eventual resolution of these difficulties will require an understanding of the principles at the heart of the problem. Furthermore, I hope to persuade that whatever the cost of a principled solution to the *Apprendi* puzzle may be, we ought to pay it. If this attitude toward the actual social cost of theoretical consistency seems cavalier, let me attempt to dispel that impression in two steps: first, by indicating how far things have gone wrong; and second, by suggesting a reason, elaborated upon at the end of this Article, why the practical cost of my solution to the *Apprendi* puzzle is not as great as one might fear.

In one of the first articles to appear in *Apprendi*’s wake,\(^{22}\) Stephanos Bibas argues that the practical effect of *Apprendi*’s holding is to make federal defendants\(^ {23}\) worse off. He points out that the combination of the Federal Sentencing Guidelines’ “real offense” approach to sentencing and the incentives to plead guilty created by the Guidelines leads to a situation in which the defendant’s only genuine opportunity to litigate the facts of the case occurs at sentencing.\(^ {24}\) Bibas makes a persuasive case that *Apprendi* eliminates this opportunity, even while it does little to change the incentives to plead guilty.\(^ {25}\) The result is that the facts of federal cases may never be litigated at all. Bibas thus concludes that *Apprendi* should be rejected at the first opportunity, because it fails to reflect the “real world” of criminal justice in which plea bargaining predominates over the

---

\(^{21}\) See Michaels, *supra* note 15, at 323–25 (describing *Apprendi* as an incremental step toward a broader rule).


\(^{23}\) Bibas repeatedly refers to “similar state [sentencing] systems” in an effort to broaden the impact of his thesis. *E.g.*, *id.* at 1153. The fact is, however, that there are no similar state systems. The states that have adopted determinate sentencing in general have not adopted the Guidelines’ real-offense approach to sentencing. *See* United States v. Shonubi, 103 F.3d 1085, 1088 (2d Cir. 1997). Bibas’s thesis, as a result, is limited in its application to federal cases.

\(^{24}\) Determinate sentencing schemes such as the Guidelines are meant to eliminate disparity in sentences by eliminating discretion in sentencing. But if a sentence follows rigidly, according to formula, from the offense of conviction, then discretion is not eliminated; it simply is shifted from the judge to the prosecutor, who can engage in “sentence bargaining.” That is, the charge to which the defendant pleads will be the sole determinant of his sentence, and the prosecutor controls the charge, or charges, offered to the defendant. For this reason, the Guidelines operate on the principle of real-offense sentencing, under which “relevant conduct” surrounding the offense of conviction—including uncharged crimes and “crimes” of which the defendant has been acquitted—is proved at a mini-trial for purposes of sentencing. The theory is that if the sentence turns on these additional facts, instead of the charge alone, then the prosecutor will not be able to exercise exclusive control over the sentence that the defendant will receive. *See* Kate Stith & José A. Cabranes, *Fear of Judging: Sentencing Guidelines in the Federal Courts* 132–33 (1998).

\(^{25}\) Bibas, *supra* note 22, at 1152–68.
trial of criminal cases.\textsuperscript{26}

I draw quite a different conclusion from Bibas’s arguments.\textit{Apprendi} attempts to restore the traditional normative architecture of the criminal law. If the result of this effort is to make defendants in the federal criminal justice system worse off, then we have a system that, simply put, is not a criminal justice system at all. If the combination of plea bargaining and the Guidelines approach to sentencing not only does not have, but also cannot be reconciled with, the attributes of a recognizable system of just punishment, then the federal system is merely an administrative regime for the efficient processing of undesirables.

The actual operation\textsuperscript{27} of the federal system before\textit{Apprendi} bears this interpretation even without the effects that Bibas anticipates.\textsuperscript{28} We have created a monster,\textsuperscript{29} and we ought to destroy it.

\begin{itemize}
\item \textsuperscript{26} Id. at 1100 ("\textit{Apprendi} is symptomatic of criminal procedure’s preoccupation with jury trials at the expense of the real world of guilty pleas and sentencing.").
\item \textsuperscript{27} Some of the principal features of Guidelines sentencing are simply shocking to the unindoctrinated in their gross variance from established criminal justice norms. See STITH \& CABRANES, supra note 24, passim. The Guidelines’ real-offense sentencing philosophy permits offenders to be sentenced for "guidelines crimes" that are not part of any democratically legislated criminal code, but are instead ad hoc assemblages of facts that probation officers and prosecutors have deemed relevant to the actual crime of conviction. The defendant can be convicted of these "guidelines crimes" without the benefit of notice, a grand jury indictment, rules of evidence, proof beyond a reasonable doubt, or any meaningful opportunity to appeal. Federal prosecutors have frankly admitted that they have pursued and obtained convictions on lesser offenses in order to sentence the offender to a more serious "guidelines crime"—such as murder—that they could not prove as a genuine crime. See infra note 164 and accompanying text. Furthermore, a "guidelines crime" can be a crime of which the offender previously has been acquitted. See United States v. Watts, 519 U.S. 148, 149 (1997) (holding that a guidelines crime premised on acquitted conduct need be proved only by a preponderance); Witte v. United States, 515 U.S. 389, 399–400 (1995) (holding that double jeopardy is no bar to sentencing on acquitted criminal conduct).
\item \textsuperscript{28} STITH \& CABRANES, supra note 24, at 103 ("Without principled foundation or application, the awesome power of the state to inflict suffering is wielded as an exercise in bureaucratic regularity, for which no one, ultimately, bears responsibility."); see also Paul H. Robinson, \textit{Punishing Dangerousness: Cloaking Preventive Detention as Criminal Justice}, 114 HARV. L. REV. 1429, 1431–32 (2001) (arguing that the Guidelines, among other sentencing schemes, falsely presents purely prophylactic incapacitation as deserved punishment).
\end{itemize}
The cost of abolishing the Federal Sentencing Guidelines and the other mandatory sentencing schemes probably would be high, but it would be astronomical if this abolition were carried out retroactively. Retroactive abolition would entail the reversal of hundreds of thousands of final judgments and lead to resentencing or release in all of them. Ordinarily, new rules of law are not applied retroactively to final judgments, but exceptions exist for new rules that embody fundamental principles of criminal law or that protect essential procedural rights. My argument appeals to the fundamental requirements of just punishment in order to restore both the due process right to proof beyond a reasonable doubt and the Sixth Amendment jury right. The rule that I advocate, then, would seem to apply retroactively. However, the exceptions to the rule of nonretroactivity are meant to identify the unusual constitutional rule on which long-standing consensus exists and that has been articulated as a new rule only because it has never before been challenged. My solution does not run afoul of these exceptions, because I do not seek to restore a constitutional consensus. My argument is that a long-standing consensus should be dissolved in order to free the Court from a stalemate. The *Apprendi* puzzle arises from a misconceived theory of punishment that the Court ought to abandon in favor of a more adequate one. Were the Court to adopt my solution to the *Apprendi* puzzle, it would mark a significant departure in the Court’s jurisprudence that could apply only prospectively. If so, then the cost of my solution would be lower than one might fear.

The holdings of both *Apprendi* and *McMillan v. Pennsylvania*—the leading statement of the rule contrary to *Apprendi*’s—easily can be reduced to absurdities. As Part I argues, both factions on the Court have performed this exercise on the other’s position, but neither side seems to have recognized the obvious implication of this exchange: that the problem has been misconceived by all concerned. Both sides recognize that the question comes down to apportioning the respective responsibilities of the jury and the judge in finding the facts that bear on punishment. Their debate has centered on the historical practice of committing some category of facts to one or another of these decisionmakers. But a categorical approach is doomed because juries and judges determine the same or similar categories of facts. The difference between them is the functional context in which each institution makes such determinations.

The Court has missed this point because of its notorious reluctance to constitutionalize the substantive criminal law—a reluctance that seems to...
involve a deep-seated aversion to the theory of punishment. With the decision in Apprendi, the time has come to take theory seriously. Without greater conceptual clarity, the Court will continue to fumble this issue. Part II therefore introduces some fundamental distinctions and concepts from the theory of punishment and reconceives the problem on this basis. Part III solves the Apprendi puzzle by these means.

The jury determines facts in the context of determining criminal fault. The judge determines the same or similar facts in the context of determining a proportionate punishment. Fault in wrongdoing and proportionality in punishment can be distinguished, as can the respective institutional roles of juries and judges in finding facts that bear on punishment. Part III goes on to elaborate and defend this argument by drawing the implications of this solution for the Federal Sentencing Guidelines, as well as for state mandatory-minimum and "three strikes" sentencing schemes.

I. THE PUZZLE

By the time Apprendi was decided, the issue it presented was finely poised between two equally defensible positions. Unfortunately, the Apprendi opinions did little more than heighten the conceptual tension.

A. BUILDING THE PUZZLE

The issue behind the Apprendi puzzle is the constitutional definition of a criminal offense. The issue arises in two different ways. In some cases, such as Apprendi itself, a legislature has taken a fact that arguably is an element of a crime and has designated it a "sentencing factor" instead. Because sentencing receives less constitutional regulation than the criminal trial, this designation has the effects of relieving the prosecution of the burden of proving its case beyond a reasonable doubt, removing the question from the jury, and depriving the defendant of other trial rights such as notice in the indictment and the right to confront witnesses. The same issue arises when the legislature takes a fact that arguably is an element of a crime and designates it an "affirmative defense." This has the effect of shifting the burden of persuasion on the issue to the defendant, again relieving the prosecution of the burden of proving its case beyond a reasonable doubt.

The notion of a sentencing factor has its roots in the 1949 case of Williams v. New York. In the course of sentencing Williams to death for murder, the trial

---

32. See Kyron Huigens, Rethinking the Penalty Phase, 32 ARIZ. ST. L.J. 1195, 1220-28 (2000).
35. 337 U.S. 241 (1949).
judge relied on a probation report that, among other things, accused Williams of having committed thirty burglaries in the vicinity of the murder, which itself had occurred during a burglary. On appeal, Williams objected to this information being used to sentence him, on due process grounds. The Supreme Court affirmed the death sentence. The Court reasoned that the trial judge had discretion in sentencing and that "no federal constitutional objection would have been possible if the judge here had sentenced appellant to death because appellant's trial manner impressed the judge that the appellant was a bad risk for society, or if the judge had sentenced him to death giving no reason at all." This being the case, there could be no constitutional objection to a procedure that made the grounds of Williams's sentence explicit. Nor was there any question for the Williams Court that the trial judge's discretion in sentencing was well-founded. The Court described that discretion as a valid feature of modern indeterminate sentencing schemes, under which the legislature sets a broad sentencing range, a judge sets a maximum sentence within this range, and a probation department determines the actual amount of time served by an inmate, depending on his progress in reformation and rehabilitation. "Retribution," the Court wrote, "is no longer the dominant objective of the criminal law. Reformation and rehabilitation have become important goals of criminal jurisprudence."

The Court seemed to contradict Williams a few years later in Specht v. Patterson. Specht was convicted in Colorado of indecent liberties, a crime carrying a maximum penalty of ten years. But Specht was not sentenced for that offense. Instead, he was sentenced under a separate statute, the Sex Offenders Act, to an indeterminate term of one day to life. This sentence rested on a finding by the trial court, based on a probation report and entered without a hearing, to the effect that Specht was a danger to the public. Specht challenged this sentence on due process grounds, and the Court reversed. Even though the purpose of the Act was to incapacitate sex offenders, the Court held, the sentence constituted punishment. That being so, due process required the presence of counsel, a hearing, and the confrontation of witnesses. The Court distinguished this case from the "radically different situation" presented in Williams, apparently on the ground that Specht involved a new and formally distinct proceeding to determine the defendant's dangerousness, instead of an

36. Id. at 245.
37. Id. at 252.
38. Id.
39. Id. at 248–49.
40. Id. at 248.
41. 386 U.S. 605 (1967).
43. Specht, 386 U.S. at 607–08.
44. See id. at 608–09.
45. Id. at 609–10.
46. Id. at 608.
ordinary sentencing.\textsuperscript{47} But it is difficult to see this distinction as making a constitutional difference. Both cases involved the imposition of punishment. \textit{Williams}'s principal rationale was that if a trial judge may impose a sentence without stating any reason, then the defendant is no worse off—indeed, he is probably better off—if the sentencing judge makes her reasons explicit. This rationale covers the standards and procedures at issue in \textit{Specht} no less than the sentencing procedure at issue in \textit{Williams}.

\textit{In re Winship}\textsuperscript{48} seemed to cement the broad principle underlying \textit{Specht}: that the Due Process Clause protects the traditional normative architecture of the criminal law. The \textit{Winship} Court held that "the Due Process Clause protects the accused against conviction except upon proof beyond a reasonable doubt of every fact necessary to constitute the crime with which he is charged."\textsuperscript{49} In its opinion, the Court stressed the fundamental unfairness of imposing criminal liability on civil standards of proof.\textsuperscript{50} It relied on the reasonable doubt standard's status under common law as an essential safeguard against erroneous deprivations of liberty,\textsuperscript{51} and it quoted Justice Frankfurter's assertion that the standard is "rightly one of the boasts of a free society."\textsuperscript{52} Likewise, Justice Harlan wrote in his concurrence that the criminal law is "bottomed on a fundamental value determination of our society that it is far worse to convict an innocent man than to let a guilty man go free."\textsuperscript{53}

\textit{Winship} stands not only for the constitutionalization of the reasonable doubt rule, but also for the rule that this requirement cannot be evaded by formalistic distinctions. Samuel Winship was a juvenile. He was convicted on proof by a preponderance of an act that "if done by an adult, would constitute a crime."\textsuperscript{54} In response to the State's contention that this was not a crime, but a matter of juvenile delinquency, the Court stressed that "civil labels and good intentions do not themselves obviate the need for criminal due process safeguards in juvenile courts."\textsuperscript{55}

\textit{Mullaney v. Wilbur}\textsuperscript{56} lined up squarely behind \textit{Specht} and \textit{Winship} even though the issue was not the elements of an offense versus sentencing factors, but instead the elements of an offense versus an affirmative defense. In \textit{Mullaney}, Maine had defined murder and manslaughter as two degrees of felonious homicide. A felonious homicide consisted of an unlawful (that is, neither justified nor excused) killing, done with intent. The felonious homicide would be punished as murder unless the defendant persuaded the jury, by a preponder-

\begin{itemize}
\item \textsuperscript{47} Id. at 607--08.
\item \textsuperscript{48} 397 U.S. 358 (1970).
\item \textsuperscript{49} Id. at 364.
\item \textsuperscript{50} See id. at 363--64.
\item \textsuperscript{51} See id. at 361--63.
\item \textsuperscript{52} Id. at 362 (quoting \textit{Leland v. Oregon}, 343 U.S. 790, 802 (1952) (Frankfurter, J., dissenting)).
\item \textsuperscript{53} Id. at 372 (Harlan, J., concurring).
\item \textsuperscript{54} \textit{Winship}, 397 U.S. at 360 (quoting N.Y. FAM. CT. ACT § 712).
\item \textsuperscript{55} Id. at 365--66.
\item \textsuperscript{56} 421 U.S. 684 (1975).
\end{itemize}
ance, that he had acted in the heat of passion; or, to put the same point in converse terms, that he had not acted with malice aforethought.\(^{57}\) If the defendant could carry this burden, then he would be punished for manslaughter instead of murder. The Supreme Court held that this Maine law violated due process because it placed the burden of proving an essential element of the crime of murder—malice aforethought—on the defendant.\(^ {58}\)

Even though Specht, Winship, and Mullaney seemed to have established a strong line of precedent, the line they established was in significant tension with Williams. The Mullaney opinion distinguished Williams on the ground that Williams concerned sentencing, whereas Mullaney concerned the defendant’s guilt.\(^ {59}\) But in a separate passage, the Court also recognized the permeability of this boundary. Arguing that substance controls over form and that a state cannot redesignate and reassign offense elements at will,\(^ {60}\) the Court wrote:

Moreover, if Winship were limited to those facts that constitute a crime as defined by state law, a State could undermine many of the interests that decision sought to protect without effecting any substantive change in its law. It would only be necessary to redefine the elements that constitute different crimes, characterizing them as factors that bear solely on the extent of punishment.\(^ {61}\)

This conflicts with Williams because that case’s central argument implies a plenary power in the states to designate offense elements as sentencing factors. If a sentencing judge has the discretion to impose a sentence for any reason whatsoever or for no reason at all, then the sentencing judge can impose a sentence implicitly on the basis of a fact that historically has been an element of an offense. To require the judge to exercise discretion on such a basis explicitly, and to require the proof of this fact to be by a preponderance, constitutes a net constitutional gain, not a loss.

McMillan v. Pennsylvania\(^ {62}\) elevated this implication of Williams to the status of constitutional doctrine under the Sixth Amendment and Due Process Clauses. Pennsylvania’s Mandatory Minimum Sentencing Act set a mandatory minimum sentence of five years for certain specified felonies provided that the prosecution proved “visible possession of a firearm.”\(^ {63}\) This fact was to be proved at sentencing, to the judge, by a preponderance. McMillan and others contended that this way of proceeding violated Winship, because the visible possession of a

\(^{57}\) Maine construed its law to be that malice aforethought and the heat of passion were two sides of a single coin. See id. at 686–87. The Court accepted Maine’s construction as binding, in accordance with long-standing constitutional doctrine. Id. at 691.

\(^{58}\) See id. at 703–04.

\(^{59}\) See id. at 697 n.23.

\(^{60}\) Id. at 699 (“Winship is concerned with substance rather than this kind of formalism.”).

\(^{61}\) Id. at 698.


\(^{63}\) Id. at 80–81 (citing 42 Pa. Cons. Stat. § 9712 (1982)).
firearm was a fact on the basis of which the State proposed to stigmatize and imprison them. But the Court rejected this argument on the rationale suggested by Williams:

Sentencing courts have traditionally heard evidence and found facts without any prescribed burden of proof at all. Pennsylvania has deemed a particular fact relevant and prescribed a particular burden of proof. We see nothing in Pennsylvania's scheme that would warrant constitutionalizing burdens of proof at sentencing.

Petitioners apparently concede that Pennsylvania's scheme would pass constitutional muster if only it did not remove the sentencing court's discretion, i.e., if the legislature had simply directed the court to consider visible possession in passing sentence. We have some difficulty fathoming why the due process calculus would change simply because the legislature has seen fit to provide sentencing courts with additional guidance.64

The McMillan dissenters stressed the point that the Court had made in Mullaney: If the state has the unrestricted power to designate facts that justify punishment as "sentencing factors" or "affirmative defenses," then it can evade Winship's due process requirement of proof beyond a reasonable doubt.65 Winship had rejected this kind of formalism: No matter what the state wished to call it, any fact that the state invoked to justify punishment had to be proved beyond a reasonable doubt. In response, the McMillan Court conceded that "there are constitutional limits to the State's power in this regard,"66 but it expressly declined to say what those limits were.67 Subsequently, lower courts in need of guidance have taken it from a passing remark in the McMillan opinion that is almost comically unsuited to the role of a constitutional standard:68 "The statute gives no impression of having been . . . a tail which wags the dog of the substantive offense."69

The way to McMillan's holding had been eased considerably by Patterson v. New York,70 a case concerning an affirmative defense and one that severely limited the holding of Mullaney. Patterson was convicted of murder, a crime that, in New York, required the State to prove that the defendant had caused the

---

64. Id. at 91–92 (citing Williams v. New York, 337 U.S. 241 (1949)) (internal citation omitted). The Court also reasoned that

[t]he Pennsylvania Legislature did not change the definition of any existing offense. It simply took one factor that has always been considered by sentencing courts to bear on punishment—the instrumentality used in committing a violent felony—and dictated the precise weight to be given that factor if the instrumentality is a firearm.

Id. at 89–90.

65. See id. at 93–94 (Marshall, J., dissenting); id. at 103 (Stevens, J., dissenting).


67. See id.


death of another person and had done so with intent. The criminal code allowed the defendant the opportunity to reduce his crime to manslaughter, provided that he could prove, by a preponderance, that he had acted under an extreme emotional disturbance.\(^{71}\) The Court recognized that extreme emotional disturbance is the modern restatement of the defense of provocation or "heat of passion."\(^{72}\) The Court perceived that "[i]t is plain enough that if the intentional killing is shown, the State intends to deal with the defendant as a murderer unless he demonstrates the mitigating circumstances."\(^{73}\) However, the Court did not recognize that this made the New York law indistinguishable from the Maine law struck down in \textit{Mullaney}. The Court thought it a sufficient distinction to say that New York had defined a set of elements constituting murder, that this set did not include "malice aforethought," and that New York had created an "affirmative defense" of extreme emotional disturbance without reference to these elements. Whereas Maine had treated heat of passion as the converse of malice aforethought, New York had never treated extreme emotional disturbance as the flip side of any element of murder.\(^{74}\) The Court decided that the real defect in \textit{Mullaney}, which was not present in \textit{Patterson}, was that Maine had shifted the burden on an acknowledged element of murder onto the defendant.\(^{75}\)

As the \textit{Patterson} dissenters pointed out, this exercise in labeling was precisely the sort of formalism that \textit{Winship} and \textit{Mullaney} had rejected in their due process analyses.\(^{76}\) But the \textit{Patterson} Court's formalism had the virtue of serving the strong, countervailing constitutional values of federalism and legislative supremacy. If a state wished to provide the benefit of a defense or mitigation in criminal cases, the Court argued, then it was surely within the state's special competence over criminal matters to condition that benefit as it chose: "The Due Process Clause, as we see it, does not put New York to the choice of abandoning those defenses or undertaking to disprove their existence in order to convict of a crime which otherwise is within its constitutional powers to sanction by substantial punishment."\(^{77}\) The Court had reasoned in this way when it recognized the states' prerogative to impose the burden of persua-

\(^{71}\) \textit{Id.} at 200.

\(^{72}\) \textit{Id.} at 202; \textit{see also} People v. Casassa, 404 N.E.2d 1310, 1314 (N.Y. 1980) (describing the state's "extreme emotional disturbance" defense as based on the Model Penal Code's revision of the heat of passion defense); \textit{Model Penal Code} § 210.3 cmt. 5(a), at 53–65 (1980) (describing "extreme emotional disturbance" as a version of provocation).

\(^{73}\) \textit{Patterson}, 432 U.S. at 206.

\(^{74}\) \textit{Id.} at 215–16. However, as Justice Powell pointed out in his \textit{Patterson} dissent, \textit{id.} at 222 n.4 (Powell, J., dissenting), and as the \textit{Mullaney} Court explicitly recognized, 421 U.S. at 689, the Maine court had decided that malice aforethought was not an element of murder, but was only a policy presumption underlying its homicide law. Thus, the elements of murder in Maine were essentially the same as the elements of murder in New York. In this light, in particular, \textit{Patterson} presented exactly the same issue as \textit{Mullaney}. \textit{See infra} Part II.B.

