Scholarship Repository
University of Minnesota Law School

Articles | Faculty Scholarship

2005

Punishment Purposes

Richard Frase
University of Minnesota Law School, frase001@umn.edu

Follow this and additional works at: https://scholarship.law.umn.edu/faculty_articles

Part of the Law Commons

Recommended Citation
Richard Frase, *Punishment Purposes*, 58 STAN. L. REV. 67 (2005), *available at* https://scholarship.law.umn.edu/faculty_articles/543.

This Article is brought to you for free and open access by the University of Minnesota Law School. It has been accepted for inclusion in the Faculty Scholarship collection by an authorized administrator of the Scholarship Repository. For more information, please contact lenzx009@umn.edu.

PUNISHMENT PURPOSES

Richard S. Frase*

INTRODUCTION 67
I. OVERVIEW OF CONTEMPORARY SENTENCING PURPOSES AND LIMITATIONS 69
A. *Utilitarian Purposes and Limitations* 70
B. *Nonutilitarian Purposes and Limitations* 73
C. *Utilitarian Proportionality and Uniformity* 74
D. *Conflicts Within and Across Punishment Principles* 75
II. RECONCILING CONFLICTING PUNISHMENT PRINCIPLES: LIMITING RETRIBUTIVISM 76
A. *The Limits of the Criminal Law as an Instrument of Crime Control* 79
B. *Illustrative Cases* 80
CONCLUSION 81

INTRODUCTION

The reform goal of promoting reasonable consistency and reducing disparity in sentencing is meaningless without a frame of reference—consistency or disparity relative to what underlying principles?[1] In order to decide that two offenders are similarly situated and thus should receive similar sentences (or that they are dissimilar and should receive different sentences) we must first define the relevant sentencing factors (the offense and offender characteristics that judges should consider in determining appropriate sentences) and the weight to be given to each of these factors. The choice and weighting of sentencing factors depends, in turn, on the punishment purposes which the sentence is supposed to