\(^{75}\) \textit{See Patterson}, 432 U.S. at 214–15.

\(^{76}\) \textit{See id.} at 224–25 (Powell, J., dissenting).

\(^{77}\) \textit{Patterson}, 432 U.S at 207–08.
sion of insanity on the defense. Once it had cast New York's recognition of extreme emotional disturbance in this mold, the Patterson Court saw no reason to revisit that decision.

The Williams-McMillan line of cases reached its high-water mark in Almendarez-Torres v. United States. Almendarez-Torres was charged with and pled guilty to the crime of being found in the United States after having previously been deported. At his sentencing, Almendarez-Torres admitted that he had been deported because of three earlier convictions for aggravated felonies. This brought his case under a different subsection of the offense statute than that stated in the indictment and subjected him to a twenty-year instead of a two-year maximum sentence. Almendarez-Torres was sentenced to eighty-five months. In his challenge to this sentence, he argued that the statute created two distinct crimes and that he had not been charged with the one that carried the higher maximum penalty. The Court rejected this interpretation of the statute and affirmed the sentence.

Invoking Congress's broad power to define crimes, defenses, and sentencing factors, the Court specifically rejected an argument based on the doctrine of constitutional doubt. Almendarez-Torres had argued that to read the statute to permit an increase in the maximum sentence on the basis of a "sentencing element" that had not been proved to a jury beyond a reasonable doubt would cast the statute in a questionable constitutional light. This argument was substantial. McMillan had turned in part on the proposition that the sentence enhancement in that case did not increase the maximum sentence. Specht had turned in part on the proposition that the sentence enhancement in that case did alter the maximum. Nevertheless, by means of stingy readings of Winship, Specht, and Mullaney, the Court found no constitutional doubt in Almendarez-Torres. It arrived at this conclusion in spite of a grudging recognition that its

78. Leland v. Oregon, 343 U.S. 790, 798 (1952), cited in Patterson, 432 U.S. at 204.
79. Patterson, 432 U.S. at 207.
81. Id. at 227.
82. Id.
83. Id. at 228.
84. Id. This doctrine of statutory interpretation consists of the rule that "[a] statute must be construed, if fairly possible, so as to avoid not only the conclusion that it is unconstitutional but also grave doubts upon that score." United States v. Jin Fuey Moy, 241 U.S. 394, 401 (1916) (Holmes, J.), cited in Almendarez-Torres, 523 U.S. at 237.
88. Regarding Winship, the Court in Almendarez-Torres reasoned that the case did not say what should count as an element, even if it did require all elements to be proved beyond a reasonable doubt. Almendarez-Torres, 523 U.S. at 240. But Winship clearly prescribed a substantive inquiry into this question in its rejection of the State's argument that the facts proved there were not elements of a crime, but only elements of a juvenile offense. See supra notes 48–55 and accompanying text. Regarding Specht, the Court relied on the distinction without a difference drawn in that case. See Almendarez-Torres, 523 U.S. at 241–42; supra notes 46–47 and accompanying text. Regarding Mullaney, the Court
cases were in significant tension. Concerning *Mullaney* and *Patterson*, the Court concluded, “The upshot is that *Mullaney*’s language, if read literally, suggests that the Constitution requires that most, if not all, sentencing factors be treated as elements. But *Patterson* suggests the exact opposite, namely that the Constitution requires scarcely any sentencing factors to be treated in that way.”  

Elsewhere, the Court wrote in a similar vein: “At most, petitioner might read all these cases, taken together, for the broad proposition that *sometimes* the Constitution does require (though sometimes it does not require) the State to treat a sentencing factor as an element.” The Court did not credit these interpretations and, therefore, found no constitutional doubt. The dissenters found constitutional doubt in more frank readings of *Winship*, *Specht*, and *Mullaney* and in a full appreciation of the resulting tension between these cases and the *Williams-McMillan* line.

One year later, in another case calling for statutory interpretation, the tide turned. *Jones v. United States* required the Court to interpret a federal criminal statute that was distinguishable, but not significantly different from the statute at issue in *Almendarez-Torres*. In both cases, good arguments could be made for and against the view that the statute created several offenses instead of one offense with several sentencing provisions. The difference was that the Court found constitutional doubt in *Jones*, whereas it had harbored no such doubts in *Almendarez-Torres*. Behind this difference lay a different view of the historic norms of the criminal law. *Almendarez-Torres* had been decided in line with *McMillan*, in which the Court had written: “We reject the view that anything in the Due Process Clause bars States from making changes in their criminal laws that have the effect of making it easier for the prosecution to obtain convictions.” But in *Jones*, the Court found constitutional doubt precisely because the Government’s reading of the statute would make it easier for the prosecution to obtain convictions.

The constitutional question in *Jones*, according to the Court, was “whether recognizing an unlimited legislative power to authorize determinations setting ultimate sentencing limits without a jury would invite erosion of the jury’s function to a point against which a line must necessarily be drawn.” The Court

---

90. *Id.* at 242.  
91. *Id.* at 256 (Scalia, J., dissenting).  
94. See *McMillan v. Pennsylvania*, 477 U.S. 79, 90 n.5 (1986). Presumably, the Court took the same view of the Sixth Amendment, given that the convictions at issue in *McMillan* had been challenged on this ground as well.  
95. *Jones*, 526 U.S. at 244.
found that there was such a danger, and it drew such a line. In doing so, the Court stressed the fact that, in the Founding era, the Crown had attempted to ensure convictions by drafting criminal statutes so as to avoid jury decisionmaking. This was a prime motivation behind the Jury Right Clause. The issue was prominent in the Founders' political value system because of the influence of Blackstone, who had condemned such measures, and because of the trial of John Peter Zenger, which had offered the colonists an object lesson in the power of juries to counter government overreaching. The Jones dissenters, of course, perceived no such danger. On the contrary, they perceived a threat to the Federal Sentencing Guidelines and state mandatory-minimum sentencing schemes and, more broadly, to the constitutional power of Congress and the states to define crimes and prescribe punishments.

The Jones dissenters demanded that the Court state the constitutional principle that raised doubts about the Government's interpretation of the statute. In order to oblige, the Court had to navigate around the holdings of McMillan and Almendarez-Torres, but it did so with relative ease. The Court wrote,

[U]nder the Due Process Clause of the Fifth Amendment and the notice and jury trial guarantees of the Sixth Amendment, any fact (other than prior conviction) that increases the maximum penalty for a crime must be charged in an indictment, submitted to a jury, and proven beyond a reasonable doubt.

The Court had to concede that no prior case rested squarely on this proposition, but it found constitutional doubt nevertheless. In Apprendi, with one minor variation, this proposition was elevated to an express requirement of the Constitution.

Apprendi had been accused of shooting at the front door of an African-American family in his neighborhood, and charged with twenty-three counts of various weapons and assault crimes. None of the twenty-three counts alleged that he had acted with a racially discriminatory purpose. Apprendi pled guilty to three counts. The maximum statutory term available for the most serious of these offenses, second-degree possession of a firearm for unlawful purposes, was five to ten years. Apprendi was sentenced to twelve years in prison. This sentence was based on the sentencing court's finding, by a preponderance of the evidence in a closely contested hearing, that Apprendi had acted "with a

96. Id. at 247-48.
97. Id. at 246 (quoting 4 WILLIAM BLACKSTONE, COMMENTARIES *277-79).
98. Id. at 246-47.
99. Id. at 271 (Kennedy, J., dissenting).
100. Id. at 266-69.
102. Apprendi, a state case, did not implicate the Grand Jury Clause, and due process limits on charging in state cases were not litigated. See Apprendi v. New Jersey, 530 U.S. 466, 477 n.3 (2000). As a result, Apprendi did not create a constitutional requirement for notice in charging.
purpose to intimidate an individual or group of individuals because of race, color, gender, handicap, religion, sexual orientation or ethnicity." The court made this finding under New Jersey's "hate crime" statute, which increased the maximum penalty for crimes committed with such a purpose to twenty years. The Supreme Court reversed Apprendi's conviction and sentence on the ground suggested in Jones: "Other than the fact of a prior conviction, any fact that increases the penalty for a crime beyond the prescribed statutory maximum must be submitted to a jury, and proved beyond a reasonable doubt." The Apprendi Court, like the Jones Court, was motivated primarily by a desire to preserve the traditional normative architecture of the criminal law. Quoting Joseph Story and William Blackstone, the Court wrote,

"[T]o guard against a spirit of oppression and tyranny on the part of rulers," and "as the great bulwark of [our] civil and political liberties," trial by jury has been understood to require that "the truth of every accusation, whether preferred in the shape of indictment, information, or appeal, should afterwards be confirmed by the unanimous suffrage of twelve of [the defendant's] equals and neighbours . . . ."

Similarly, the Court noted, as it had in Winship, that the reasonable doubt standard of proof had been a part of the law in England and the United States at least since 1798. In his concurrence, Justice Scalia noted that "the jury-trial guarantee was one of the least controversial provisions of the Bill of Rights." 107

103. See id. at 469 (quoting N.J. STAT. ANN. § 2C:44-3(e) (West Supp. 2000)) (internal quotation marks omitted).
104. Id. at 490.
105. Id. at 477 (alterations in original) (emphasis omitted) (quoting 2 JOSEPH STORY, COMMENTARIES ON THE CONSTITUTION OF THE UNITED STATES 540–41 (4th ed. 1873); and 4 BLACKSTONE, supra note 97, at *343).
106. Id. at 478.
107. Id. at 498 (Scalia, J., concurring). Justice Scalia's regard for the Sixth Amendment jury guarantee as a cornerstone of the original constitutional structure is clear in his dissent in Neder v. United States, wherein he wrote:

Article III, § 2, cl. 3 of the Constitution provides: "The Trial of all Crimes, except in Cases of Impeachment, shall be by Jury . . . ." The Sixth Amendment provides: "In all criminal prosecutions, the accused shall enjoy the right to a speedy and public trial, by an impartial jury . . . ." When this Court deals with the content of this guarantee—the only one to appear in both the body of the Constitution and the Bill of Rights—it is operating upon the spinal column of American democracy. William Blackstone, the Framers' accepted authority on English law and the English Constitution, described the right to trial by jury in criminal prosecutions as "the grand bulwark of [the Englishman's] liberties . . . secured to him by the great charter." 4 W. Blackstone, Commentaries *349. One of the indictments of the Declaration of Independence against King George III was that he had "subject[ed] us to a Jurisdiction foreign to our Constitution, and unacknowledged by our Laws" in approving legislation "[f]or depriving us, in many Cases, of the Benefits of Trial by Jury." Alexander Hamilton wrote that "[t]he friends and adversaries of the plan of the convention, if they agree in nothing else, concur at least in the value they set upon the trial by jury; Or if there is any difference between them, it consists in this, the former regard it as a valuable safeguard to liberty, the latter
The Court made it clear that it regarded the notion of a sentencing factor to be of a much more recent and distinctly inferior vintage. 108

Commentators and courts had read *McMillan* as a grant of power to the states, and by implication to Congress, to define crimes and punishments as they saw fit. 109 *Apprendi* turned this interpretation on its head by stressing two aspects of *McMillan* that had passed largely unnoticed in the intervening years. First, the *McMillan* opinion did acknowledge that due process imposes some limits on legislative power to create sentencing factors, even if it left those limits vague. 110 Second, the sentence enhancement law at issue in *McMillan* did not authorize a sentence in excess of the statutory maximum for the crime; it merely limited the sentencing judge's discretion within that range. 111 By emphasizing these two features of *McMillan*, the Court actually could invoke *McMillan* in support of its holding in *Apprendi*—a point of bitter irony to the dissenters. 112 Statutory maxima would serve as the due process boundary on the use of sentencing factors. In response to the dissenters' accusation that the Court had overruled *McMillan*, Justice Stevens pointed out that the holding in *Apprendi* merely narrowed the rule of *McMillan*, confining it to the authorization of the use of sentencing factors to support enhancements within a statutory sentencing range. 113

*Almendarez-Torres* had been regarded by many members of the Court as a logical and uncontroversial extension of *McMillan* beyond this limit. 114 In that case, the sentencing factor of Almendarez-Torres's prior convictions for aggravated felonies had been used to justify a sentence of eighty-five months

---


108. See *Apprendi*, 530 U.S. at 478 ("Any possible distinction between an 'element' of a felony offense and a 'sentencing factor' was unknown to the practice of criminal indictment, trial by jury, and judgment by court as it existed in the years surrounding our Nation's founding."); id. at 483 n.10 ("Rather than offer any historical account of its own that would support the notion of a 'sentencing factor' legally increasing punishment beyond the statutory maximum—and Justice Thomas' concurring opinion in this case makes clear that such an exercise would be futile—the dissent proceeds by mischaracterizing our account."); id. at 485 ("It was in *McMillan v. Pennsylvania* that this Court, for the first time, coined the term 'sentencing factor' to refer to a fact that was not found by a jury but that could affect the sentence imposed by the judge."); id. at 494 ("That point applies as well to the constitutionally novel and elusive distinction between 'elements' and 'sentencing factors.'")).

109. See, e.g., Bilionis, supra note 31, at 1305.


111. Id. at 88–89.


114. See *Jones v. United States*, 526 U.S. 227, 248–49 (1999); id. at 268–69 (Kennedy, J., dissenting); see also *Apprendi*, 530 U.S. at 535 (O'Connor, J., dissenting) ("[In *Almendarez-Torres*] we squarely rejected the 'increase in the maximum penalty' rule.'").
following his plea of guilty to a crime that carried a maximum sentence of two years. The Apprendi Court, however, chose to treat Almendarez-Torres as “at best an exceptional departure from the historic practice that we have described.” The Court regarded criminal history as a genuine sentencing factor, given that “recidivism . . . is a traditional, if not the most traditional, basis for a sentencing court’s increasing an offender’s sentence,” and given that criminal history is not part of the commission of the offense. The Court also stressed that, unlike other facts that may serve as sentencing factors, criminal history is itself the product of judicial proceedings that presumably met due process requirements and is thus supported by “substantial procedural safeguards of [its] own.” In this way the Apprendi Court put in place the second of its two limitations on the broad principle that the facts that are used to justify punishment must be proved beyond a reasonable doubt to a jury.

The dissenters doubted, with reason, that the Court intended to observe either of these two limitations for long. Even while Justice Stevens insisted that Apprendi did not overrule McMillan, he refused to rule out that possibility. He wrote, “Conscious of the likelihood that legislative decision may have been made in reliance on McMillan, we reserve for another day the question of whether stare decisis considerations preclude reconsideration of its narrower holding.” Justice Stevens signaled that Almendarez-Torres might be confined to its facts, when he pointedly noted that Almendarez-Torres had not contested the fact of his prior convictions, and he cited this as one of the reasons that no constitutional doubt had been raised in that case. Justice Stevens also refused to rule out the possibility that Almendarez-Torres might be overruled outright, in stronger terms than he had used in connection with McMillan. He noted that “it is arguable that Almendarez-Torres was incorrectly decided, and that a logical application of our reasoning today should apply if the recidivist issue were contested.” In short, the dissenters’ anxiety over the fate of the Williams-McMillan line of cases was well-founded.

B. CONCEPTUAL TENSION

If Apprendi represents an unstable balance of power between two credible lines of authority, then it is worth considering whether the entire problem has not been misconceived. A close examination of the Apprendi opinions indicates the depth of the stalemate in which the Court has been locked, because it reveals

115. Apprendi, 530 U.S. at 487.
116. Id. at 488 (quoting Almendarez-Torres, 523 U.S. at 243) (internal quotation marks omitted).
117. Id. (quoting Almendarez-Torres, 523 U.S. at 244).
118. Id.
119. Id. at 487 n.13.
120. Id. at 488; see also id. at 490 (“Given its unique facts, [Almendarez-Torres] surely does not warrant rejection of the otherwise uniform course of decision during the entire history of our jurisprudence.”).
121. Id. at 489–90.
that the antagonists lack the conceptual resources they need to escape it. The source of this deficiency is the subject of the next Part. For now, I wish only to show how this deficiency is manifested in the *Apprendi* opinions.

1. Dueling Reductios ad Absurdum

Foremost among the signs of fundamental confusion is the fact that the basic position of each side in this controversy can be reduced to absurdity in a few quick steps. Concurring in *Apprendi*, and writing for the *Winship-Mullaney* faction on the Court, Justice Scalia suggested the reductio ad absurdum of the *Williams-McMillan* line when he wrote,

> What ultimately demolishes the case for the dissenters is that they are unable to say what the right to trial by jury does guarantee if, as they assert, it does not guarantee... the right to have a jury determine those facts that determine the maximum sentence the law allows. They provide no coherent alternative. 122

The *Williams-McMillan* argument obliterates not only the right to a jury, but also due process and other trial right guarantees. If a state legislature or Congress has the authority to make any fact that formerly has served as an element of a crime into a sentencing factor instead, then there is no logical reason that it cannot make all offense elements into sentencing factors. For example, a legislature may make any and all persons found in the jurisdiction guilty of murder and then provide for sentencing based on whether an individual can demonstrate by a preponderance to the sentencing court that he did not purposely cause the death of another person. This point was recognized as far back as *Mullaney*, of course, when the Court noted that "if *Winship* were limited to those facts that constitute a crime as defined by state law, a State could undermine many of the interests that decision sought to protect without effecting any substantive change in its law." 123

The essential point is not that there must be some limit on legislative power in this regard. The *McMillan* Court acknowledged this much, even though it refused to specify the limit. 124 The essential point is that the *McMillan* Court refused to specify the limit because it could not do so. The Court lacked the conceptual means to limit *McMillan* because the seminal argument of *Williams* does not imply any such limits. The *Williams-McMillan* argument is that a judge who has sentencing discretion may impose a sentence for any reason or for none at all, and that it is therefore a constitutional gain to require the judge to make her findings and conclusions explicit—even if the rights that adhere at trial are not provided for in the sentencing procedure. If it is true that there are no limits

---

122. *Id.* at 498–99 (Scalia, J., concurring).
on a sentencing judge’s discretion, then a sentencing judge can pick any courtroom observer and sentence that person to life imprisonment. One objects, of course, that even in a discretionary sentencing system the sentencing judge must have a conviction of a crime on the record before she can sentence anyone, and even then she may not exceed the statutory maximum allowed for such a conviction. But it is important to notice that these are limits that the Williams argument supposes, not limits that it implies. If we take the unreviewable discretion of a sentencing judge as the standard by which we measure what the legislature may do, then the legislature may evade any constitutional limitation on punishment at will. The point is not that we are in danger of such oppression, but only that the Williams argument logically does not rule it out.

The Apprendi opinion is vulnerable to a mirror-image reductio ad absurdum. Justice Stevens’s opinion does not remotely suggest a return to fixed felony sentencing, in which the judge exercises no discretion in the determination of a sentence and only imposes a sentence prescribed by law for each given offense. On the contrary, the Apprendi opinion supposes that discretionary judicial sentencing is constitutionally valid. But if the sentencing judge exercises any discretion at all, then inevitably the judge will determine facts in the course of exercising that discretion. It was incumbent upon the Apprendi Court, then, to draw some boundaries between the jury’s role in determining facts and the judge’s role in determining facts. In the Apprendi opinion, the statutory maximum serves this purpose: The jury determines the facts that establish which statutory maximum applies by virtue of the offense of conviction; the judge determines the facts that set a sentence within this statutory maximum.

The difficulty is that the statutory maximum is entirely within the control of the legislature, which means that the legislature retains all the power it needs to evade Winship or to depart from the traditional norms of criminal law in any other respect. The legislature needs only to increase the statutory maxima throughout its criminal code, and then to use sentencing factors to mitigate punishments downward. In her Apprendi dissent, Justice O’Connor pointed this out and used New Jersey’s hate crime law as an example. New Jersey had defined a weapons offense and, in the offense statute itself, had provided for a statutory maximum of ten years. In a separate statute, the state had provided for a sentence of up to twenty years for this same offense, among others, if the defendant acted with a discriminatory purpose. Apprendi struck this system down. But now the state could easily amend its statutory scheme to provide for the twenty-year maximum in the offense statute itself. In order to obtain a sentence at the lower end of this range, the defendant then could be required to prove—to the judge, by a preponderance, and without trial rights—that he had


126. See Apprendi, 530 U.S. at 540–41 (O’Connor, J., dissenting).
not acted with a racially discriminatory purpose. This would comply with the rule of Apprendi and would reach the same result that the state originally sought to attain. However, as Justice O’Connor pointed out, “[i]t is difficult to understand, and the Court does not explain, why the Constitution would require a state legislature to follow such a meaningless and formalistic difference in drafting its criminal statutes.”

Justice Stevens offered the same reply to this argument that he had offered fifteen years before in his McMillan dissent: There are political constraints that will forestall the legislature’s increasing statutory maxima to draconian levels across the board. As a matter of fact, this assumption is doubtful. In recent years, the public has been not only willing, but also positively eager to impose draconian sanctions on criminals. But it is less important to see how this reply is contingent on fact, than to see that it is contingent on fact. The problem with Justice Stevens’s reply lies not with the people, desperate and irrational as their punitive impulses may be, but with the logic of using statutory maxima to define a criminal offense for constitutional purposes. It is impossible to restrain positive law by means of some other feature of positive law. Constitutional adjudication requires a firmer foundation in principle and theory.

2. Weakness All Around

If these mirror-image reductios of the factions’ principal arguments indicate fundamental confusion or conceptual inadequacies, then one consequence of these deficiencies is the weakness of the more detailed arguments that the factions advanced in the Apprendi opinions.

The Apprendi Court preserved the holding of Almendarez-Torres, taking care to distinguish recidivism as a sentencing factor from the offense element of a discriminatory purpose. Recidivism is a sentencing factor, the Apprendi opinion said, not only because it is “a traditional, if not the most traditional, basis for a sentencing court’s increasing an offender’s sentence,” but also because it is not a part of the commission of the offense and it carries its own due process imprimatur from the proceedings that produced it. The Apprendi dissenters doubted that either the holding of Almendarez-Torres or the limitation on the Apprendi rule that reflects that holding would survive for long. Their

127. Id. at 541.
128. Id. at 541 n.16; see also McMillan, 477 U.S. at 100–02 (Stevens, J., dissenting) (arguing that political constraints exist that prevent legislatures from evading Winship’s rule at will).
130. See Apprendi, 530 U.S. at 486–90.
131. Id. at 488 (quoting Almendarez-Torres v. United States, 523 U.S. 224, 244 (1998)).
132. See id.
133. See id. at 541 (O’Connor, J., dissenting) (“Given the pure formalism of the above readings of the Court’s opinion, one suspects that the constitutional principle underlying its decision is more far reaching.”).
fear was justified because none of these reasons for excepting criminal history from the *Apprendi* rule withstands scrutiny.

The notion that criminal history is sufficiently reliable for due process purposes, so that trial-like standards of proof can be omitted, turns entirely on the assumption that the person to be sentenced is the same person who collected the criminal history in question. But this fact itself needs to be proved, and this question of identity is sufficiently uncertain often enough to merit the protections of due process and other trial rights—not to mention that identity is as typical an offense element as one may imagine.\(^{135}\)

A more fundamental objection to the *Apprendi* Court’s position on recidivism is that it takes a categorical approach to the problem of offense elements versus sentencing factors. Even if criminal history is a “traditional” sentencing factor (though the Court clearly views this notion itself as an oxymoron), it also can be an element of an offense. Modern analysis of offense elements recognizes that

\[^{136}\] the last category includes circumstances that are also features of the offender, such as his age in the offense of statutory rape or his identity in all criminal cases. Criminal history can work like this: It is an attendant circumstance element in a prosecution for an offense premised in whole or in part on recidivism.\(^{137}\) In such a case, criminal history is not traditionally part of the sentencing decision and is a part of the commission of the offense.

The sterility of a categorical approach affected a number of arguments made in the *Apprendi* opinions. The New Jersey hate crime statute imposed an enhanced sentence if the defendant “acted with a purpose to intimidate an individual or group of individuals because of race, color, gender, handicap, religion, sexual orientation, or ethnicity.”\(^ {138}\) According to the Court, the word “purpose” denoted a mens rea.\(^ {139}\) Such a culpability element, the Court concluded, was manifestly one of the elements of the offense. As a constitutional matter, the prosecution could not be relieved of its obligation to prove this element beyond a reasonable doubt to a jury. However, to the dissenters, such a purpose was nothing more than a motive, which Justice O’Connor characterized as a “traditional” basis for an increased sentence\(^ {140}\) and as “one factor that has always been considered by sentencing courts to bear on punishment.”\(^ {141}\) The Court pointed to the use of the word “purpose” in the New Jersey criminal code

\[^{134}\] See *Almendarez-Torres*, 523 U.S. at 268–69 (Scalia, J., dissenting) (discussing evidence from the legislative history of alien defendants with multiple aliases).

\[^{135}\] Cf. *Almendarez-Torres*, 523 U.S. at 230 (stating that recidivism “is as typical a sentencing factor as one might imagine”).


\[^{140}\] *Id.* at 554 (O’Connor, J., dissenting).

\[^{141}\] *Id.* at 553.
to describe an important kind of culpability. But there was no reason to think that the New Jersey Legislature had this usage in mind when it passed the law, particularly because the word had not been used in the definition of an offense. The dissenters pointed to a previous hate crime case, Wisconsin v. Mitchell, in which the Court had referred to such a discriminatory purpose as a motive. But nothing in Mitchell particularly turned on this characterization, and in any event, the purpose in the Wisconsin statute had to be proved to a jury. Thus, neither side in Apprendi could adduce any compelling reason why a discriminatory purpose should be categorized as either an element or a sentencing factor.

Justice Breyer's dissent makes another argument that reveals the incoherence of a categorical approach to the sentencing-factor-versus-offense-element question. He argues that some facts must be sentencing factors simply as a practical matter. To make some facts elements, he argues, "could easily place the defendant in the awkward (and conceivably unfair) position of having to deny he committed the crime yet offer proof about how he committed it, e.g., 'I did not sell drugs, but I sold no more than 500 grams.'" This argument is a variant on one that the Court made in Almendarez-Torres. There, the claim was that recidivism is better treated as a sentencing factor because to prove the defendant's criminal history before the jury would prejudice him in the eyes of the jurors. The standard answer to both of these arguments invokes the possibility of bifurcated proceedings, in which the jury would consider such potentially prejudicial facts in a separate proceeding, following its consideration of the main charge.

142. Apprendi, 530 U.S. at 493 n.17 (citing N.J. STAT. ANN. § 2C:2-2(b)(1) (West 1999)); see also MODEL PENAL CODE § 2.02(2)(a) (1985) (setting forth the definition of "purpose" adopted by New Jersey and listing it among the "Kinds of Culpability").

143. Apprendi, 530 U.S. at 553 (O'Connor, J., dissenting).

144. Apprendi, 530 U.S. at 493 n.18.

145. These fruitless efforts to categorize recapitulated a similar, and similarly hollow, exercise in Jones v. United States, 526 U.S. 227 (1999): The issue was joined in that case over the term "serious bodily injury." The Court acknowledged that Congress had treated serious bodily injury as a sentencing factor in some statutes, but it noted that serious bodily injury historically has been an element of aggravated robberies. The Court considered this classification to be decisive in a case involving car-jacking. Id. at 234–35. Serious bodily injury was therefore an offense element. The dissent, however, was equally certain that "the harm from a crime—including whether the crime, after its commission, results in the serious bodily injury or death of a victim—has long been deemed relevant for sentencing purposes. Like recidivism, it is 'as typical a sentencing factor as one might imagine.'" Id. at 256–57 (Kennedy, J., dissenting) (quoting Almendarez-Torres v. United States, 523 U.S. 224, 230 (1998)). Both factions could and did cite extensive authority in support of their positions. It seems not to have dawned on anyone that serious bodily injury might be both an offense element and a sentencing factor, and that a categorical approach to the question was bound to be a failure all around.

146. See Apprendi, 530 U.S. at 556–57 (Breyer, J., dissenting).

147. Id. at 557.

148. See Almendarez-Torres, 523 U.S. at 234–35.

149. See, e.g., Bibas, supra note 22, at 1143.
not be characterized as anything other than a penalty phase—but also that such bifurcated proceedings never have been constitutionally required.\(^{150}\) As Justice Breyer, citing Williams, points out, we have opted instead for the more efficient alternative of discretionary judicial sentencing.\(^{151}\)

The real flaw in Justice Breyer’s argument is that the dilemma he describes does not distinguish sentencing factors from offense elements. Many offenses contain compound allegations that a defendant may wish to deny in part and concede in part. For example, in an ordinary burglary prosecution, the allegation is that the defendant entered upon property with the intent to commit some felony.\(^{152}\) Many of those accused of burglary are perfectly willing to concede that they entered the property, but wish to deny that they had any felonious intent. The position of such an alleged burglar is not logically different from that of the alleged drug dealer who wishes to concede that he delivered drugs but wishes to deny that he delivered a lot of drugs. The difference is that we have not drafted our drug offense statutes in a way that makes this distinction. The burglar’s dilemma is resolved by means of the lesser included offenses doctrine.\(^{153}\) The alleged burglar can concede the unlawful entry without conceding that he committed burglary, because he can ask for an instruction on criminal trespass.\(^{154}\) Whether or not an alleged drug dealer can ask for a similar instruction depends on whether the drug offense statutes are drafted in a similar way. There is nothing in the nature of drugs or drug dealing that precludes drafting the drug offenses in this way, or that necessarily assigns drug quantities to the category of sentencing factors because of the unavailability of a lesser included offense instruction.

The prejudicial effect of criminal history or recidivism also has been overstated. If criminal history is an element of an offense, then in a prosecution for that offense the jury will hear the defendant’s prior crimes alleged and will be presented with evidence that he committed these crimes.\(^{155}\) Justice Breyer would argue that the defendant will be prejudiced by this. But how would this differ from a prosecution for murder, in which the jury hears the allegation that the defendant killed another human being and is presented with evidence that this is so? Of course the defendant is prejudiced by this allegation, but not unfairly so. There is no other way to conduct a murder trial. We do what we can to minimize unfair prejudice by means of an instruction telling the jury that allegations are no more than allegations and that nothing is to be inferred from

---

151. See Apprendli, 530 U.S. at 557–58 (Breyer, J., dissenting).
155. See, e.g., State v. Gill, 997 P.2d 710, 714 (Kan. 2000) (ruling that in prosecution for felon in possession of handgun, right to stipulate to prior felony is waived if no stipulation is offered).
the accusation itself.\textsuperscript{156} If criminal history were an element of some offense, then the same instruction would be given with respect to criminal history in any trial on that offense.\textsuperscript{157} There is nothing in an allegation of criminal history that necessarily distinguishes it, as an allegation, from an allegation of murder, or that precludes it from being an offense element instead of a sentencing factor.

Elsewhere in his dissent, Justice Breyer stressed the \textit{Williams} argument and offered an interesting variation on it. He acknowledged the danger that a legislature may transform elements of crimes into sentencing factors, thereby evading constitutional limits on punishment. However, Justice Breyer argued, the problem is endemic: Any judge possessed of an ordinary amount of discretion over sentencing has a similar power. A judge may, for example, impose consecutive maximum sentences on a multiple count conviction for embezzlement, based on her own finding that the defendant also had murdered his employer.\textsuperscript{158} From this “egregious example,”\textsuperscript{159} Justice Breyer concluded that the solution to the problem the Court faced did not lie in restricting legislative power to define elements and select sentencing factors, but rather “in sentencing rules that determine punishments on the basis of properly defined relevant conduct, with sensitivity to the need for procedural protections where sentencing factors are determined by a judge.”\textsuperscript{160} In other words, the solution lies in a system such as the Federal Sentencing Guidelines.\textsuperscript{161}

The difficulty is that the Guidelines themselves encourage and ratify egregious violations of fundamental fairness in sentencing. Determine sentencing schemes such as the Guidelines are meant to eliminate disparity in sentences by eliminating discretion in sentencing. But if a sentence follows rigidly, according to formula, from the offense of conviction, then discretion is not eliminated; it simply is shifted from the judge to the prosecutor, who can engage in “sentence bargaining.”\textsuperscript{162} For this

\begin{itemize}
  \item \textsuperscript{156} See \textit{Pattern Criminal Jury Instructions for the Sixth Circuit} § 1.03(1) (1991) (“The indictment is not any evidence at all of guilt.”); \textit{see also} United States v. Schanerman, 150 F.2d 941, 946 (3d Cir. 1945) (failure to give such an instruction is reversible error).
  \item \textsuperscript{157} \textit{See}, e.g., State v. Green, 493 So. 2d 588, 592 (La. 1986) (failure to give such an instruction in a recidivist theft prosecution is reversible error). Of course, rules of evidence barring prior bad acts and prior convictions would not apply to such an offense. \textit{See \textit{Fed. R. Evid.}} 404(b), 609. These rules are simply particular applications of the fundamental probative-versus-prejudicial calculus. When the prior offense is an element of a crime, the prior offense is manifestly probative, and the problem of undue prejudice must be addressed by an instruction to the effect that the allegations of the information or indictment are not evidence.
  \item \textsuperscript{158} \textit{See \textit{Apprendi v. New Jersey}, 530 U.S. 466, 562 (2000) (Breyer, J., dissenting).}
  \item \textsuperscript{159} \textit{Id.}
  \item \textsuperscript{160} \textit{Id.} at 562–63.
  \item \textsuperscript{161} \textit{See Stephen Breyer, The Federal Sentencing Guidelines and the Key Compromises upon Which They Rest, 17 Hofstra L. Rev. 1, 8–14 (1988) (describing the Guidelines in similar terms). This conclusion is not a surprising one, coming as it does from the principal architect of the Federal Sentencing Guidelines. \textit{See \textit{Stith \& Cabrantes, supra} note 24, at 58 (describing how the “Breyer draft” was adopted by the original Sentencing Guidelines Commission and ultimately by Congress).}
  \item \textsuperscript{162} \textit{See \textit{Stith \& Cabrantes, supra} note 24, at 132.}
\end{itemize}
reason, the Guidelines operate on the principle of real-offense sentencing, under which “relevant conduct” surrounding the offense of conviction—including uncharged crimes and “crimes” of which the defendant has been acquitted—is proved at a mini-trial for purposes of sentencing.

The effects of real-offense sentencing include precisely the kind of sentencing that Justice Breyer called “egregious” in his dissent:

[T]he Guidelines provide an incentive for the government to go to trial on only one count—the one most easily proved—and then effectively to “convict” the defendant on further “counts” through application of the “relevant conduct” principle. . . . One brief recently filed by the government in the federal appeals court on which one of the authors serves noted, with candor, that the government had elected to prosecute the defendant only for robbery and to wait until the sentencing hearing to prove that the defendant should also be held responsible for an associated murder.

Justice Breyer’s solution to the problem of legislative evasions of Winship is thus no solution at all. As we will see, the reason that Justice Breyer’s argument is vulnerable in this way is the fact that, from a theoretical perspective, the Federal Sentencing Guidelines are indistinguishable from the discretionary sentencing system they replaced.

Justice Thomas’s concurrence explicitly endorses the overruling of Almendarez-Torres and McMillan, and implicitly acknowledges that this may entail declaring unconstitutional the Federal Sentencing Guidelines. He advocates the broad rule that “a ‘crime’ includes every fact that is by law a basis for imposing or increasing punishment (in contrast with a fact that mitigates punishment).” Justice Thomas discusses an impressive body of authority dating from the early nineteenth century in support of his argument. In her dissent, Justice O’Connor criticized this authority as irrelevant to the meaning of the Fifth and Sixth Amendments, given that the Amendments substantially predate Justice Thomas’s cases. However, the problem with Justice Thomas’s authority is not that it is too young, but that it is too old. His nineteenth-century authority, including that collected in several authoritative treatises by Joel Prentiss Bishop, is from the era of jury sentencing and fixed felony sentenc-

---

163. But see Yellen, supra note 29, at 438–53 (arguing that the reasons for the adoption of real-offense sentencing are not clear and that the system in fact has little effect on prosecutorial power or the abuse of it).
164. STITH & CABRANES, supra note 24, at 140.
165. See infra Part II.A.1.
167. Id. at 523 n.11.
168. Id. at 501.
169. Id. at 501–18.
170. Id. at 527–28 (O’Connor, J., dissenting).
With the advent of discretionary judicial sentencing in the late nineteenth century, the rules changed. Cases having to do with jury sentencing are irrelevant to current sentencing practices. For example, Justice Breyer argues that under a discretionary judicial sentencing system a judge may sentence a person for an uncharged murder. This claim could be rebutted by a quotation from Bishop’s treatise—were it not for the last clause of the relevant passage, and but for the authority that Bishop cites.

This discrepancy in his authority means that Justice Thomas’s argument proves too much. In effect, he argues not only against the idea of sentencing factors, but also against discretionary judicial sentencing of any kind. This is just as the Williams-McMillan faction would have it: There can be no objection to a legislature’s picking out sentencing factors if discretionary judicial sentencing is constitutionally valid at all. Furthermore, and on a more fundamental level, Justice Thomas’s contention that aggravating facts must be proved to a jury while mitigating facts can be proved to the sentencing court is vulnerable to the same reductio ad absurdum that the Apprendi dissenters leveled against Justice Stevens’s opinion for the Court. Facts that aggravate punishment easily can be rephrased as facts that mitigate punishment. Which way the fact will be phrased in a criminal code is entirely up to the legislature. As a result, it is reasonable to ask Justice Thomas, as well as the Court, “why the Constitution would require a state legislature to follow such a meaningless and formalistic difference in drafting its criminal statutes.”

II. RECONCEIVING THE PUZZLE

The Court’s inability to draw a straight line from Williams through Apprendi is a consequence of its inattention to the substantive criminal law and the theory of punishment. The Court has used concepts—culpability, offense element, mens rea, mitigation—that are poorly defined or ambiguous. These concepts

---


172. Bishop writes that “the aggravating matter must not be of a crime separate from the one charged in the indictment—a rule perhaps not applicable where the court determines, after verdict, the punishment.” 1 JOEL PRENTISS BISHOP, NEW COMMENTARIES ON THE CRIMINAL LAW § 948, at 572 (5th ed. 1872).

173. See id. at 572 n.4. All three cases clearly involve jury sentencing. In Ingram v. State, 39 Ala. 247 (1864), the trial judge “instructed the jury that they might look to the evidence of such [uncharged] acts, ‘in aggravation of the fine, but for no other purpose.’” Id. at 253. In Baker v. State, 4 Ark. 56 (1842), the court stated, “Here the prisoner was found guilty upon both counts in the indictment; for one of which, maiming, confinement in the penitentiary is the proper punishment, and to which he was properly sentenced by the jury.” Id. at 64–65. In Skains v. State, 21 Ala. 218 (1852), the trial court charged the jury that “they could not fine either of the defendants less than one hundred dollars, and might fine them five hundred, or imprison them six months.” Id. at 219.

174. Apprendi, 530 U.S. at 543 (O’Connor, J., dissenting).
have led the Court into simplistic and untenable solutions to the problem of defining a criminal offense for constitutional purposes. This Part will examine five points from the theory of punishment that clarify the problem significantly. Then, for purposes of illustration, this more sophisticated conceptual vocabulary will be applied to the problem of offense elements versus affirmative defenses. This discussion will point toward a solution of the *Apprendi* puzzle, to be advanced in Part III.

A. THE RELEVANCE OF PUNISHMENT THEORY

1. Determinate and Indeterminate Sentencing

The *Williams-McMillan* faction is particularly anxious about the fate of the Federal Sentencing Guidelines because real-offense sentencing is threatened by the unrestricted version of *Apprendi* that these Justices see in the offing. 175 Obviously, a real-offense sentencing system is inconsistent with a rule that requires all facts that authorize or increase punishment to be proved to the jury. However, several commentators have noted an irony in a reliance on *Williams* to defend real-offense sentencing under the Guidelines. *Williams* is a product of the indeterminate sentencing philosophy of the mid-twentieth century. As *Williams* itself shows, penal theory and policy in that era were oriented toward rehabilitation and characterized by a conscious, indeed strident, rejection of retribution or just deserts as the leading aim of punishment. 176 The law and institutions of punishment reflected this philosophy: Judges set indeterminate sentences so that prison and probation officials would have latitude in setting rehabilitative goals and in determining when the offender had met those goals. 177 In contrast, the adoption of determinate sentencing schemes such as the Guidelines is said to have been motivated by a change in penal policy toward retribution and away from rehabilitation. 178 The aim was “truth in sentencing” so that criminals would serve the actual sentences they had received from the sentencing court. 179 Thus, as one federal judge has noted, “*Williams*’ whole underpinning is the . . . very system the Sentencing Reform Act of 1984 was

175. See id. at 549–50.


178. See STITH & CABRANES, supra note 24, at 53, 55 (describing the influence of Commissioner Paul Robinson’s “retributive” theory). But see infra note 187.

179. STITH & CABRANES, supra note 24, at 40 (describing the elimination of parole as one of the objectives of the Sentencing Reform Act of 1984).
passed to destroy!"\textsuperscript{180}

From a theoretical perspective, the truth is both more complex and more dismaying. In fact, the Sentencing Guidelines do not pursue a genuine retributive philosophy. Retribution as a goal of punishment implies strict proportionality. A consistent retributivist would find the infliction of any punishment greater than the deserved punishment to be a violation of moral duty.\textsuperscript{181} As Justice Breyer, the Guidelines’ principal architect,\textsuperscript{182} frankly acknowledged, the Sentencing Guidelines Commission adopted no such philosophy and made no attempt to determine proportionality in this sense because of the “inherent subjectivity” of desert and proportionality.\textsuperscript{183} Instead, the Commission purported to model the Guidelines on prevailing sentences in the federal courts.\textsuperscript{184} But as Kate Stith and José Cabranes have shown, the Commission systematically set sentencing ranges higher than these prevailing sentences.\textsuperscript{185} Furthermore, as Stith and Steve Y. Koh have demonstrated, the excessive sentencing ranges in the Guidelines reflect a legislative purpose to “get tough” with criminals, regardless of proportionality and a principled retributivism.\textsuperscript{186}

The theory that actually animates the Guidelines is a simple-minded consequentialism.\textsuperscript{187} The aim is neither rehabilitation nor retribution,\textsuperscript{188} but rather

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{180} United States v. Wise, 976 F.2d 393, 408 (8th Cir. 1992) (en banc) (Arnold, C.J., dissenting); \textit{see also} Stith & Cabranes, supra note 24, at 152 (quoting Judge Arnold and concluding that “[t]he old and new regimes of ‘real-offense’ sentencing thus differ both in theory (rehabilitative versus punitive) and in practice (discretionary versus mandatory”).
\item \textsuperscript{181} \textit{See}, e.g., \textit{Andrew von Hirsch, Censure and Sanctions} 15–16 (1993) (describing the principle of proportionality in a retributive punishment scheme as ruling out undeservedly severe punishments).
\item \textsuperscript{182} \textit{Breyer, supra note 161, at 16.}
\item \textsuperscript{183} \textit{See Stith & Cabranes, supra note 24, at 58.} The basic approach of a strict, harm-based sentencing calculus was proposed in a first draft by Paul Robinson, but that draft failed to gain much support because it was perceived as overly complex. \textit{See id. at 53–54}.
\item \textsuperscript{184} \textit{Breyer, supra note 161, at 16.}
\item \textsuperscript{184} \textit{See id. at 17–18.}
\item \textsuperscript{185} \textit{See Stith & Cabranes, supra note 24, at 59–66.}
\item \textsuperscript{186} \textit{See generally Kate Stith & Steve Y. Koh, The Politics of Sentencing Reform: The Legislative History of the Federal Sentencing Guidelines, 28 Wake Forest L. Rev. 223 (1993).}
\item \textsuperscript{187} \textit{Stith and Cabranes make the common mistake of taking the Guidelines’ focus on the harm done in offenses as evidence of a retributive approach to punishment. They describe Commissioner Paul Robinson’s original proposal for a strict, harm-based sentencing calculus as being aimed at “retribution,” Stith & Cabranes, supra note 24, at 53, and as “a deontological, just-deserts approach,” id. at 59. Just the opposite is true. A focus on harm is characteristic of a consequentialist theory of punishment, whereas a deontological theory of punishment focuses on wrongdoing. See infra notes 267–69 and accompanying text (explaining one of the philosophical origins of this difference). Robinson has made his consequentialist philosophy quite clear in his subsequent work. See infra note 192 and accompanying text.}
\item \textsuperscript{188} \textit{A great deal of confusion in the theory of punishment can be avoided by distinguishing between the aims of punishment and theories of punishment. References to the deterrence theory of punishment or the retributive theory of punishment are common, but deterrence and retribution—like incapacitation, education in social norms, public catharsis, and so on—are merely aims or functions of punishment. Whether and how any of these functions justifies punishment depends on a moral or ethical theory that attributes a justifying effect to them. Theories of punishment, then, come in consequentialist, deontological, and virtue ethics models. Nor is the justification of punishment the only question}
\end{itemize}
\end{footnotesize}
lengthy incapacitation and an "economy of threats" deterrence.\textsuperscript{189} The irony is that this fact gives credence to Justice Breyer's reliance on Williams. The rehabilitative sentencing philosophy of the Williams era also reflected a consequentialist theory of punishment—just a more sophisticated one. The aim of punishment under both mid-twentieth-century indeterminate sentencing schemes and late-twentieth-century determinate sentencing schemes was the maximization of social welfare. The difference lay only in the chosen means: a conscientious effort to reintegrate the offender into civilized society versus brute deterrence and dishartened warehousing.

Given that a consequentialist theory of punishment lies beneath both Williams and the Federal Sentencing Guidelines, a better understanding of this theory may help us solve the Apprendi puzzle. One feature of the consequentialist theory of punishment is particularly relevant: It has no independently viable conception of criminal fault.\textsuperscript{190} Fault, desert, and culpability are retrospective considerations: They pose the question whether blame and punishment are appropriate in light of past wrongdoing. Because consequentialism is inveterately forward-looking, it uses such concepts, if at all, only in a derivative fashion. Thus, H.L.A. Hart argued that the concept of desert does not serve any justifying purpose in punishment; it serves only as a side constraint that maximizes liberty and social welfare.\textsuperscript{191} Similarly, Paul Robinson (one of the Guidelines' architects) has argued that culpability is a feature of the criminal law because, by satisfying the widespread intuition that only the culpable should be punished, the criminal law is more effective than it otherwise would be in promoting social welfare.\textsuperscript{192}

One consequence of the uninterrupted dominance of consequentialist conceptions of punishment through most of the last century was a failure to develop an adequate understanding of desert and fault, or an adequate vocabulary with which to translate these fundamental concepts into doctrine.\textsuperscript{193} This deficiency affected the theory and practice of both indeterminate and determinate sentencing. Thus, the vocabulary with which the

dealt with by a theory of punishment. Other outstanding issues include the nature of fault, the rationale of excuse, the outlines of prohibitory norms, proportionality in punishment, and so on.

\textsuperscript{189} The term "economy of threats" was coined by H.L.A. Hart, and used to describe deterrence that operates by explicit instrumental reasoning aimed at avoiding pain. H.L.A. HART, Legal Responsibility and Excuses, in PUNISHMENT AND RESPONSIBILITY: ESSAYS IN THE PHILOSOPHY OF LAW 29, 40-44 (1968). Hart contrasts this with deterrence that involves the internalization of legal rules as obligations that people incorporate into their other aims and plans. See id. at 44-50. The great achievement of Hart's consequentialist theory of punishment was to explain this latter, more pervasive and important, kind of deterrence. See Kyron Huigens, The Dead End of Deterrence, and Beyond, 41 WM. & MARY L. REV. 943, 984–87 (2000).

\textsuperscript{190} See Huigens, supra note 189, at 956–80.

\textsuperscript{191} See Hart, supra note 189, at 39–40.


\textsuperscript{193} See Huigens, supra note 189, at 966–71.
Williams-McMillan and Winship-Mullaney factions have carried out their argument—mitigation, culpability, element, sentencing factor, affirmative defense—is plagued by vagueness and riddled with ambiguity. Therefore, in the four subsections that follow, I propose to examine some of these terms and replace them with a more rigorous vocabulary drawn from a nonconsequentialist theory of punishment.

2. Mitigation

The Apprendi dissenters feared that the rule stated by the Court—"Other than the fact of a prior conviction, any fact that increases the penalty for a crime beyond the prescribed statutory maximum must be submitted to a jury, and proved beyond a reasonable doubt"—eventually would be stripped of its limitations. The real Apprendi rule would then be the one stated in Justice Thomas’s concurrence: "[A] ‘crime’ includes every fact that is by law a basis for imposing or increasing punishment (in contrast to a fact that mitigates punishment)." Certainly the Thomas formulation represents the core principle of Winship and Mullaney. However, the concepts of aggravation and mitigation on which the Thomas formulation turns are problematic.

Any fact that seems to aggravate an offense can be framed as a mitigating fact, depending on how the offense is defined. This is the point at the heart of Justice O’Connor’s reductio. If a state may not employ the aggravating fact of serious bodily injury to enhance a sentence from a statutory maximum of ten years to twenty years, then it can, instead, increase the statutory maximum to twenty years and provide for the mitigation of the sentence if the offense did not result in serious bodily injury. This mitigating fact need not be proved to a jury beyond a reasonable doubt under Justice Thomas’s formulation of the rule because mitigation is expressly excepted from that rule. But clearly the spirit of the rule—and with it the constitutional requirements of Winship and Apprendi—will have been evaded.

Obviously, some term or terms other than “mitigation” and “aggravation” are needed to get at the concern that lies behind the Apprendi rule, but as yet we have no clear idea of what that concern is. As a starting point, we may notice a further difficulty with the terms “mitigation” and “aggravation.” While it is true that all aggravating factors can be rephrased as mitigating factors, the converse is not true. Some mitigators cannot be rephrased as aggravators. For example, mental deficiency or disease not amounting to legal insanity is a common ground for mitigating punishment. It appears as a statutory mitigator in most if not all capital punishment schemes. But there is no meaningful converse of

195. Id. at 501 (Thomas, J., concurring).
this mitigator. No capital punishment statute lists ordinary intelligence or mental health as an aggravator. To take another example, youth is a mitigating factor in punishment. Youth, too, appears as a mitigator in most if not all capital punishment statutes. However, none of these statutes lists maturity as an aggravator, and it would be odd to find any sentence enhanced on the ground that the offender is middle-aged or elderly. The term "mitigation" is, then, ambiguous. Sometimes it refers to a consideration that has an aggravating converse, and sometimes it does not.

This ambiguity plays a role in Justice Stevens's opinion for the Court in *Apprendi* and may account for some of the *Winship-Mullaney* faction's misplaced confidence that its basic position is defensible. One of Justice Stevens's responses to Justice O'Connor's reductio is the argument that "the principal dissent ignores the distinction the Court has often recognized between facts in aggravation of punishment and facts in mitigation." This is a surprising argument given that, as we have seen, the interchangeability of aggravation and mitigation is one reason that Justice O'Connor's argument works. But Justice Stevens has in mind the other kind of mitigation, the kind that does not have an aggravating converse, and he chooses an example that makes his argument work. He writes,

If the defendant can escape the statutory maximum by showing, for example, that he is a war veteran, then a judge that finds the fact of veteran status is neither exposing the defendant to a deprivation of liberty greater than that authorized by the verdict according to statute, nor is the Judge imposing upon the defendant a greater stigma than that accompanying the jury verdict alone.

Veteran status, of course, has no aggravating converse. It is inconceivable that any judge or legislature would increase punishment on the ground that the offender never has served in a war. As a result, a statute that treated veteran status as a sentencing factor would not be vulnerable to Justice O'Connor's argument.

In order to see why mitigation and aggravation work in these ways, and to get at the fundamental concerns of the two factions on the Court, we need to examine the idea behind mitigation and aggravation—namely, culpability.

draft version of section 210.6 served as the model for most states' redrafting of their capital sentencing statutes following the decision in *Furman v. Georgia*, 408 U.S. 238 (1972).


198. Undoubtedly the words, "You're old enough to know better," have been spoken in sentencing, but they obviously would explain a refusal to mitigate because the offender is young, but old enough to be held responsible, rather than an intention to enhance a sentence because the offender is middle aged or elderly. For example, the sentencing judge would not go on to say, "You're lucky you're only 30. If you were 40, I'd throw the book at you, and if you were 50, I'd have to give you a life sentence."

199. *Apprendi*, 530 U.S. at 490 n.16.

200. *Id.*
3. Culpability

It is generally agreed that the question of defining a criminal offense for constitutional purposes concerns not only the justification of punishment in the individual case, but also the justification of a particular degree of punishment.201 This concern ordinarily is called culpability—criminal liability, unlike civil liability, requires culpability; punishment for a given offense generally increases with the level of culpability associated with the offense.202 Culpability is commonly defined as blameworthiness, but both culpability and blameworthiness are ambiguous terms.203 The resolution of this ambiguity sheds light on the fundamental concerns of the Williams-McMillan and Winship-Mullaney factions, in part because it parallels the ambiguity of mitigation.

One kind of culpability turns on the definition of a criminal offense because it consists of the presence of particular elements. For example, if a person shoots and kills a figure in the woods, believing that it is a deer, and the figure turns out to be a human being instead, then the killing is not a murder. A human being is dead, but the shooter’s belief that he was shooting at an animal precludes the requisite finding that he had a purpose to kill a human being.204 If the victim was walking in the woods during hunting season dressed in brown clothing, then the shooter probably would not be reckless or negligent either, precluding liability for manslaughter.205 Or suppose that a person suffers from insane delusions and kills a human being in the belief that the victim is an evil spirit in human form. This insane belief, too, will preclude a finding that the killer acted with any of the kinds of culpability, such as a purpose to kill a human being, required by the homicide statutes.206 The kind of culpability that is missing in these cases is commonly called mens rea, but as we will see below, a better term for it is “fault.”

A different kind of culpability, or its absence, can be discerned in cases in which the elements of an offense, including fault, are present. Suppose, for example, that a first-grader is angry with a classmate. He brings his father’s gun to school the next day and shoots the other child, fortunately not fatally. When he is asked why he did this, the child answers honestly that he wanted to hurt

201. See id. at 479–82; id. at 498–99 (Scalia, J., concurring); id. at 499–501 (Thomas, J., concurring); id. at 539–41 (O’Connor, J., dissenting); id. at 561–62 (Breyer, J., dissenting).
203. Huigens, supra note 32, at 1238 (distinguishing two senses of culpability or blameworthiness: fault and eligibility for punishment).
204. See 1 PAUL H. ROBINSON, CRIMINAL LAW DEFENSES § 62(b), at 246 (1984) (citing MODEL PENAL CODE § 2.04(1)(a) (1985)).
205. See MODEL PENAL CODE § 210.3 (1980) (defining manslaughter as a homicide committed recklessly); id. § 210.4 (defining negligent homicide).
206. See 1 ROBINSON, supra note 204, § 64(a), at 272–79 (describing the defense of diminished capacity in terms of the negation of a mental-state fault element of an offense by the defendant’s mental illness).
the other child. All the elements of assault are present here, but no criminal liability will be imposed, because a child of six is not subject to the criminal law at all. Or suppose that a person kills a coworker because he suffers from the insane delusion that he is John Wilkes Booth, that his coworker is Abraham Lincoln, and that Lincoln’s murder is historically inevitable. In this case, the purpose to kill a human being is present, but in most jurisdictions the defendant will be acquitted on grounds of insanity. The feature missing from these cases usually is referred to as culpability or blameworthiness, but both of these terms are also commonly applied to cases in which fault is missing. In order to avoid this ambiguity, it is better to refer to the kind of culpability that is missing from these cases as “eligibility for punishment.”

We refer to the absence of culpability in cases in which a person is simply ineligible for punishment, such as a child’s assault, and yet it seems odd to refer to the mere fact that one is eligible for punishment as culpability. In contrast, in cases in which fault is at issue, we talk not only about the absence of culpability, but also, on the positive side (so to speak), about the varying degrees of culpability. In this respect, the ambiguity in culpability between eligibility and fault parallels the ambiguity in mitigation between mitigators for which there is no aggravating converse and those for which there is a corresponding aggravator. The reason for this parallel, of course, is that culpability is the idea behind mitigation and aggravation. A crime is aggravated when it is accompanied by a high degree of fault—for example, when a homicide is committed purposely. A crime is mitigated when the actor is less at fault, such as when a homicide is committed negligently. And a crime also is mitigated or simply excused when some attribute of the actor makes him ineligible for the punishment that ordinarily would be imposed.

Mens rea is a poor term for fault because it denotes a mental state, whereas in fact fault is not limited to mental states. Negligence, for example, is by definition not a mental state, but it is a well-recognized form of criminal fault. There is a school of thought that has contended that criminal liability ought to be limited to prohibited acts accompanied by

---

207. See Model Penal Code § 211.1(2)(b) (1980) (defining aggravated assault as purposely or knowingly causing bodily injury to another with a deadly weapon).

208. Common law categorically prohibited criminal liability for a child under age seven, and presumptively prohibited it for a child under age fourteen. 4 Blackstone, supra note 97, at *23–24. States generally provide juvenile court jurisdiction over young wrongdoers, with various age limits. See 2 Robinson, supra note 204, § 175(a)–(b), at 321–30. A number of states also carry forward the common-law limitations, although the effect of these provisions in light of juvenile court jurisdiction is unclear. See Model Penal Code § 4.10 cmt. 2, at 273–74 (1985).

209. See Christopher Slobogin, An End to Insanity: Recasting the Role of Mental Disability in Criminal Cases, 86 Va. L. Rev. 1199, 1208–14 (2000) (describing the insanity defense). Slobogin notes that five states—Idaho, Kansas, Montana, Nevada, and Utah—have abolished the insanity defense. Id. at 1214.

mental states.\textsuperscript{211} The Model Penal Code's inclusion of criminal liability for negligence was opposed on this ground.\textsuperscript{212} But the prevailing view is that negligence is a kind of fault, and this objective, or unintentional, kind of fault is in fact pervasive in the criminal law. Those who would have barred negligence as a kind of fault have lost similar battles over felony murder,\textsuperscript{213} depraved heart murder,\textsuperscript{214} transferred intent,\textsuperscript{215} voluntary intoxication, and so on.

\begin{itemize}
  \item \textsuperscript{211} See, e.g., Jerome Hall, \textit{Negligent Behavior Should Be Excluded from Penal Liability}, 63 \textit{COLUM. L. REV.} 632, 635 (1963).
  \item \textsuperscript{212} See \textit{MODEL PENAL CODE} § 2.02 cmt. 4, at 243–44 (1985).
  \item \textsuperscript{213} Under the doctrine of felony murder, a defendant can be convicted of murder if a death occurs in the course of his commission of a specified felony, even if the defendant is unaware of the risk of death. See \textit{MODEL PENAL CODE} § 210.2 cmt. 6, at 30 (1980). In spite of vilification by generations of commentators, see, e.g., James J. Tomkovicz, \textit{The Endurance of the Felony-Murder Rule: A Study of the Forces That Shape Our Criminal Law}, 51 \textit{WASH. & LEE L. REV.} 1429, 1431 n.9 (1994) (collecting critical commentary), and reformers, see, e.g., \textit{MODEL PENAL CODE} § 210.2 cmt. 6, at 29–42 (1980) (offering a critical historical assessment of the rule, which in its pristine form is "indesensible in principle"), the doctrine endures. The Model Penal Code itself incorporates the felony murder doctrine into its definition of murder that requires a showing of recklessness under circumstances manifesting an extreme indifference to human life. See \textit{id.} § 210.2(1)(b) (providing that recklessness and extreme indifference will be presumed when death occurs as the result of some felonies). It also is to be found in its original form in most states. See, e.g., \textit{ALABAMA CODE} § 13A-6-2(a)(3) (1994); \textit{ARKANSAS CODE ANN.} § 5-10-101(a)(1) (Michie 1997); \textit{COLORADO REV. STAT. ANN.} § 18-3-102(1)(b) (West 1999 & Supp. 2000).
  \item \textsuperscript{214} The doctrine of depraved heart murder authorizes punishment for murder when

\begin{itemize}
  \item the act [is] prompted by, or the circumstances indicate that it sprung from, a wicked, depraved or malignant mind—a mind which, even in its habitual condition, and when excited by no provocation which would be liable to give undue control to passion in ordinary men, is cruel, wanton or malignant, reckless of human life, regardless of social duty.
\end{itemize}

Maher v. People, 10 Mich. 212, 218 (1862). From the point of view of the intentional-states construction of fault, this doctrine is utterly baffling. See Bernard E. Gegan, \textit{More Cases of Depraved Mind Murder: The Problem of Mens Rea}, 64 \textit{ST. JOHN’S L. REV.} 429, 432–37 (1990) (describing the unsuccessful efforts of some courts to reduce the doctrine to recklessness). This is because the doctrine addresses a state of character instead of a state of mind. In such cases,

\begin{itemize}
  \item [it] may become clear that the actor has a character flaw more blameworthy than that shown by a single indiscretion; it may even be established that he simply holds human life without value. This is not a specific mental state formed at the moment of action, such as intent or reckless disregard. Rather, it is an immoral predisposition to harm.
\end{itemize}


accomplice liability, unreasonable mistake, and strict liability. In each of these instances, fault is to be found not in a discrete mental

216. The construction of fault as intentional states implies that a severely intoxicated offender is innocent because of his severe intoxication. That is, if a person is extremely drunk, he cannot act with a fixed reference to any consequences. Logically, then, he cannot commit crimes that require him to have an intentional state regarding prohibited consequences. The Model Penal Code braves this implication when crimes of purpose and knowledge are concerned. MODEL PENAL CODE § 2.08(1) cmt. 1, at 354 (1985) ("[W]hen purpose or knowledge . . . must be proved as an element of the offense, intoxication may generally be added in disproof if it is logically relevant."). But the Code avoids this implication when recklessness is concerned. The Code contains a prohibition on proof of voluntary intoxication to negate the intentional state of recklessness, even though intoxication may well render a person unable to recognize and consciously disregard risks. See id. § 2.02(2)(c) (defining recklessness as the conscious disregard of a substantial and unjustifiable risk). This inconsistency is justified with the vague claim of a "general equivalence" between recklessness in getting drunk and recklessness while one is drunk. See id. § 2.08 cmt. 1, at 359. This solution parallels that of the common law, which permitted proof of intoxication to rebut specific intent and barred its use to rebut general intent. Id. at 353–54. Constitutionally, a state may go further and prohibit proof of intoxication for crimes premised on purpose and knowledge as well. See Egelhoff v. Montana, 518 U.S. 37 (1996). Compare Stephen J. Morse, Fear of Danger, Flight from Culpability, 4 PSYCHOL. PUB. POL'Y & L. 250, 251 (1998) (expressing fear that Egelhoff represents a turn toward the incapacitation of dangerous persons in disregard of fault or desert), with Huigens, supra note 189, at 981–82, 1030–33 (defending Egelhoff and nonintentional fault against similar rule-of-law objections).

217. The Model Penal Code requires a purpose to facilitate the principal's wrongdoing for accomplice liability, having rejected a strong argument that knowledge of the principal's planned illegal conduct should suffice. MODEL PENAL CODE § 2.06 cmt. 6(c), at 313–20 (1985). However, many jurisdictions punish accomplices on a mere showing that the accomplice should have known that the principal would act illegally. See generally People v. Luparello, 231 Cal. Rptr. 832 (1987); Grace E. Mueller, Note, The Mens Rea of Accomplice Liability, 61 S. CAL. L. REV. 2169 (1988). This is the equivalent of liability for negligence. Given that, for most crimes, the principal ordinarily must act with a higher level of culpability than negligence, this is anomalous. It is not, however, constitutionally objectionable. In the area of accomplice liability, as in the area of intoxication and as a constitutional matter, the Supreme Court has separated the concept of criminal fault from mental states. See Tison v. Arizona, 481 U.S. 137, 158 (1987) (holding that "major participation in the felony committed, combined with reckless indifference to human life, is sufficient" to constitute fault for capital murder); cf: Andrew H. Friedman, Note, Tison v. Arizona: The Death Penalty and the Non-Triggerman: The Scales of Justice Are Broken, 75 CORNELL L. REV. 123, 145–50 (1989) (arguing that Tison authorizes liability without fault because the Tison brothers' conduct was neither purposeful nor reckless).

218. An unreasonable mistake precludes the formation of a culpable mental state, just as does a reasonable mistake. One implication of this fact is that an alleged rapist who is so stupid and callous that he cannot comprehend a woman's nonconsent to sex is, for that reason, not guilty of rape. Many jurisdictions have resisted this implication of the mental state's construction of fault. See SANFORD H. KADISH & STEPHEN J. SCHULHOFER, CRIMINAL LAW AND ITS PROCESSES: CASES AND MATERIALS 327 (6th ed. 1995) ("Most of the recent American cases permit a mistake defense [to rape], but only when the defendant's error as to consent is honest and reasonable."); Harriett R. Galvin, Shielding Rape Victims in the State and Federal Courts: A Proposal for the Second Decade, 70 MINN. L. REV. 763, 849 (1986) ("[A] majority of courts here have adopted the rule that only a reasonable mistake is a valid defense." (citing People v. Mayberry, 542 P.2d 1337, 1344–47 (Cal. 1975); and State v. Dizon, 390 P.2d 759, 769 (Haw. 1964))); Victoria J. Dettmar, Comment, Culpable Mistakes in Rape: Eliminating the Defense of Unreasonable Mistake of Fact as to Victim Consent, 89 DICK. L. REV. 473, 483–89 (1985) (summarizing cases outlining the unreasonable mistake defense in rape cases); see also Regina v. Morgan, 1976 A.C. 182 (affirming rape convictions of unreasonably mistaken rape defendants); Lani Anne Remick, Comment, Read Her Lips: An Argument for a Verbal Consent Standard in Rape, 141 U. PA. L. REV. 1103, 1108 ("At least one court has even gone so far as to suggest that there is no mens rea for rape.").

219. Strict liability usually is defined as criminal liability without culpability or as criminal liability without proof of a mental state, see, e.g., United States v. Cordoba-Hincapie, 825 F. Supp. 485 (1993)
state, but in a broader set of facts surrounding the offense. In the drafting of crimes, we usually do state the requirement of criminal fault in terms of discrete mental states, but we do this for reasons of legality—because such rules are more likely to keep criminal liability within appropriate bounds in a constitutional democracy—and not because of the nature of criminal fault. In some instances, we have found that rules about mental states actually are not the best rules about fault. Death sentences, for example, have seemed to be more defensible when they are the product of a detailed inquiry into fault-based aggravating and mitigating factors than when they are the product of a broad and vague inquiry into “malice aforethought” or “premeditation.”

(Weinstein, J.) (invoking both of these definitions), and usually is said to be unjust for precisely these reasons, see, e.g., MODEL PENAL CODE § 2.05 cmt. 1, at 283 (1985). But this is not quite right. As the foregoing examples indicate, we impose criminal liability without proof of a mental state in many situations. And in most cases of strict liability, one can identify some fault, usually negligence, somewhere in the transaction. See Mark Kelman, Strict Liability: An Unorthodox View, in 4 ENCYCLOPEDIA OF CRIME AND JUSTICE 1512, 1516–17 (1983); Kenneth W. Simons, When Is Strict Liability Just?, 87 J. CRIM. L. & CRIMINOLOGY 1075, 1121–22 (1997). Strict liability is unjust because the prosecution is relieved of the task of proving fault explicitly, while the defendant simultaneously is deprived of any formal avenue of rebutting the inference that he was at fault. In other words, strict liability is objectionable because it violates due process or the principle of legality.

220. Sometimes, the identification of intentionality with fault leads the devotees of an intentional-states construction of fault simply to decree the existence of facts to match intentional states. For example, suppose that a man has sex with a nine-year-old girl and believes that she is eleven. The law defines sex with a person who is younger than ten as a class A felony, and sex with a person who is at least ten and under twelve as a class B felony. Our defendant is guilty of neither crime on a strict intentional-states construction of fault: The act occurred, but no corresponding intentional state can be proved for the class A felony; the intentional state is present, but no corresponding act can be proved for the class B felony. Section 2.04(2) of the Model Penal Code decrees guilt for the class B felony in such cases because the defendant would have been guilty of the offense “had the situation been as he supposed.” MODEL PENAL CODE § 2.04(2) (1985).

221. The Supreme Court has recognized this point, at least regarding homicide:

A narrow focus on the question of whether or not a given defendant “intended to kill,” however, is a highly unsatisfactory means of definitively distinguishing the most culpable and dangerous of murderers.... Some unintentional murderers may be among the most dangerous and inhumane of all—the person who tortures another not caring whether the victim lives or dies, or the robber who shoots someone in the course of the robbery, utterly indifferent to the fact that the desire to rob may have the unintended consequence of killing the victim as well as taking the victim’s property. This reckless indifference to the value of human life may be every bit as shocking to the moral sense as an “intent to kill.”


222. Benjamin Cardozo offered this classic assessment of capital murder statutes that employ premeditation to distinguish the most culpable murderers:

If intent is deliberate and premeditated whenever there is choice, then in truth it is always deliberate and premeditated, since choice is involved in the hypothesis of the intent. What we have is merely a privilege offered to the jury to find the lesser degree when the suddenness of the intent, the vehemence of the passion, seems to call irresistibly for the exercise of mercy. I have no objection to giving them this dispensing power, but it should be given to them directly and not in a mystifying cloud of words.

BENJAMIN N. CARDOZO, What Medicine Can Do for Law, in LAW AND LITERATURE 70, 100 (1931). The
As noted, both sides in the *Apprendi* controversy recognize that the degree of punishment varies with the degree of culpability. What the Justices have in mind here, apparently, is degrees of fault. But the point requires further elaboration. Both kinds of culpability—fault and eligibility—are considered in the determination of a proportionate punishment. For example, a defendant who suffers from a mental illness that is not severe enough to merit an excuse may be ineligible for a punishment as severe as that which would be imposed on a healthy offender. In other words, the offender’s mental illness mitigates his punishment and does so in a way that does not have an aggravating converse. However, we cannot determine proportionality solely on considerations of this kind. Obviously, the gravity of the wrongdoing and the degree of fault entailed in it also must be considered in the determination of a proportionate punishment. All else being equal in the cases of two mentally ill offenders, a proportionate punishment for the one who murders will be higher than a proportionate punishment for the shoplifter.

4. Elements

Both factions on the Court frame the *Apprendi* controversy as a matter of designating which facts are elements of offenses and which are sentencing factors. The *Apprendi* rule refers to facts that increase the available sentence, and the Thomas formulation—which represents the unqualified core of *Apprendi*—refers to facts that are the basis for imposing or increasing punishment. But neither the idea of offense elements nor any of these categories of fact is particularly useful in understanding the idea of a criminal offense or its corollaries concerning burdens of proof. Punishment is justified for criminal wrongdoing, but criminal wrongdoing is a matter of norms, not facts, and the definition of criminal wrongdoing is never entirely confined to the elements of an offense.

An emphasis on discrete facts in the analysis of the criminal law might be appropriate if justified punishment followed automatically from the infliction of certain harms. But it does not. As any competent analysis of inchoate offenses and criminal fault will make clear, the infliction of harm is neither a necessary

drafters of the Model Penal Code dealt with this difficulty by means of Section 210.6 of the Code, which provides for a bifurcated hearing following conviction for murder, at which the jury weighs aggravating and mitigating features of the offense against one another in order to determine whether the sentence will be death or something less. Model Penal Code § 210.6 (1980). This solution was adopted by most states in the wake of *Furman v. Georgia*, 408 U.S. 238 (1972), and was approved by the Supreme Court in *Gregg v. Georgia*, 428 U.S. 153 (1976). Most aggravating and mitigating factors in capital sentencing statutes are fault considerations. See Huigens, supra note 32, at 1268–69, 1272–75.


224. See id. at 486–90; id. at 551–55 (O’Connor, J., dissenting); id. at 557–58 (Breyer, J., dissenting).

225. See *Apprendi*, 530 U.S. at 488–90.

226. See id. at 499–501 (Thomas, J., concurring).
nor a sufficient condition for justified punishment.\textsuperscript{227} We impose punishment because a person has violated a prohibitory norm that has been properly adopted as a positive law. Not all prohibitory norms involve harm, but all prohibitory norms in the criminal law include the notion of fault.\textsuperscript{228}

The enactment of prohibitory norms into criminal law is not limited to the statutory definition of criminal offenses. Very often, part of the prohibitory norm is split off and enacted under the rubric of a justification defense. For example, in the matter of homicide, the killing of another human being usually is not done for a reason that makes the killing right instead of wrong. It would make no sense to require the prosecution, in every case, to plead and come forward with evidence that no such reason exists. However, if the defendant acted in self-defense or within the scope of his legal authority, then indeed the killing might have been right instead of wrong.\textsuperscript{229} Because these instances are rare, we require the defendant to plead and come forward with evidence of this justification—which we could not do if the matter of justification were included in the offense definition.\textsuperscript{230} For this reason, we state justifications such as self-defense and the lawful use of deadly force in separate defense statutes.

It is important to note, however, that even though the justification is not included in the offense definition, and even though we impose the burdens of pleading and production on the defense, the burden of persuasion on justification defenses rests with the prosecution.\textsuperscript{231} That is, the prosecution must prove the absence of self-defense, of lawful authority, or of other justification. The reason for this distribution is very simple. The prosecution bears the burden of proving that a crime has occurred, and a crime consists of the violation of a prohibitory norm. The absence of justification is part of the prohibitory norm, even though we enact it into law separately from the offense for the sake of clarity and convenience.

Fault is an even more distinctive part of the prohibitory norm. Fundamental fairness and the principle of legality—also known as due process—require us to state

\textsuperscript{227} In inchoate crimes, we impose criminal liability when no harm has been done. For example, in its attempts provision, the Model Penal Code does not assess the proximity of the defendant's actions to a completed crime. Instead, the Code asks only if the defendant has engaged in some conduct, however minimal, that corroborates his criminal intentions. \textit{Model Penal Code} § 5.01 (1985). This is not only criminal liability without harm; it is very nearly criminal liability for thoughts alone. Conversely, we do not impose criminal liability when a harm clearly has been inflicted, if we can find no fault or culpability. \textit{See Staples v. United States}, 511 U.S. 600, 605–06 (1994) (discussing the rule of statutory construction that requires proof of fault in criminal offenses); \textit{see also Model Penal Code} § 2.02(1) (1985) (requiring proof of some kind of culpability regarding each material element of all Code offenses).

\textsuperscript{228} The exception would be strict liability offenses, but even there one usually can find objective fault considerations that justify punishment. \textit{See generally Kelman, supra note 219}.

\textsuperscript{229} \textit{See 2 Robinson, supra note 204, § 131(a), at 69–74} (explaining self-defense in terms of the balance of evils); \textit{id.} § 142(e), at 135–38 (explaining defense of law enforcement authority in terms of the balance of societal and personal interests).

\textsuperscript{230} \textit{Id.} § 142(a), at 123–24 (explaining that the burden of production for the defense of law enforcement authority is on the defendant).

\textsuperscript{231} \textit{Id.} at 124.
in advance, in legislation, everything in the prohibitory norm that a person needs to know in order to comply with that norm.\footnote{232} But this does not exhaust the prohibitory norm. The criminal law not only prohibits certain conduct or the infliction of certain harms; it prohibits certain conduct done with fault, and the infliction with fault of certain harms. Even though fault is necessary to criminal wrongdoing, fault is an issue ex post, in adjudication. Fault is an aspect of wrongdoing—having a purpose is not against the law; having a purpose to kill is—but fault is a rule of decision or adjudication, not a conduct rule. The matter of fault is addressed to legal authorities, in order to guide their legal decisionmaking, not to the general public, in order to guide ordinary conduct.\footnote{233} Even though mental states representing fault usually are included in the definition of an offense, this is not done in order to guide conduct, but in order to guide the adjudication of the charge.\footnote{234}

Finally, there is a dimension of fault that is not stated ex ante in rules at all. Just as there is more to fault than mens rea, there is more to the adjudication of fault than the finding of mental states. As I will explain at greater length below, fault is an inference to the effect that the person who has violated a prohibitory norm has done so in a way that calls the quality of his practical reasoning into question.\footnote{235} This inference is drawn in the course of the jury’s adjudication of wrongdoing, and for this reason, it escapes notice and analysis if one focuses, as both sides in \textit{Apprendi} did, on the statutory definition of offenses.

It is important to understand fault’s relationship to the adjudication of wrongdoing because it tells us where the burden of persuasion for fault should lie. Just as the burden to disprove justification rests with the prosecution because the absence of justification is part of the prohibitory norm, so it is with fault. The burden of persuasion on fault rests with the prosecution because fault is part of the prohibitory norm and the adjudication of fault is inseparable from the adjudication of wrongdoing.

5. Affirmative Defenses

The \textit{Williams-McMillan} faction also would do well to examine more closely the notions of “element” and “facts that justify punishment.” This faction’s principal argument is vulnerable to the objection that it reduces due process and other trial guarantees to nullities because legislatures can redesignate elements

\begin{footnotes}
\footnote{232. See Model Penal Code § 2.04(3)(a) (1985) (providing a defense of ignorance of law when a law has not been published or otherwise made available prior to the conduct alleged).}
\footnote{234. See Dan-Cohen, supra note 233, at 649 (noting that offense definitions contain decision rules as well as conduct rules); Paul H. Robinson, \textit{Rules of Conduct and Principles of Adjudication}, 57 U. Cin. L. Rev. 729, 739 (1990) (distinguishing between “culpability mental elements, which function as part of the principles of adjudication, and criminalization mental elements, which are a necessary component of the rules of conduct”).}
\footnote{235. See infra Part III.c.}
\end{footnotes}
as sentencing factors or affirmative defenses at will. This difficulty could be avoided if the argument appealed to some notion that was similar to an "element" or "facts that justify punishment," but that was not entirely a matter of positive law.

Consider the category of "affirmative defenses." The Justices define an affirmative defense, as they do a sentencing factor, in opposition to an offense element. Whereas the prosecution must prove offense elements, the defendant must prove affirmative defenses. But when it is used in this way, the notion of an affirmative defense is essentially meaningless. The term "affirmative defense" is triply ambiguous, because it can mean a defense that is not merely the negation of the elements of the offense, a defense on which the defendant has the burden of production, or a defense on which the defendant has the burden of persuasion. Given that the question at issue in the Mullaney and Patterson cases was who ought to carry the burden of persuasion, the term "affirmative defense" was irrelevant on the second meaning and question-begging on the first and third.

The Williams-McMillan faction would do better to focus on the nature of wrongdoing, the ambiguity of culpability, and the different implications that these theoretical points have for the assignment of burdens of persuasion. We have seen that if the accused defends on the ground of justification, then he contends in essence that his action was not wrong at all, but right. The prosecution must disprove this claim because the prosecution always must prove wrongdoing in order to justify punishment. Self-defense is an example of this kind of defense. Similarly, one kind of culpability, fault, is a matter for the prosecution to prove, by virtue of its being an aspect of criminal wrongdoing. As a consequence, if the accused defends on the ground that he is not at fault, then the prosecution must disprove this claim. Mistake of fact is an example of this kind of defense. But the situation is quite different when defenses of eligibility for punishment are concerned. On eligibility defenses—defenses based on claims such as insanity and minority, for example—and on claims of mitigation that are analogous to eligibility defenses, the defendant carries the burden of persuasion.

Eligibility for punishment, or nonexcuse, is a fact that justifies punishment, but it is less a discrete doctrinal requirement than a necessary background assumption that a system of just punishment makes. The general application of the penal law to those who fall within its jurisdiction reflects a presumption that

237. See MODEL PENAL CODE § 2.04(1)(a) (1985) ("Ignorance or mistake as to a matter of fact or law is a defense if: (a) the ignorance or mistake negatives the purpose, knowledge, belief, recklessness, or negligence required to establish a material element of the offense.").
238. See Leland v. Oregon, 343 U.S. 790, 797 (1952) (due process permits state to impose burden of persuasion of insanity on defendant); In re Manuel L., 865 P.2d 718, 728 (Cal. 1994) (defendant bears the burden of persuasion on claim of minority status); see also ARIZ. REV. STAT. ANN § 13-703 (West 2001) (defendant in penalty phase of a capital case bears burden of persuasion on mitigating factors).
each member of society is capable of bearing the burdens that the law imposes—to respect property, not to kill other people, and so on. To claim an excuse is to claim that one is extraordinary in this regard. It is to claim an exemption from the general application of the penal law because the presumption of ordinary capability does not hold. The burden of persuasion is properly assigned to the person who makes such a radical claim. The defendant who invokes an excuse invokes a special kind of legal rule that effectively negates the force of vast stretches of law. It makes sense for the law to guard such a rule jealously and only rarely to permit a person to impose a legal obligation on all others to free him from his own legal obligations. For this reason, the burden of persuasion on defenses such as insanity is imposed on the defendant.

If the term “affirmative defense” has any meaning at all, then it is confined to defenses that pertain to eligibility. Only for these defenses can one give a reason grounded in theory and principle—as opposed to positive law—that the defendant, instead of the prosecution, should bear the burden of persuasion.

If the Williams-McMillan faction were to confine its use of the term “affirmative defense” to eligibility defenses and its use of the term “sentencing factor” to eligibility-based mitigators, then its principal argument would not be subject to the reductio advanced by the Winship-Mullaney faction. A legislature that created such an affirmative defense, or that defined such a sentencing factor, could give reasons for doing so. Even if the fact in question had served as, or could be portrayed as, an offense element, these reasons would make it clear that the offense had not been set on a slippery slope, and that the remaining facts that justify punishment—those pertaining to wrongdoing and fault—were not about to slide out of the offense definition and into the defense case or sentencing. The legislature, and a court reviewing its work, could point to a principled distinction between the fact that the legislature has chosen to treat as an affirmative defense or sentencing factor, and those facts that it has not, and logically cannot, remove from the offense definition.

B. PATTERSON RECONSIDERED

The foregoing paragraph describes a preliminary solution to the Apprendi puzzle in its suggestion that matters of wrongdoing and fault are for the jury and matters of eligibility are for the judge to decide. This preliminary solution is unlikely to be embraced by the Williams-McMillan faction because, in the simplest terms, it seems to cede most of the disputed territory to the Winship-Mullaney faction. Things are more complex than this objection supposes because the territory ceded to the Williams-McMillan faction is not insignificant, and in any event, the territory of each faction (to push the metaphor one step too far) overlaps considerably. But the objection does contain an important kernel of truth. My solution to this puzzle does not involve a reconciliation of the cases as they stand; instead, it entails overruling some of the cases because they are misconceived and misleading. Choices of this kind are a necessary part of any solution to this problem because the profound equipoise of competing lines of
authority *is* the problem. No solution to the *Apprendi* puzzle can retain the sense and objectives of both lines of authority without perpetuating most of the difficulty.

*Patterson v. New York,*239 among other cases, comes out differently when properly analyzed. A demonstration of this reversal will serve to summarize and illustrate some of the ideas presented in the foregoing section. The offense-elements-versus-affirmative-defense problem is simpler than the offense-elements-versus-sentencing-factor problem, yet the basic issue—the definition of an offense for constitutional purposes—is the same. Therefore, the following analysis of *Patterson* should set the stage for the similar but more complex analysis that *Apprendi* requires.

The central distinction drawn in *Patterson,* by means of which the Court freed the case from the governance of *Mullaney,* was patently unsound. The distinction started from the premise that Maine had defined murder by reference to the element of “malice aforethought,” and then shifted the burden to the defendant to disprove malice aforethought in order to mitigate the offense from murder to manslaughter.240 In contrast, New York had provided for an independently defined affirmative defense of “extreme emotional disturbance” that would mitigate the offense from murder to manslaughter without shifting the burden of persuasion on an element of murder to the defendant.241 This distinction begs the question. The distinction turns on the idea that extreme emotional disturbance is not an element of the offense, which in turn rests only on the idea that it is an “affirmative defense.”242 Meanwhile, the notion of an affirmative defense was defined, if at all, to mean only that the elements of the defense were not elements of the offense.243

The only way to escape this circularity is to give some content to the notion of an affirmative defense. If extreme emotional disturbance were an eligibility-based mitigator, then it would be a consideration genuinely independent of the wrongdoing defined by the homicide statutes and a matter that a legislature could, in a principled way, require the defendant to prove. The term “affirmative defense” would be apt and meaningful and would genuinely distinguish New York’s law from Maine’s. On the other hand, if extreme emotional disturbance were a matter of fault, then it necessarily would be an aspect of wrongdoing that, like the rest of the wrongdoing alleged, would be for the prosecution to prove.

*Patterson* was decided incorrectly because extreme emotional disturbance is unquestionably a matter of fault, not eligibility. To begin with, the mitigator of extreme emotional disturbance has an aggravating converse. The Model Penal Code’s homicide provisions, on which the New York law in *Patterson* was

240. *Id.* at 212–14.
241. *Id.* at 215–16.
242. *Id.* at 206–07.
243. *Id.* at 205–06.
Based, define murder initially as a purposeful or knowing killing and manslaughter as a killing done recklessly. This basic structure of homicide degrees defined by mental states is supplemented with provisions that adjust the crime upward and downward, in mirror-image fashion, according to other, objective fault criteria—specifically, certain attitudes of the offender. A reckless killing that would otherwise be manslaughter is murder if it is done under circumstances that manifest an extreme indifference to human life. A purposeful or knowing killing that would otherwise be murder is manslaughter if it is done under the influence of "extreme mental or emotional disturbance." Fault, as we will see, is a matter of character, and the contrast here is between the killer who is so coldhearted that he does not know or care that he kills and the ordinary person of good character whose homicidal purpose is born of extraordinary circumstances. Within New York's statutory scheme for homicide, then, "extreme emotional disturbance" was a mitigator that had a corresponding aggravator of "extreme indifference to human life." This set of mitigating and aggravating attitudes indicates that extreme emotional disturbance is a fault-based mitigator.

One may object that extreme emotional disturbance describes an incapacitating condition that sounds more in eligibility and excuse. But this misses the historical significance of the term, which is descended from the law of provocation. Originally, provocation was simply "a logical inference from the early meaning of 'malice aforethought.'" The idea behind provocation was simply that one who acted in the heat of passion because of some provoking event could not have acted with malice aforethought. The Model Penal Code expanded the category of provocation from its traditional basis in circumstances—such as catching one's wife in flagrante delicto—that gave rise to a "heat of

244. Id. at 218-19 (Powell, J., dissenting).
245. See Model Penal Code § 210.2(1)(a) (1980) (defining murder); id. § 210.3(1)(a) (defining manslaughter).
246. Id. § 210.2(1)(b).
247. Id. § 210.3(1)(b).
248. Id. § 210.3 cmt. 5, at 54 (1980).
249. The history of these categories gives further support to the notion that the Model Penal Code's "extreme indifference" murder is the aggravating converse of its "extreme emotional disturbance" mitigator. The term "malice aforethought" was coined to distinguish those homicides that were sufficiently egregious in the circumstances or manner of their commission to merit death. Malice aforethought could be either express or implied malice. The doctrine of implied malice gave rise to the offense of depraved heart or depraved mind murder. Gegan, supra note 214, at 437. Depraved heart murder was committed when the act was "prompted by, or the circumstances indicate that it sprung from, a wicked, depraved or malignant mind—a mind which, even in its habitual condition, and when excited by no provocation which would be liable to give undue control to passion in ordinary men, is cruel, wanton, or malignant, reckless of human life, regardless of social duty." Maher v. People, 10 Mich. 212, 218 (1862) (emphasis added). The direct descendant of deprived heart murder, and the aggravating converse of provocation, is the Model Penal Code's category of murder committed "recklessly under circumstances manifesting an extreme indifference to the value of human life," the aggravating converse of extreme emotional disturbance. Model Penal Code § 210.2(1)(b) (1980).
passion." In doing so, the Code drafters gave the mitigation a more subjective cast, focusing on the emotional disturbance that such events cause. But in doing this, the drafters did not change the mitigator’s essential function as a fault element.

Historically, conceptually, and normatively, then, the crime and defense at issue in Patterson were indistinguishable from the crime and defense at issue in Mullaney. The fact that the Maine homicide law in Mullaney defined heat of passion and manslaughter explicitly in terms of negating malice aforethought reflected nothing more than Maine’s adherence to a less modern homicide scheme than that of New York. The basic Winship analysis ought not to have been affected by this circumstance, and it would not have been had the Court looked beyond the positive law to the history and theory of the substantive criminal law.

Furthermore, if the Patterson Court had distinguished fault from eligibility, then it would not have committed the pivotal error of equating extreme emotional disturbance with insanity. In Leland v. Oregon, the Court had decided that the Due Process Clause permitted states to place the burden of persuasion of insanity on the defense. In Rivera v. Delaware, the Court reaffirmed this holding against a Winship challenge. The Patterson Court supposed that it would have to reconsider these holdings if it were to bar New York from placing the burden of proving extreme emotional disturbance on the defendant. Nothing was further from the truth. Had the Court not hobbled itself with the ambiguous and question-begging concept of an “affirmative defense,” it easily could have distinguished Leland and Rivera from Patterson. The distinction turns on the fundamental difference between proving an eligibility-based defense such as insanity and proving a fault-based mitigator such as extreme emotional disturbance. Patterson ought to have held, consistently with Mullaney, that the prosecution must disprove a prima facie case of extreme emotional disturbance because the prosecution must prove a violation of the prohibitory norm in all its aspects, whether they appear in the definition of an offense, in a justification defense, or in a fault-based mitigation such as provocation or extreme emotional disturbance.

III. SOLVING THE PUZZLE

The solution to the Apprendi puzzle turns on the distinction between fault and eligibility, a related distinction between the functions of the jury and the functions of the sentencing court, and rule-of-law considerations that are unique to determinations of fault. Enough has been said to permit a statement of the

251. 343 U.S. 790 (1952).
254. See supra text accompanying notes 236–38.
solution in the first section of this Part. But the solution is not a simple one, and it requires considerable elaboration, defense, and illustration. This will be the business of the remaining sections of this Part. Sections III.B and III.C will draw out the implication that the Federal Sentencing Guidelines and state mandatory sentencing schemes, respectively, are unconstitutional if the Constitution protects traditional criminal law values at all. Section III.D will show how this solution to the *Apprendi* puzzle nevertheless preserves discretionary judicial sentencing. These conclusions follow from the fact that the *Williams-McMillan* argument breaks down under a proper understanding of the respective roles of the jury and the court in the adjudication of wrongdoing, fault, and proportionality in punishment. Section III.E considers four objections to this argument. Section III.F explains why this solution to the *Apprendi* problem, unlike that of the *Apprendi* Court itself, is not one that applies retroactively to invalidate millions of sentences.

A. THE SOLUTION, BRIEFLY STATED

Any fact may serve as an element of an offense, an element of a defense, or a sentencing factor. Accordingly, any fact may be determined by the jury or the judge, depending on the nature of its relevance to the case. The first step in solving the *Apprendi* problem, then, is to abandon the Court’s focus on categories of fact and to clarify the respective roles of the jury and the sentencing court.

If we have a meaningful jury right at all, then the jury’s function includes at least the determination of whether wrongdoing has been committed. Because fault is an aspect of wrongdoing, fault can be determined only by the body conducting the deliberations in which wrongdoing is determined, in those same deliberations. These are, of course, highly fact-specific determinations. The determination of wrongdoing and fault, therefore, is the jury’s province, as is the determination, beyond a reasonable doubt, of the facts on which wrongdoing and fault are predicated. No one disputes these principles where the mens rea elements of an offense are concerned. Much of the argument that follows turns on a central point—that criminal fault is not confined to those mental states. The manner and circumstances of the offender’s wrongdoing also serve as indicators of fault, and the jury’s determination of fault is inextricable from its determination of wrongdoing. The jury right that unquestionably extends to both the determination of wrongdoing and mens rea extends to all objective indicators of fault as well.

The sentencing court’s job is to determine a proportionate punishment. Questions of proportionality are like questions of excuse because proportionality in punishment turns in part on the capabilities and other personal characteristics that make the offender eligible or ineligible for a particular punishment. A sentencing court makes the proportionality determination and also determines the facts concerning the offender’s capabilities and characteristics as they bear on the proportionality of his sentence. Because these eligibility-based mitigators
are not aspects of wrongdoing, they need not be determined by the jury along with wrongdoing.

Unlike the excuses, the proportionality of a sentence also turns on the nature of the wrongdoing that the offender has committed and on the kind of fault with which he did it. The sentencing court thus considers wrongdoing and fault in arriving at a proportionate sentence. However, the sentencing court does not determine wrongdoing or fault, nor does it determine the facts on which those findings are predicated. It is bound by the jury’s decision in these matters and has no occasion to make its own determinations.

This last point is subject to an important qualification that preserves discretionary judicial sentencing, but not the Federal Sentencing Guidelines or state mandatory sentencing schemes. There is a distinction to be made between positive fault considerations and interstitial fault considerations. Positive fault considerations are those that are so closely related to some wrongdoing that they have been stated in positive law—matters such as extreme indifference to human life or a purpose to kill—as statutorily defined elements of an offense. Interstitial fault considerations are not enacted into positive law; they arise only ex post, in the course of adjudication. They consist of precisely those fault considerations that are so context-dependent and fine-grained\textsuperscript{255} that they cannot be stated ex ante in rules without distorting their significance and function.\textsuperscript{256} The jury makes determinations of both positive and interstitial fault in the adjudication of wrongdoing. But whereas the sentencing court is bound by the jury’s determination of positive fault considerations, as well as the jury’s determination of other elements of the offense, the sentencing court is free to make its own determinations of interstitial fault.

The basic implications of these principles for the \textit{Apprendi} puzzle can be stated succinctly. Fault considerations cannot be enacted into positive law as sentencing factors simply because fault is an aspect of wrongdoing and because the sentencing court never determines wrongdoing. If a fault consideration is enacted into positive law at all, it can only be as an offense element,\textsuperscript{257} that is, as a part of the prohibitory norm, the violation of which constitutes criminal

\textsuperscript{255. See infra notes 335–38 and accompanying text (offering a definition of “fine-grainedness” in this context).}

\textsuperscript{256. For example, consider the difference between a mercy killing and a contract killing. Both of these killings are murder, by virtue of the positive fault consideration of purpose or premeditation regarding death. But the two murders differ in fault because of the vastly different circumstances surrounding the two killings: acceding to the request of a loved one who is in intolerable pain versus making a profit from the coldhearted killing of a stranger. These surrounding circumstances that make such an enormous difference in how we evaluate the murders are interstitial fault considerations.}

\textsuperscript{257. A fault consideration may also be enacted as an element of a justification because the prosecution must prove that the defendant has acted with fault with respect to justifications as well as with respect to the material elements of an offense. This is why, for example, a mistake regarding justifying circumstances will defeat the justification defense if it is unreasonable, but will not defeat the defense if it is reasonable. See, e.g., People v. Goetz, 497 N.E.2d 41 (N.Y. 1986) (holding unreasonable mistake regarding self-defense deprives defendant of that defense). A positive fault consideration related to a justification is also for the prosecution to prove and for the jury to determine.}
wrongdoing. These positive fault considerations, like the other offense elements, are for the jury to determine. In contrast, interstitial fault considerations—those that, by definition, have not been enacted into positive law—can be determined not only by the jury, but also by the sentencing court. Similarly, the sentencing court has the authority to determine an offender’s capabilities and other characteristics. Any such eligibility-based mitigators could be enacted into positive law as sentencing factors that are to be determined by the sentencing court along with the facts on which they are predicated. However, these could be only advisory sentencing factors within a discretionary sentencing system, not binding sentencing factors within a mandatory sentencing scheme. A proportionate sentence is a value judgment that resists simplistic consequentialist formulations because it involves the simultaneous consideration of eligibility-based mitigators, wrongdoing, fault, and a variety of other matters as well. Sentencing involves the exercise of expressive—not instrumental—rationality, and this precludes the ex ante prescription of sentences by the legislative branch.

This set of principles solves the Apprendi puzzle because it breaks the Williams-McMillan syllogism. Sentencing courts have extensive discretion in sentencing and the authority to determine some facts. However, it does not follow from this alone that the legislature may specify any fact and enact it into positive law as a sentencing factor. The sentencing court can determine eligibility-based mitigators and can decide matters of interstitial fault in the course of deciding a proportionate punishment. But the sentencing court cannot determine wrongdoing or positive fault, and the determination of a proportionate sentence overall is a matter of expressive rationality that cannot be determined ex ante by a system of rules. The enactment of sentencing factors within a mandatory sentencing scheme is a constitutional loss, not a gain. The Due Process Clause and the Sixth Amendment prohibit such laws.

B. DETERMINATE SENTENCING REJECTED

As I suggested in the previous Part, my solution to the Apprendi puzzle takes sides. In the end, the Winship-Mullaney faction wins and the Williams-McMillan faction loses. Among the greatest fears of the latter faction is the possibility that the Federal Sentencing Guidelines and state determinate sentencing schemes may be rendered unconstitutional. This is indeed inevitable if the Constitution protects the traditional normative architecture of the criminal law at all. A further inquiry into the nature of proportionality in punishment helps to explain why this is so. More generally, this section explains the notion of expressive rationality and the relevance of this idea to the distribution of responsibility for sentencing.

The conclusion that eligibility-based mitigators could be enacted as sentencing factors may seem to leave room for a version of the Federal Sentencing Guidelines to operate. The distinctions drawn in the foregoing section suggest a reform that Stith and Cabranes have advanced. Stith and Cabranes argue that the Guidelines are flawed because they emphasize aspects of the offense but neglect aspects of the offender other than his criminal history, such as his youth or social background, that may also serve as sentencing factors. In the terms that I have advanced, aspects of the offense and criminal history usually are fault considerations, whereas matters such as youth or social background usually are eligibility considerations. My argument, then, would take the Stith and Cabranes reform one step further and confine the Guidelines sentencing factors to aspects of the offender other than his criminal history.

However, not even this much of the Guidelines can survive because any determinate sentencing scheme misconceives and distorts the proportionality determination that is at the heart of criminal sentencing. Again, Stith and Cabranes get at some of the difficulty in their critique of the Guidelines. They argue that the Guidelines are the product of an Enlightenment tradition that places too much faith in quantification and calculation, and that determinate sentencing neglects the ceremonial and aesthetic aspects of sentencing—the public theater that is essential to give punishment its proper gravity as a transaction between the individual and society. A more rigorous way to make these points is to say that the Guidelines are the product of the consequentialist theory of punishment and that, like most current versions of consequentialism, the Guidelines misconceive value and value judgments in noncognitive terms. The Guidelines distort sentencing because they fail to recognize that proportionality in sentencing is a matter of expressive rationality instead of a matter of instrumental rationality. Even if we were to limit sentencing factors to eligibility-based mitigators, the sentencing decision by nature cannot be given the precision that the Guidelines pretend to achieve.

These claims obviously require further elaboration and defense. Let me begin with the simple observation that the Guidelines and state determinate sentencing schemes consist of rules for sentencing. The difficulty is that sentencing is characteristically resistant to formulation in terms of rules. Rules are by nature over- and under-inclusive. When any rule confronts the complexity of real life, it tends to capture some cases that ought not to be captured and to miss some cases that ought to be captured, relative to the rule’s background justification. This inherent inaccuracy in rules is more tolerable in some areas of life and law than in others, and we tend to have little tolerance for it in the area of

259. STITH & CABRANES, supra note 24, at 74–75.
260. Id. at 169.
261. Id. at 81.
sentencing. For example, in the constitutional regulation of capital punishment, the Supreme Court initially demanded a rule-like approach to death sentencing, with the objective of reducing inequality and arbitrariness. Subsequently, when faced with some egregious cases of overinclusion, the Court demanded the individualized consideration of desert in death sentencing. These two demands cannot be reconciled, of course: Individualized consideration necessarily raises the possibility of arbitrariness and inequality—as when a jury exercises mercy—and gives up the benefits of a rule-based approach along with the detriments of over- and under-inclusion. Similarly, in Guidelines sentencing, the federal district courts have exhibited intolerance for the over- and under-inclusiveness of rules in sentencing. The judges have complained in the strongest terms of the Guidelines’ insensitivity to the individual merits of cases, and many have acquiesced in manipulations of the sentencing rules aimed at avoiding anomalous and disproportionate sentences.

Is this incompatibility between sentencing and rules necessary and inevitable, so that determinate sentencing systems are misconceived from the start? It is, and they are. Just sentencing results in a proportionate sentence, and a proportionate sentence is a value judgment. The central flaw of the Guidelines and other determinate sentencing systems is that their drafters and proponents misunderstand the nature of value judgments. They adhere to a noncognitivist conception of value, under which values are “subjective” in the way that matters of taste are said to be subjective: They are matters of feeling that are beyond the reach of reason. A resulting aspiration to minimize the role of value judgments and to give greater influence to matters of fact leads to the adoption of consequentialism in the theory of punishment: an emphasis on harm prevention instead of retribution for wrongdoing as the organizing principle of the criminal law, on the ground that harms are observable facts whereas the norms that inform the

265. See STITH & CABRANES, supra note 24, at 5 n.12 (citing and quoting extensive criticism of the Guidelines by judges).
266. MOLLY TREADWAY JOHNSON & SCOTT A. GILBERT, THE UNITED STATES SENTENCING GUIDELINES: RESULTS OF THE FEDERAL JUDICIAL CENTER’S 1996 SURVEY 10 (1997) ("Overall, then, these results suggest that respondents believe that parties often manipulate the guidelines through plea agreements—in part by stipulating facts—and that judges rarely scrutinize or reject such agreements.").
retributive approach are value judgments. Because value judgments are subjective and unreasoned, the argument goes, their enactment into law always threatens to be arbitrary and oppressive. In the noncognitivist and consequentialist tradition, retributive punishment long has been criticized in just these terms. The Guidelines describe the severity of crimes and the proportionality of punishment in terms of quantifiable harms. This approach reflects a consequentialist theory of punishment, in spite of the drafters’ pretense (common among consequentialists) of having adopted no theory at all, in favor of attention to the facts.

The difficulty is that valuation is not unreasoned, all value judgments cannot be assimilated into matters of taste, and the claim to a greater rationality that consequentialism makes on the basis of these mistaken premises is an empty pretense. The assumption that value judgments fall outside the realm of rational criticism, debate, or consensus reflects a recognition that valuation is premised in emotion. The mistake is to suppose that emotion is unreasoned. Feelings, bare somatic states, do not have a rational dimension, but emotions are not mere feelings. Emotions can distort rational thinking, but it is also demonstrably true

---

268. See, e.g., Breyer, supra note 161, at 16 ("Although guidelines motivated by a just deserts rationale would be cloaked in language and form that evoke rationality, using terms such as 'rank order of seriousness,' the rankings would not, in substantive terms, be wholly objective.").

269. K.G. Armstrong described the prevailing view of retributivism in intellectual circles at the mid-twentieth century in this way:

Retributive punishment is only a polite name for revenge; it is vindictive, inhumane, barbarous, and immoral. Such an infliction of pain-for-pain's-sake harms the person who suffers the pain, the person who inflicts it, and the society which permits it; everybody loses, which brings out its essential pointlessness . . . . The only people who today defend the retributive theory are those who, whether they know it or not, get pleasure and a feeling of virtue from seeing others suffer, or those who have a hidden theological axe to grind.

Armstrong, supra note 176, at 139.

270. STITH & CABRANES, supra note 24, at 55, 68–70. David Yellen explains how this calculus-of-harms approach has led to discrepancies in sentencing of the kind that the Guidelines were supposed to eliminate. One source of discrepancies in Guidelines sentencing is the difference between section 3D2.1(d) offenses, to which real-offense sentencing principles apply, and those which are not subject to section 3D2.1(d), for which the offense of conviction determines the sentence. See Yellen, supra note 29, at 434–38 (describing this difference and the resulting discrepancies). The only difference between these categories is that the harm done in the former category of offenses is readily quantifiable. “The Commission decided that for certain types of crimes, amounts would largely determine the sentence. Then, because they could easily count amounts for nonconviction offenses, they did. This mechanical approach, apparently unguided by principle, seems destined to accomplish very little of value.” Id. at 445.

271. See, e.g., Breyer, supra note 161, at 17; see also STITH & CABRANES, supra note 24, at 53, 55 (attributing the Guidelines' emphasis on quantifiable harms to Commissioner Paul Robinson’s influence); supra note 192 and accompanying text (explaining Robinson’s consequentialist theory of punishment). It is interesting to note that Richard Posner sees the Guidelines as one of the few places in which the economic analysis of law has had an impact on substantive criminal law. See Steve Kurtz, Sex, Economics, and Other Legal Matters: Judge and Scholar Richard A. Posner Speaks Out on the Clinton Impeachment, the Microsoft Case, and Nude Dancing, REASON, Apr. 1, 2001, at 36, 39 (quoting Posner as saying “there is definitely an economic flavor in the federal sentencing guidelines”).
that rational thought is impossible without the capacity for emotion.\textsuperscript{272} Emotions have truth value; that is, one can be mistaken in one's emotions.\textsuperscript{273} Emotions can be changed through a process involving their articulation and rational criticism.\textsuperscript{274} And just as emotions are subject to rational evaluation, so are the valuations and motivations that are premised in emotion.\textsuperscript{275} Desires have propositional content, whereas mere feelings do not.\textsuperscript{276} We are fallible regarding our desires, as we are not regarding our feelings.\textsuperscript{277} Even a valuation that ordinarily is deemed a matter of pure taste, such as a liking for a particular food, is not beyond the reach of reason.\textsuperscript{278} In politics and law, questions involving fundamental values often seem to be beyond rational resolution. They may be intractable as a matter of fact, but in principle they are not. Indeed, the persistence of a controversy over value is itself a sign that the participants believe they are advancing rational arguments. And the historical record of seemingly intractable controversies about which a rational consensus has been reached on questions of value suggests that they are right.\textsuperscript{279}


\textsuperscript{273} Gerald Gaus, \textit{Value and Justification: The Foundations of Liberal Theory} 136 (1990). For example, suppose a hurricane passes through an area, killing dozens, and I stand on the beach watching it go out to sea. I am crying, not because of the deaths of dozens of people, but because the hurricane is dissipating into an ordinary storm. My emotions are mistaken, and I could benefit from treatment in which I was challenged to articulate my feelings and defend them rationally.

\textsuperscript{274} See id. at 31-34 (noting that psychotherapy assumes that emotions are cognitive); Richard H. Pildes, \textit{Conceptions of Value in Legal Thought}, 90 Mich. L. Rev. 1520, 1546 (1992) (same).


\textsuperscript{276} Michael Smith, \textit{The Moral Problem} 104-111 (1994) (arguing that desires are not mere feelings, because they have propositional content). For example, we say that I feel a desire and that I feel a pain. But whereas we also say that "I feel a desire to N," we never say "I feel a pain to N."

\textsuperscript{277} Id. For example, I may realize, after years of studying the violin, that, whereas I thought I was motivated by a love for music, I was motivated all along by a sibling rivalry with my younger sister, a naturally gifted cellist. Cf. Hilary du Pré & Piers du Pré, \textit{Hilary and Jackie} (1998) (describing a similar relationship).

\textsuperscript{278} Wine connoisseurship is the most familiar example of a reasoned approach to a matter of taste. \textit{See generally Michael Bonadies, Sip by Sip: An Insider's Guide to Learning All About Wine} (1998). Another example is the fact that in Western societies four basic tastes are recognized (bitter, sour, sweet, salty), but in Asian societies, a fifth is recognized (in Japanese, \textit{umami}). John Willoughby, \textit{A Chemical Mystery That Excites the Taste Buds}, N.Y. Times, Jan. 14, 1998, at F3 ("It turns out to be largely a cultural or a linguistic problem," Professor [Michael] O'Mahony said, because \textit{umami} is a new word, it's a Japanese word and the taste basically comes from Japanese flavorings.").

\textsuperscript{279} The universal condemnation of slavery is an example of this phenomenon. Slavery still exists, \textit{see generally Kevin Bales, Disposable People: New Slavery in the Global Economy} (1999); Samuel Cotton, \textit{Silent Terror: A Journey into Contemporary African Slavery} (1999), but whereas it was once thought of as an honorable institution in many societies, \textit{see, e.g., Aristotle, Politics} 118 (Ernest Barker ed. & trans., 1946), it is now condemned, \textit{see Universal Declaration of Human Rights, G.A. Res. 217 III(A), U.N. GAOR, 3d Sess., at 71, U.N. Doc. A/810, art. 4 (1948), reprinted in Innovation and Inspiration: Fifty Years of the Universal Declaration of Human Rights} 217, 218 (Peter Baehr et
The misconception of value in noncognitive terms is a leading cause of consequentialism. That is, to misunderstand value in this way usually causes people to confine their conception of value to states of affairs that are desired as a matter of fact; to conceive of the rational response to value as the maximization of these states of affairs; and to think of practical reasoning as coextensive with instrumental, means-ends reasoning toward this maximization. In this way, the failure to appreciate the rational dimension of value leads to a misconception of rationality itself. It causes one to overlook the rationality involved in choosing the ends toward which instrumental reasoning is directed. Our ends are not merely states of affairs that we happen to desire; they are matters that we rationally deliberate upon, mostly with an eye toward the construction of a stable identity and the maintenance of a certain character.280 A serviceable (though not entirely satisfactory) name for this dimension of rationality is expressive rationality.281

Expressive rationality and its cognate, expressive value, in fact have been recognized widely in legal theory282—even by the law and economics
school, though that treatment of these ideas is fatally flawed. This flaw has to do with another misunderstood feature of value—one that bears on the problem of proportionality in punishment as well. Economic analysis of any variety supposes that value is transitive and therefore commensurable. But these assumptions have been undermined by recent experimentation that shows that value is highly context-dependent. Because value arises in an infinite variety of practical contexts, value cannot be transitive. This intransitivity means that these valuations are incommensurable. I cannot rank them on a single scale, though I can rationally


285. Cf. LEONARD J. SAVAGE, THE FOUNDATIONS OF STATISTICS 69 (2d ed. 1972) (“A function U that thus arithmetizes the relation of preference among acts will be called a utility. It will be shown that the multiplicity of utilities is not complicated, every utility being simply related to every other.”); James P. Spica, The Rationality of Normative Expectations, 24 J. Contemp. L. 259, 277 & n.74 (“Ramsey’s third axiom (concerning the transitivity of preference relations) is characteristically innocuous: ‘If option A is equivalent [i.e., indifferent] to option B and B to C then A to C.’” (citing F.P. RAMSEY, TRUTH AND PROBABILITY, in PHILOSOPHICAL PAPERS 52, 75 (D.H. Mellor ed., 1990))).


287. Transitivity means that if I do not prefer A to B, but I do prefer C to B, then I must prefer C to A. Intransitivity means that this relationship does not hold. For example, suppose that I am a novelist and a mountaineer. I may find it difficult to say whether I would rather win the Booker Prize or conquer Mount Everest. If so, it may seem that this is because these achievements are roughly equivalent in value. But the truth is more complex than this. Suppose, in addition to the fact that I do not prefer the Booker Prize to the conquest of Everest, that I have a decided preference for winning the Nobel Prize for Literature over winning the Booker Prize. It does not follow from this that I would rather win the Nobel Prize than conquer Everest. My valuations of the Booker Prize, the Nobel Prize, and the conquest of Everest are intransitive, and they are intransitive because they arise from the vastly disparate practical contexts of literature and mountaineering.

288. See RAZ, supra note 282, at 324–26 (describing incommensurability in terms of intransitivity).
compare them.\textsuperscript{289} I value my various activities not because they provide me with generic satisfactions that I can trade off against one another, but because of the things that each activity distinctively allows me to be and to express.\textsuperscript{290} My practical reasoning is not a matter of instrumentally maximizing a generic state of satisfaction; it is a matter of creating and sustaining a particular identity and character.

Just sentencing is a rational enterprise, but it is not one that is subject to calculation because proportionality in punishment is a question of expressive, not merely instrumental, rationality. Wrongdoing and fault are normative questions that cannot be reduced to facts in the form of discrete, quantifiable harms. The personal capacities and characteristics of the offender are relevant to his sentence, but their precise relevance is difficult to articulate, let alone to fit into a determinate sentencing grid. The various social ends of punishment—retribution, deterrence, public catharsis, education in social norms, incapacitation, rehabilitation—are considerations in a just sentence, as are matters more closely concerned with the individual offender, such as the impact of a sentence on his dependents and the possibility of mercy. But all of these competing considerations have widely divergent origins, and they are neither transitive nor commensurable. A proportionate punishment can be determined in only one way: by the exercise of mature judgment immersed in the human context of the particular case.

Determinate sentencing schemes, especially the Federal Sentencing Guidelines, fall well short of this minimum requirement of justice. In the words of Stith and Cabranes:

The new regime of bureaucratic sentencing inadvertently mocks the moral premises upon which the traditional ritual was based, while denying both sentencing judges and appellate judges the opportunity to develop a principled sentencing jurisprudence. In order for this regime of comprehensive, \textit{ex ante} sentencing rules to function effectively, the defendant must be reduced to an inanimate variable in an equation, the probation officer must operate as the "special master" of Guidelines facts, the sentencing judge must weight the crime according to the Sentencing Guidelines Commission's calculus, and the role of the courts of appeals is simply to police sentencing judges. Without principled foundation or application, the awesome power of the state to inflict

\begin{footnotesize}
\begin{enumerate}
\item \textsuperscript{290} The economist will argue that if I actually make a choice between winning the Nobel Prize and conquering Everest, then I have revealed a preference that can be assessed with economic tools. Moreover, I have revealed that the values are transitive and commensurable by making this choice. But these conclusions do not follow, because value has a tragic dimension. Any choice necessarily foregoes some value that does not show up in the analysis of revealed preferences. See Huigens, \textit{supra} note 284, at 545–47.
\end{enumerate}
\end{footnotesize}
suffering is wielded as an exercise in bureaucratic regularity, for which no one, ultimately, bears responsibility.²⁹¹

However, Stith and Cabranes go on to suggest that the alternative to the rigid bureaucracy of determinate sentencing schemes is a reliance on “highly subjective judgments”²⁹² and intuition.²⁹³ This is equally mistaken. To say that a matter cannot be reduced to ex ante prescriptions toward certain consequences is not to say that it falls beyond the realm of reason. If we can engage in meaningful reflection on the matter ex post and incorporate the results of those reflections into our future conduct and into the construction and maintenance of our identity and character, then this is a rational matter.²⁹⁴ Sentencing is such a matter.

Proportionate punishment has as much to do with the character of the punishing majority as it has to do with the character of the offender. In the individual case, the sentencing judge cannot commensurate and calculate the many considerations that bear on a proportionate sentence. However, none of this implies that she cannot rationally compare these considerations according to governing moral and legal norms.²⁹⁵ Ultimately, the question the sentencing judge faces is whether she can impose a given sentence consistently with her sense of her own personal and professional character. She will avoid harsh and arbitrary sentences in order to avoid crediting a view of herself as a harsh and arbitrary judge. At the level of legislation, the question of proportionality comes down to whether we, as individuals who take a large part of our self-conception from the society to which we belong, can bear to be part of a society that is arbitrary and unnecessarily cruel. Disproportionate punishments are barbaric. We may at times lack the political will to disavow such punishments, but the political question for each of us is whether we will remain part of a barbaric society.

C. MANDATORY-MINIMUM AND THREE-STRIKES SCHEMES REJECTED

Mandatory-minimum sentencing schemes are laws such as that at issue in McMillan: They identify some feature of the offender’s crime that is not an element of the offense and, on the basis of a judicial finding that this feature is present, require the sentencing court to impose a sentence above a certain floor that is not the statutory minimum for the offense. The holding of Apprendi does not apply to such laws, and this was quite deliberate on the Court’s part. The hate crimes law at issue in that case did not impose a mandatory-minimum

²⁹¹ STITH & CABRANES, supra note 24, at 103.
²⁹² Id. at 150.
²⁹³ See id. at 169.
²⁹⁴ Rosalind Hursthouse explains this kind of rationality in the context of the choice over abortion. See Rosalind Hursthouse, Virtue Theory and Abortion, in VIRTUE ETHICS 217, 224 (Roger Crisp & Michael Slote eds., 1997).
²⁹⁵ See Chang, supra note 289, at 1588–91.
sentence, and the Court's conservative approach to the question before it preserved the holding of _McMillan_. However, as Justice O'Connor pointed out in her dissent, and as Justice Thomas asserted in his concurrence, the logic of _Apprendi_ applies just as well to the imposition of mandatory minimums as it does to increases in the statutory maximum. In both kinds of determinate sentencing, the offender faces a more onerous sentence than he otherwise would, based on a sentencing factor that has not been proved to a jury beyond a reasonable doubt. If _Winship_ is evaded by increases in the statutory maximum sentence, then it is just as clearly evaded by legislative impositions of mandatory minimums. For this reason, the _Williams-McMillan_ faction on the Court doubted that the space that the _Apprendi_ opinion cleared for _McMillan_ would remain open for very long.

The _Apprendi_ dissenter's fear is justified because a principled treatment of mandatory-minimum sentencing schemes will not preserve _McMillan_ or mandatory-minimum sentencing schemes generally. Just as _Patterson_ cannot withstand a careful analysis, so _McMillan_ comes out the other way when its vague and ambiguous terminology is replaced with a more precise vocabulary. To commit an offense with a firearm is not merely an aggravating factor, it is a matter of fault, and this precludes, as a constitutional matter, our removing it from the jury's province. If Pennsylvania were to conclude that to have a particular intention in the commission of a crime indicated a degree of fault that was not reflected in existing law, then Pennsylvania ordinarily would include that mental state in the definition of a new offense or a new degree of an existing offense. Similarly, if Pennsylvania thought that the use of a firearm in the commission of an offense indicated a degree of fault that was not reflected in existing law, then Pennsylvania should have included the use of a firearm in the definition of a new offense or a new degree of an existing offense. In the half-century since _Williams_ initiated the debate over sentencing factors versus offense elements, no one has argued that the mens rea elements of an offense could be removed from the offense definition and treated as sentencing factors. This is why the Justices argued in _Apprendi_ over whether a racial purpose could be characterized as a motive instead of a mens rea. All the Justices agreed, implicitly, that if a racial purpose were a mens rea, then it could not be converted into a sentencing factor consistently with due process and the jury guarantee. But criminal fault is not exclusively a matter of mental states, either as a conceptual or as a practical matter. The reasons that lead us consistently to treat mental-states indicators of fault as a part of the offense definition should lead us to treat objective fault considerations in the same way.

297. _Id._ at 532–33 (O'Connor, J., dissenting).
298. _Id._ at 520–23 (Thomas, J., concurring).
299. _See id._ at 543–45 (O'Connor, J., dissenting).
300. _See Apprendi_, 530 U.S. at 490–93; _id._ at 553–54 (O'Connor, J., dissenting).
301. _See supra Part II.A.3._
These reasons are similar to those expressed in the Apprendi opinion and concurrences. Fault justifies the imposition of punishment, just as wrongdoing justifies punishment, because fault is an aspect of wrongdoing. If the justification provided by a violation of the elements of an offense ought to be determined by a jury, then the justification provided by fault also ought to be determined by a jury. The Apprendi Court framed this idea in terms of the features of positive law; that is, in terms of the facts that the legislature actually has designated as grounds for punishment or as grounds for a degree of punishment. 302 But in order to avoid the weaknesses of these arguments, we need to tie the argument to a theoretical account of the way in which fault justifies punishment. There are competing theoretical accounts of fault, of course, but one demonstrates better than the others do the intricate connections between wrongdoing and fault.

Neither consequentialist theories of punishment (usually called deterrence theories) nor deontological theories of punishment (usually called retributive theories) distinguish adequately between eligibility and fault. 303 Consequentialist theories recognize no particular connection between fault and wrongdoing, because they treat fault as an exogenous side constraint on punishment. 304 Deontological theories of punishment focus on the will as the seat of moral agency, in accordance with their Kantian origins. 305 The resulting conception of culpability stresses the idea of voluntariness as a necessary condition of just punishment and treats the intentional states that most often indicate culpability as various modes of voluntariness. 306 Because eligibility predominates over fault in this volitional conception of culpability, deontological theories also fail to give much of an account of fault as an aspect of wrongdoing. The strongest sign that consequentialist and deontological theories of punishment are inadequate is the inability of either theory to explain why criminal fault is not in fact confined to the presence or absence of mental states. 307

A virtue ethics 308 theory of punishment is superior in these re-

303. See supra note 188 (explaining why "deterrence theory of punishment" and "retributive theory of punishment" are misnomers).
304. See supra notes 191–92 and accompanying text.
306. Huijgens, supra note 32, at 1233–35; see Stephen J. Morse, The "Guilty Mind:" Mens Rea, in Handbook of Psychology and Law 207, 211 (Dorothy K. Kagehiro & William S. Laufer eds., 1992) (describing the "just deserts" position as holding that "there is no blameworthiness unless there is an appropriate mental state, such as intent or knowledge, that marks an actor's offending conduct as 'hers'").
307. See supra notes 210–22 and accompanying text.
308. No one should be put off by the word "virtue" in this context. It has nothing to do with the political and cultural agenda of some conservative political figures. See, e.g., WILLIAM BENNETT, THE BOOK OF VIRTUES (1990). Virtue ethics is the third major branch of philosophical ethics. See generally, e.g., THREE METHODS OF ETHICS: A DEBATE (Philip Pettit et al. eds., 1997). Just as consequentialism and deontological morality have served as the bases of theories of punishment, so can virtue ethics serve as the basis of a theory of punishment. Virtue ethics begins with Aristotle. See generally ARISTOTLE, THE
spects. The virtue ethics theory is a character-based theory of punishment. It holds that the inculcation and maintenance of the capacity for sound practical reasoning (virtue in its correct, technical sense) is a prominent objective of punishment that can justify punishment. The basic idea behind virtue ethics is that we govern ourselves at the level of our motivations and that we are responsible for the state of our motivations. The virtue ethics theory of punishment posits that the criminal law enforces this responsibility and plays a role—along with families, religious institutions, schools, and systems of social norms—in the inculcation of sound practical


309. The virtue ethics theory of punishment begins with this passage from The Nicomachean Ethics:

To obtain the right training for virtue from youth up is difficult, unless one has been brought up under the right laws. To live a life of self-control and tenacity is not pleasant for most people, especially for the young. Therefore, their upbringing and pursuits must be regulated by laws; for once they have become familiar, they will no longer be painful. But it is perhaps not enough that they receive the right upbringing and attention only in their youth. Since they must carry on these pursuits and cultivate them by habit when they have grown up, we probably need laws for this too, and for the whole of life in general.

Aristotle, supra note 308, at 295–96; see also id. at 34 ("Lawgivers make the citizens good by inculcating (good) habits in them, and this is the aim of every lawgiver."). Aristotle aside, the virtue ethics theory of punishment first appeared in the 1970s, see Edmund L. Pincoffs, Legal Responsibility and Moral Character, 19 Wayne L. Rev. 905, 918 (1973) (arguing that punishment represents a demand that one develop and exhibit certain character traits), and the theory has been given several full or partial treatments since then, see Peter Arenella, Character, Choice, and Moral Agency: The Relevance of Character to Our Moral Culpability Judgments, 7 Soc. Phil. & Pol’y 59 (1990), reprinted in Crime, Culpability, and Remedy 59, 61 (Ellen Frankel Paul et al. eds., 1990) (arguing that so-called rational choice theory is inadequate to describe the criminal law’s concern with character); Claire O. Finkelstein, Duress: A Philosophical Account of the Defense in Law, 37 Ariz. L. Rev. 251, 252–53 (1995) (relying on Aristotle’s conception of judgment to give an account of duress in terms of states of character); Huigens, supra note 189; Huigens, supra note 32; Kyron Huigens, Virtue and Inculpation, 108 Harv. L. Rev. 1423 (1995). The virtue ethics theory of punishment is one of several theories that conceptualizes criminal fault in terms of the actor’s desires, motivations, or preferences, instead of discrete mental states. See Guyora Binder, Meaning and Motive in the Law of Homicide, 3 Buff. Crim. L. Rev. 755, 766–69 (2000) (review essay) (describing the recent work of Huigens, Samuel Pillsbury, Dan Kahan, Alan Michaels, and Kenneth Simons in such a manner).


311. See Aristotel, supra note 308, at 170–73; Huigens, supra note 309, at 1452–56 (describing Aristotle’s identification of virtue with phronesis, or sound practical judgment).

The theory seeks to explain the principal features of just punishment and the criminal law, including fault, from this perspective.

The virtue ethics theory explains what it means to say that fault is an aspect of wrongdoing and, thus, the reason that fault must be determined by the jury. It also explains why fault is not limited to mental states and, thus, the reason that the features of the offense that have been designated as sentencing factors must be enacted and adjudicated as offense elements instead.

In a virtue ethics theory of punishment, fault is an inference, drawn in adjudication, to the effect that the person who has violated a prohibitory norm has done so in a way that calls the quality of his practical reasoning into question. The inference of fault in adjudication is a necessary complement to the legislation of offenses. The prohibitory norms that make up the criminal law are generalizations about proper and improper conduct, drawn from generations of shared experience. Action-guiding legislation takes as its models those judgments about the proper course of action in problematic practical situations that are widely regarded as sound judgments. The generalizations of these sound judgments are familiar to us, and almost uncontroversial: that property ought to be secure, that life and bodily integrity ought to be preserved, that individual autonomy ought to be respected, and so on. A criminal code is a relatively detailed set of positively enacted ethical generalizations of this kind. Their violation constitutes criminal wrongdoing.

Because they are generalizations of sound practical judgments in the relevant sphere of human conduct, each of the conduct rules of the criminal law presents an implicit demand that the accused should engage in sound practical reasoning in the circumstances in which such a crime may be committed. The requirement of fault entails a retrospective, adjudicative inquiry into this question. The question before the jury is whether the acts of the accused, in the particular circumstances of the alleged crime, displayed inadequate or flawed practical reasoning, including the deliberations on ends that have gone toward establishing and maintaining the attitudes and standing motivations of the accused. This determination by the jury turns on a comparison between the defendant's judgment in the relevant circumstances—as evidenced by the particular manner and circumstances of his wrongdoing—and a sound judgment in the relevant circumstances—as evidenced by the applicable conduct rules. This comparison between the defendant's judgment in the relevant circumstances and the sound judgment that is implicit in the applicable conduct rule leads to the inference of fault or no-fault. In this way, the justifying rationale of the conduct rule is brought to bear on the punishment of the individual offender.

Both intentional and nonintentional, or objective, fault are determined in this

---

313. See Huijgens, supra note 189, at 1016–34.
314. See id. at 1028–31.
315. See id. at 1024–25.
316. See id. at 1028–31.
way. In most cases, an intentional state regarding harm or a risk of harm denotes fault. However, this intentional state does not constitute fault. Fault consists of the way in which the accused has come to do wrong and the particular way in which he has done wrong. This fault may or may not be denoted by an intentional state on the occasion of action.\textsuperscript{317}

The inference of fault is a necessary step in the imposition of just punishment. Sound practical reasoning is context-dependent.\textsuperscript{318} That is, we cannot say what the right decision in any situation is, or was, unless we know not only what alternatives were available, but also the circumstances under which the choice was made. If part of the question before the jury is whether the defendant engaged in sound practical reasoning on the occasion of wrongdoing, then the conduct rule has to be returned to the level of specific, context-rich practical judgment from which it arose and in which the offender acted. The jury performs this specification of the conduct rule when it applies the law to the facts before it. This adjudicative specification of the norm complements the legislative generalization of the norm. If part of the justification of punishment is the inculcation and maintenance of sound practical judgment, then just punishment cannot be imposed unless and until this step is taken.\textsuperscript{319} We refrain from punishing those who do wrong without fault because the inculcation of sound practical judgment is one of the principal functions of punishment. It would be expressively irrational—even if it might be consequentially effective—to punish one whose actions do evince sound practical reasoning, even if these actions constitute a nominal violation of the prohibitory norm.\textsuperscript{320}

This intricate, complementary relationship between wrongdoing and fault suggests that fault always must be determined by the jury, just as wrongdoing is. This is too broad an inference because fault is a consideration in setting a proportionate sentence and the sentencing court also makes some fault determinations. I will explore this qualification and its implications in the next section. But the validity of mandatory-minimum sentencing schemes is unaffected by this qualification. Before moving on, then, I will stop to apply the principles that I have laid out so far to that question, and to the similar case of so-called “three strikes, you’re out” laws.

Enough probably has been said about mandatory-minimum sentencing schemes. \textit{McMillan} was incorrectly decided, and the limitation on the \textit{Apprendi} holding that is based on \textit{McMillan} cannot be maintained. It makes no difference that the law at issue in \textit{McMillan}, unlike the law at issue in \textit{Apprendi}, raised the available minimum punishment instead of the available maximum punishment. In either case, the law provided for the judicial determination of a matter of fault. If the Constitution requires the jury to determine wrongdoing beyond a

\begin{flushright}
317. See id.
318. See supra Part III.b.
320. See Huigens, supra note 32, at 1246–51.
\end{flushright}
reasonable doubt, then the Constitution requires the jury to determine fault beyond a reasonable doubt because fault is an aspect of wrongdoing. No one would suggest that the mens rea elements of an offense could be made into sentencing factors for purposes of either a mandatory-minimum sentencing or a sentencing in excess of the statutory maximum. Objective fault considerations that are employed for these purposes are objectionable for the same reasons and to the same degree.

By the same logic, *Almendarez-Torres* was incorrectly decided, and the limitation on the *Apprendi* holding that is based on that case cannot be maintained. The distinction between fault and eligibility does not correspond to a simple distinction between the offense and the offender. Not all personal characteristics are eligibility considerations. Indeed, the difference between fault and eligibility considerations is a functional one, so that a single fact may appear as one or the other kind of consideration, depending on context.\(^{321}\) In the context of a prosecution for illegal entry into the country after a previous deportation, the fact that an offender was deported because he had been convicted of violent offenses is a fault consideration. That circumstance makes the wrongdoing worse than it otherwise would be, and a matter of greater social concern, because the offender is a dangerous person. Criminal history aggravates the wrongdoing, and this aggravation has a mitigating converse. That the offender has little or no substantial criminal history appears as a mitigator in many capital sentencing statutes.\(^{322}\) More to the point, the circumstance of criminal history indicates flawed or defective practical judgment because the offender evidently cannot maintain a peaceful, law-abiding character. If the wrongdoing of crossing the border must be proved to the jury beyond a reasonable doubt, then the Constitution requires criminal history to be proved in the same way because it is a related fault determination that is a necessary step in the imposition of just punishment.

If criminal history or recidivism is a fault consideration in cases such as *Almendarez-Torres*, then the three-strikes sentencing schemes adopted by many states in the early 1990s are unconstitutional. Three-strikes schemes provide that an offender must be sentenced to life in prison without the possibility of parole if he is being sentenced for a felony of a specific class and if he also has two or more previous convictions for felonies of that class on his record.\(^{323}\) The most recent conviction is his "third strike." Were it not for its limitation

321. For example, the age of the offender can be a fault consideration in a statutory rape offense. The age of the offender, relative to that of the victim, makes the offense better or worse than it may otherwise be. *See, e.g.*, WASH. REV. CODE ANN. § 9A.44.073(1) (West 2000) (defining rape of a child in the first degree as intercourse between a victim under age twelve and a perpetrator twenty-four months older than the victim).


concerning recidivism, many three-strikes schemes would violate Apprendi because they authorize life sentences for offenses that otherwise carry a lower maximum penalty. But even if Almendarez-Torres were overruled, this would invalidate three-strikes schemes only in part, and in a way that, following Justice O’Connor’s advice, state legislatures would find easy to remedy. A more sophisticated analysis produces a much stronger argument against these schemes.

In three-strikes schemes, as in Almendarez-Torres, the previous convictions are used as fault considerations. The proponents of three-strikes schemes stress the benefits of permanently incapacitating repeat offenders, as well as these schemes’ supposed enhancement of the criminal law’s deterrent power. This consequentialism matches that of the Federal Sentencing Guidelines for crudeness of conception and disdain for the traditional normative architecture of the criminal law. The objective is to warehouse undesirables; the moral niceties of punishment be damned. Nevertheless, as in the case of the Guidelines, the basic features of genuine punishment are discernable. The essential charge against the repeat offender is not just that he is dangerous, but that he cannot maintain a peaceful, law-abiding character. His criminal history is a fault consideration because it tells us that he has flawed or defective practical judgment. If the wrongdoing of the current offense must be proved to the jury beyond a reasonable doubt, then the Constitution requires this aspect of his wrongdoing, the fault implicit in his criminal history, to be proved in the same way.

D. DISCRETIONARY JUDICIAL SENTENCING RETAINED

The intricate, complementary relationship between wrongdoing and fault that I described in the preceding section suggests that fault always must be determined by the jury, just as wrongdoing is. This is too broad an inference, because fault is a consideration in setting a proportionate punishment. In the course of sentencing, the judge makes her own determinations of fault and finds the facts on which these determinations turn. Fault’s particular relevance is to character. A sentencing court inquires into character even more broadly than the jury does, and in doing so the judge draws inferences about the quality of the offender’s practical judgment from the manner and circumstances of his wrongdoing—the quality of a person’s practical judgment being a prominent constituent of character.

On the logic of Williams and McMillan, the fact that the judge draws

324. See, e.g., id. § 9.94A.020(25) (listing seventeen offenses that can constitute a “third strike” but that would not otherwise carry a statutory maximum penalty of life imprisonment); see also supra notes 120–21 and accompanying text (describing the likelihood of the Court’s overruling Almendarez-Torres and thus eliminating the exception for recidivism in Apprendi’s holding).

325. See supra Part I.B.1.

inferences of fault implies that fault considerations can be enacted into positive law as sentencing factors instead of as offense elements. If the judge can find facts that determine fault in her discretion, then to constrain that discretion by the specification of sentencing factors and mandatory terms is a net constitutional gain. Conversely, if determinate and mandatory sentencing schemes are unconstitutional, then the institution of discretionary judicial sentencing cannot be maintained.

However, once we understand wrongdoing, fault, and proportionality, the Williams-McMillan argument breaks down. If fault considerations are enacted into positive law at all, then this is evidence that they pertain to wrongdoing and to the jury’s determination of whether punishment is justified in the case. In contrast, the fault determinations that the sentencing court makes involve considerations that are not positively enacted into rules, because they are relevant to a broader inquiry into character that bears on the determination of a proportionate punishment. This division of labor preserves the discretion of the sentencing court to make determinations of fault and to find the facts on which those determinations turn. But it also implies that the very idea of a positive determinate sentencing factor is contradictory. This point is at the heart of the solution to the Apprendi puzzle, so allow me to develop it a little further in the next six paragraphs.

Both the jury and the sentencing court determine fault as well as the facts on which these fault determinations turn. But these determinations of fault occur in the context of two different, lexically ordered, questions. The jury’s task is to determine whether any punishment is justified, and as a necessary step in making this threshold decision, the jury determines whether the offender’s practical judgment is flawed. Punishment is justified by wrongdoing, but only if the wrongdoing consists of more than a nominal violation of the prohibitory norm. Because the inculcation and maintenance of sound practical judgment are justifying purposes of punishment, we need to know that these purposes hold in each particular case. If the manner and circumstances of the offender’s wrongdoing do not indicate that flawed or defective practical judgment lies behind the act, then to punish in that case would be irrational and unjust. The jury is concerned with the defendant’s character only with respect to the quality of his practical judgment; it is concerned with his defective practical reasoning—that is, his fault—only insofar as it is an aspect of wrongdoing that bears on the justification of punishment in his case.

The sentencing court assumes that punishment is justified in the offender’s case. The judge is concerned with the defendant’s character as such, and only in part with the quality of his practical judgment. The defective practical reasoning that makes the offender’s conduct more than a nominal violation of the prohibitory norm brings him within the sentencing court’s authority and responsibility and plays a role in the determination of a proportionate punishment. But this

327. See supra Part I.A.
latter task involves a far broader set of concerns than the single fact that the offender's practical judgment is flawed. The determination of a proportionate punishment is a matter of expressive rationality, in which a number of incommensurable concerns are compared with an eye to the imposition of a punishment that accords with the character and self-conception of the punishing majority. The presence and the particulars of fault in the case are relevant, but not predominant, considerations in determining a proportionate punishment.

The enactment of a fault consideration into positive law is evidence that its role is to justify punishment and not to be pertinent to proportionality. Enactment into positive law tells us that the fact is of particular relevance to the prohibitory norm and to the justification of punishment in individual cases of the prohibitory norm's violation. By providing for the adjudication of a particular fault consideration, the legislature has ensured that the wrongdoing in question will be clearly condemned in all important respects, that the offender and others will be deterred in this particular regard, that the community will experience catharsis regarding this particular aspect of the wrongdoing, and so on. Furthermore, the very fact that the particular fault consideration has been tied to the wrongdoing by and within a structure of rules indicates an apparent recognition on the part of legislators that the consideration bears on the justification of punishment in the individual case. The determination of guilt or innocence always has been hemmed in by rules because it is a threshold question; it is the point at which a citizen's standing in society will or will not undergo a radical transformation. This is not the case with sentencing. If the principle of legality requires a matter such as the quantity of drugs or the weapon involved in an offense to be committed to a structure of rules, then this is strong evidence that the matter is a fault consideration that bears on the justification of punishment.

The notion of a positive determinate sentencing factor is, from this perspective, unsustainable. If the fault consideration is so salient to the wrongdoing that it must be expressly condemned, and so salient to the threshold justification of punishment that it must be framed in rules, then it is illogical to express that consideration in rules that have nothing to do with either the determination of wrongdoing or the threshold justification of punishment. If the matter in question is a fault consideration that bears on the justification of punishment, then the structure of rules to which it belongs is that of offense definitions, and the proper method of proof is that which is required for the proof of offenses. In other words, the very idea of a positive determinate sentencing factor is a contradiction: It is a fault consideration that, on the evidence of its being positively enacted, is a matter for the jury to determine, that nevertheless is not a matter for the jury to determine. Positive fault considerations, therefore, can be neither determined by the sentencing court, nor enacted as sentencing factors.

The fault considerations that the sentencing court properly determines neces-

328. See supra notes 285–95 and accompanying text (distinguishing incommensurability from incomparability and describing sentencing as an expressively rational matter).
sarily are those that have not been enacted into positive law. The number and
gradation of facts that may bear on fault are virtually infinite, and not all of
them need be or can be positively enacted. In any given case, many consid­
erations that are pertinent to the quality of the offender’s practical judgment can
be found in the interstices of the offense definition. Many of these aspects of the
offense are simply too fine-grained to be captured in rules. These interstitial
fault considerations are determined by the jury when it draws the inference of
fault; that is, when the jury specifies the prohibitory norm and brings it down
from the general level of the offense definition to the level of particulars at
which the prohibition originated, at which the defendant made his decisions,
and at which his practical reasoning ought to be assessed retrospectively.
Interstitial fault considerations also are determined by the sentencing court
when it performs its distinctive function. In arriving at a proportionate sentence,
the court considers the positive fault that the jury has determined in the course
of rendering a conviction. But it also goes beyond this delimited determination
of fault and considers the manner and circumstances of the particular offense
and the more subtle signs and gradations of fault to be found in the interstices of
the positive law.

The distinction between positive and interstitial fault considerations preserves
discretionary judicial sentencing, but it breaks the Williams-McMillan syllogism
and reinforces the conclusion that determinate sentencing, mandatory-min­
imum, and three-strikes sentencing schemes are invalid. The sentencing court
can hear and determine the existence of eligibility-based mitigators that pertain
to the capabilities and other characteristics of the offender. The court can hear
and determine facts concerning the offense that bear on interstitial fault consid­
erations. But the litigation of positive fault considerations is barred at sentencing.
The Federal Sentencing Guidelines are invalid because real-offense sentencing
depends on the litigation of fault considerations that have been positively
enacted in the Guidelines, most of which should not have been enacted into
positive law in any form. Mandatory-minimum sentencing schemes, such as
that involved in McMillan, are invalid for the same reason. If the sentencing
factor is so salient to the offense that the legislature feels it necessary to enact it
into positive law, then it is a fault consideration that bears on wrongdoing and
on the threshold determination that punishment is justified in the individual
case. If such a consideration is enacted into law at all, it must be as part of an
offense that is to be proved to a jury beyond a reasonable doubt, and to which
other constitutional trial rights apply. Otherwise, it is an interstitial fault consid­
eration that either the jury or the sentencing court or both may determine, each
within the context of its own institutional responsibilities.

329. See infra notes 335–38 and accompanying text (defining “fine-grained” in this context).
330. See supra Part II.B; see also supra note 256 (giving examples of positive and interstitial fault
considerations).
331. I include in this description not only the Guidelines themselves, but also federal offense statutes
The case of three-strikes schemes is perhaps the most interesting of the three kinds of sentencing schemes from this perspective. The Almendarez-Torres Court pointed out that criminal history is as traditional a sentencing factor as one can imagine.\(^{332}\) This is both true and untrue. In the sentencing of most offenders, criminal history is an interstitial fault consideration. It is not a consideration that has been enacted as part of the offense for which the offender is being sentenced, but it is a fault consideration: It indicates the relative ability of the offender to maintain a law-abiding character and the proper level of social concern over his actions. The sentencing court naturally has the power to hear and determine the relevance of this consideration in the course of setting a proportionate sentence. However, when an offender’s criminal history reaches a certain level, the legislature may conclude that this aspect of the wrongdoing is particularly salient. The final act of wrongdoing in a series of criminal acts ought to be particularly condemned, so that others will be deterred in this particular regard, so that the community will experience catharsis regarding this particular aspect of the wrongdoing, and so on. If the legislature reaches this conclusion, then criminal history is no longer merely pertinent to proportionality; it is primarily relevant to wrongdoing and to the threshold determination that punishment is justified in the individual case. This is a positive fault consideration, and the Constitution requires that it be pled and proved as an element of an offense.

E. FOUR OBJECTIONS, BRIEFLY CONSIDERED

The foregoing argument can be further clarified and defended with brief responses to four objections.

The first objection is that my distinction between positive fault considerations and interstitial fault considerations turns on the state of positive law,\(^{333}\) which means that my argument will be vulnerable to Justice O’Connor’s reductio ad absurdum, in the same way that Justice Stevens’s and Justice Thomas’s arguments are vulnerable.\(^{334}\) My argument is not vulnerable in this way, because positive law plays only an evidentiary, not a determinative, role in a normative analysis. Justices Stevens and Thomas retain the category of “sentencing factor,” and draw only a positive law distinction between sentencing factors and offense elements. Given that the legislature controls positive law, the legislature retains the power to convert offense elements to sentencing elements at will, provided only, as Justice O’Connor points out, that the legislature goes about this in the right way. My argument avoids Justice O’Connor’s reductio because it draws a normative line between the adjudication of offenses and sentencing.

---

333. My argument, again, is that the legislature’s enactment of a fault consideration into positive law is evidence that it is a positive fault consideration; if it is a positive fault consideration, then it must be enacted as part of an offense definition if enacted into positive law at all. See supra Part III.D.
If a legislature enacts a fault consideration into law as a sentencing factor, then it has made a mistake. Its enacting the matter into positive law at all is evidence that the consideration is particularly salient to the wrongdoing at issue; but if this is so, then this is all the more reason to adjudicate the matter in conjunction with wrongdoing, and not separately. On my argument, the enactment of a sentencing factor implies that the legislature ought to enact a new offense or offense definition instead; whereas on the argument of Justices Stevens and Thomas, the legislature’s enactment of a sentencing factor is, if done cleverly, conclusive.

There is another way to describe the normative line between positive fault and interstitial fault, and it draws a second useful objection. It may be that a fault consideration should not be enacted into positive law at all, either as a sentencing factor or as an offense element. Some features of wrongdoing are so fine-grained that to enact them into rules at all will distort their significance and function as indicators of fault. In other words, there is an upper limit to the complexity that a system of rules can accommodate, and the boundary between positive and interstitial fault considerations tracks this upper limit. The objection to this argument is simply that the notion of a fault consideration’s being too fine-grained is unclear. It seems to be no more useful than the McMillan “wags the dog” standard. But it is difficult both to articulate a test for fine-grainedness and to show that determinate sentencing schemes such as the Federal Sentencing Guidelines cross the line so defined. We know that a matter is too fine-grained for rules when it is written into a rule and one of two things happens. First, when elaborations of the rule are permitted, elaborations proliferate and begin to conflict. Second, when elaborations of the rule are not permitted, the authorities who apply the rules begin to apply them in result-oriented ways. Both of these effects are discernible in the case of the Guidelines. The Guidelines permit a limited role for common-law development, and within these boundaries, minute and often conflicting elaborations of rules pertaining to fault have proliferated. On the whole, however, the Guidelines

---

335. In an abundance of caution, I will note that, for the reasons given in the preceding paragraph, the line between positive and interstitial fault is not the same line as the one that the “wag the dog” standard governs, which is the line between sentencing factors and offense elements.

336. For example, scores and sentences under the Guidelines are affected significantly by whether an offender is “substantially less culpable than the average participant.” U.S. SENTENCING GUIDELINES MANUAL § 3B1.2 (1995) [hereinafter U.S.S.G.]. Courts have differed over whether this involves a comparison to both the other participants in the offense and the “average participant” in such a crime, see United States v. Snoddy, 139 F.3d 1224, 1228 (8th Cir. 1998); United States v. Daughtrey, 874 F.2d 213, 216 (4th Cir. 1989); either the other participants in the offense or the “average participant” in such a crime, see United States v. Andrus, 925 F.2d 335, 338 (9th Cir. 1991); the other participants in the offense but not the “average participant” in such a crime, see United States v. Andrus, 925 F.2d 335, 338 (9th Cir. 1991); or the “average participant” in such a crime but not the other participants in the offense, see United States v. Zaccardi, 924 F.2d 201, 203 (11th Cir. 1991). See Jefri Wood, Federal Judicial Center, Guideline Sentencing: An Outline of Appellate Case Law on Selected Issues 135–36 (2000). Similarly, it is far from clear when an offender has engaged in a “crime of violence” in a “controlled substance offense” as opposed to a “violent felony” in a “serious drug offense.” Compare
have been interpreted so as to bind the lower federal courts into a rigid framework. And, on the whole, one finds a strong tendency toward result-oriented applications of the Guidelines. When rules follow rather than lead an analysis in this way then rules have overreached their useful role, and the matter is too fine-grained for rules.

A third objection denies that sentencing factors are matters of fault at all. The objection is that the enactment of an aspect of wrongdoing into positive law does not necessarily mean that the legislature wishes to condemn or to deter the wrongdoing in this particular respect. The legislature may have some other reason to enact the consideration into positive law: notably, a good faith effort to avoid discrepancies and inequalities in sentencing. If this is the case, the argument goes, then the consideration is not a matter of fault, nor must it be adjudicated at trial, by the jury, in conjunction with wrongdoing. But this conclusion does not follow. Any aspect of wrongdoing that is singled out for purposes of determining the relative severity of the wrongdoing is by definition a matter of fault. If it is singled out by designation as a sentencing factor, then this is because a legislature has chosen to provide for the adjudication of a matter of fault separately from the adjudication of the rest of the wrongdoing. That the legislature has made such a choice does not imply that the matter is not one of fault, that the legislature has acted wisely, or that the separate adjudication of fault and wrongdoing is theoretically or morally defensible.

Finally, one may object that, if determinate sentencing schemes are abolished, then we will be deprived of the means to address problems of discrepancy or inequality in sentencing. This is not true. To the extent that it is feasible to address problems of discrepancy and inequality through the enactment of detailed sentencing factors, it is feasible to address these problems through the enactment of detailed offense definitions. A conviction for such an offense may authorize a sentence within only a very narrow range, with the specific sentence to be determined by a sentencing judge exercising her discretion. If truth in sentencing is an important consideration, then the offender could be required to serve this sentence without parole. Certainly there is an upper limit on the complexity in an offense definition that a jury can be expected to handle. But if

a consideration is too fine-grained to serve as an offense element, then it is too fine-grained to serve as a sentencing factor. Our experience with Guidelines sentencing hardly has demonstrated that judges are more adept at overcoming the natural limits of rules than juries are. Defenders of the Guidelines frequently tout their system’s superiority to a pure offense-based system, under which the offense of conviction immediately determines the sentence.\textsuperscript{339} They may claim that the system that I have just suggested is offense-based. It is offense-based, but it does not entirely eliminate discretionary judicial sentencing. Like the Guidelines themselves, it strikes a balance between determinacy and discretion. The difference is that it does not sacrifice constitutional trial rights in the process.

F. THE NONRETROACTIVITY OF THE SOLUTION

The principal objection to the analysis presented above may be a practical one: The Federal Sentencing Guidelines and state mandatory-minimum and three-strikes schemes have been in force for many years, and hundreds of thousands of offenders have been sentenced under these systems.\textsuperscript{340} If the Supreme Court were to strike down these schemes on the grounds that I have laid out, then it is arguable that this holding should be applied retroactively and that every mandatory sentence ever handed down should be reviewed and possibly overturned on habeas corpus. If this were the case, then it would count as a good reason not to adopt my solution to the \textit{Apprendi} puzzle.\textsuperscript{341} But it is not the case: My solution to the puzzle is nonretroactive.

Ordinarily, a newly created or newly adopted rule of constitutional law cannot be applied retroactively to a judgment and sentence that has become final.\textsuperscript{342} Two exceptions to this rule of nonretroactivity create the difficulty for my argument. First, a new rule of constitutional law can be invoked by a


\textsuperscript{340}. See supra notes 13, 20.

\textsuperscript{341}. The reasons for the rule of nonretroactivity include judicial efficiency, the psychological value of finality, and the value of comity between state and federal courts. See Teague v. Lane, 489 U.S. 288, 308 (1989) (plurality opinion); Jordan Steiker, Innocence and Federal Habeas, 41 UCLA L. REV. 303, 361 (1993). These reasons apply with particular force to a retroactive abolition of state and federal mandatory sentencing. The financial cost of relitigating these cases would be daunting for the richest society. Even if resentencing at roughly the previous level is a far more likely outcome than release, the psychological cost of unsettling these convictions would fall heavily on victims and offenders alike because members of both groups would overestimate the likelihood of release. The impact of abolishing mandatory sentencing would fall mostly on the states because of the sheer number of state mandatory sentencing cases, which vastly outnumber Federal Sentencing Guidelines cases. See supra note 20. Because it would be the task of the federal courts to undo these otherwise final sentences, federalism would be affected in a way that has been of particular concern.

\textsuperscript{342}. Only those defendants who have not been convicted and those offenders whose appeals have not been decided can invoke the new rule. 2 JAMES S. LIEBMAN & RANDY HERTZ, FEDERAL HABEAS CORPUS PRACTICE AND PROCEDURE § 25.1, at 934 (3d ed. 1998). A new rule of constitutional law is not a valid ground for a habeas corpus petition simply because the number of final judgments affected by the new rule is likely to outnumber the nonfinal judgments affected by a large factor.
prisoner whose conviction is final, if the rule "places 'certain kinds of primary, private individual conduct beyond the power of the criminal law-making authority to proscribe.'" This exception has been interpreted to mean that new rules of substantive criminal law always apply retroactively; the nonretroactivity doctrine applies only to new rules of criminal procedure. For example, the exception has been applied to the question of an Eighth Amendment ban on the death penalty for the mentally impaired. If new substantive criminal law rules on proportionate punishment apply retroactively, then my solution to the Apprendi puzzle would seem to apply retroactively. Second, a new rule also applies retroactively if it is a fundamental rule that ensures accuracy in the outcome of the trial. This retroactivity principle applies to both trial and sentencing. It has been applied to new rules concerning the due process requirement of proof beyond a reasonable doubt and to new rules that protect the jury right. Because my solution to the Apprendi puzzle draws on the fundamental principles of just punishment in order to preserve both of these rights and to adjust the basic relationship between trial and sentencing, it would seem to apply retroactively.

However, my solution to the Apprendi puzzle does not apply retroactively, and in this respect it is significantly different from the arguments of the Apprendi majority. The exceptions to the rule of nonretroactivity for new rules of law are not really exceptions to that rule at all. The point of both of the exceptions is to identify new rules of law that are in fact old rules, by virtue of their status as fundamental law. They are rules "that never before were announced precisely because they are so fundamentally the law that no one previously thought of violating them." The arguments of the Apprendi opin-

344. See 2 Liebman & Hertz, supra note 342, § 25.7, at 1021.
346. See Teague, 498 U.S. at 313 (plurality opinion); id. at 320–21 (Stevens, J., concurring in the judgment).
347. See 2 Liebman & Hertz, supra note 342, § 25.7, at 1022 n.22.
348. See, e.g., Humphrey v. Cain, 138 F.3d 552, 553 (5th Cir. 1998) (en banc) (applying retroactively the rule that jury instructions diluting the reasonable doubt standard violate due process (citing Cage v. Louisiana, 498 U.S. 39 (1990) (per curiam))); see also In re V. v. City of New York, 407 U.S. 203 (1972) (holding In re Winship to be retroactive).
350. 2 Liebman & Hertz, supra note 342, § 25.7, at 1027. Liebman and Hertz apply this description to rules that fall under the second of the two exceptions to nonretroactivity, but it applies equally well to the first. In pre-Teague cases, the Supreme Court has taken the view that, in situations in which the first Teague exception applies, the statute that purported to outlaw the conduct in question was never of any force or effect, so that no one can be punished for violating it. See United States v. Addonizio, 442 U.S. 178, 186–87 (1979); Davis v. United States, 417 U.S. 333, 347 (1974). In other words, the apparent new rule is so fundamentally the law that no crime was ever committed under a statute that the apparent new rule strikes down. See Chambers v. United States, 22 F.3d 939, 942–43 (9th Cir. 1994) (applying
ion and concurrences are framed so as to bring the case within this description. The Justices in the *Apprendi* majority plainly regard the notion of a sentencing factor to be a recent innovation that contravenes a well-settled understanding to the effect that all facts necessary to justify or increase punishment must be proved beyond a reasonable doubt to the jury.  

If, as the dissenters insist, there are few if any explicit statements of this principle in the case law, then this is due to the fact that no one thought to challenge the principle until recently, with the result that no court had occasion to articulate it. On this reading of *Apprendi*, the case clearly would have retroactive application, as would the extension of *Apprendi* advocated in Justice Thomas's concurrence.

My argument is different from the arguments of the *Apprendi* Court in this respect because it does not seek to restore a constitutional consensus. Instead, it seeks to overturn a consensus. The *Apprendi* dissenters were right, perhaps more right than they knew, when they asserted that *Apprendi* upset a settled understanding.  

The settled understanding extends farther back than *Patterson* and reaches more broadly than *McMillan*'s approval of sentence enhancements. Those cases, like *Williams*, are part of the consequentialist consensus in the theory of punishment, a consensus that held throughout most of the twentieth century and that both determinate and indeterminate sentencing philosophies exemplify. The mandatory sentencing schemes of the states and Congress—and the decisions of the Supreme Court that encouraged those schemes—are the products of this consequentialist consensus. They are the results of good faith efforts over more than half a century to grapple with the moral and practical complexities of crime and punishment. I have argued, here and elsewhere, that the consequentialist theory of punishment is deeply mistaken, but whether mistaken or not, it set the course of constitutional law and public policy in this area throughout most of the last century.

That being the case, the retroactive annulment of the flawed results of that constitutional consensus is not constitutionally required. The exceptions to the rule of nonretroactivity are meant to identify and apply retroactively the unusual constitutional rule on which long-standing consensus exists and that has been articulated as a new rule only because it has never before been challenged. My solution to the *Apprendi* puzzle advocates a sharp departure from a consensus view. The strong version of *Apprendi*'s holding that I advocate requires a fundamental shift in the constitutional analysis of the criminal law—a shift not only toward a more explicit reliance on the theory of punishment, but toward reliance on a different theory from the one that is implicit in the Court's decisions up until now. Because it would overthrow the consequentialist

---

*Davis and Addonizio* to a case falling under *Teague*'s first exception), vacated on other grounds, 47 F.3d 1015 (9th Cir. 1995).

351. See supra note 108.


353. See supra Part II.A.1.

354. See Huigens, supra note 189, passim.
consensus, my solution to the Apprendi puzzle should be treated as a new rule and applied only prospectively.

CONCLUSION

The Supreme Court is understandably averse to adopting a particular theory of punishment for purposes of constitutional analysis. Its institutional integrity seems to require it to avoid pronouncements on contested moral matters, and the theory of punishment is a branch of moral philosophy on which little consensus has been in evidence. However, the Apprendi puzzle is one of several points at which constitutional law and substantive criminal law intersect, and the Court has no hope of deciding the constitutional question definitively if it ignores the assistance that the theory of punishment can provide. The Court’s institutional integrity also is threatened by the use of ill-defined and ambiguous conceptual tools and by the inconsistent decisionmaking that the Court has displayed between Williams and Apprendi. Ultimately, the Court’s integrity rests not on the avoidance of controversy, but on consistent, principled decisionmaking.

The Court should abandon once and for all the terminology that it has used up until now for the constitutional analysis of substantive criminal law questions—terms such as culpability, aggravation, mitigation, and affirmative defense. It should employ the more sophisticated terminology of a coherent theory of punishment and speak in terms of wrongdoing, fault, eligibility, and proportionality. It could then, as I have argued above, recognize the common threads that run through seemingly different statutory schemes, such as the homicide laws of Maine and New York. More importantly, the Court could look beyond the state of positive law, find its footing in theory and principle, and provide defensible answers to questions such as the constitutional definition of a criminal offense.

The Court has avoided taking such a step up until now, and the inevitable results have been confusion and the erosion of constitutional standards. The Court’s avoidance of controversy leads only to legislative overreaching and to crises such as the “erosion of the jury’s function to a point against which a line must necessarily be drawn.”\textsuperscript{355} The only way to resolve such crises and to avoid them in the future is to draw defensible lines. As I have argued above, the Court should not hesitate to declare determinate sentencing, mandatory-minimum sentencing, and three-strikes schemes invalid. These steps are the logical outcome of a theoretically sound and principled approach to the question of the constitutional definition of a criminal offense. If the cost of drawing such lines is enormous, then the Court’s timidity is as much to blame as the legislatures’ opportunism. The only way out of this expensive mess is to take a principled approach that will provide consistent guidance to lawmakers and minimize further damage to the Court’s integrity and prestige.

\textsuperscript{355} Jones v. United States, 526 U.S. 227, 244 (1999).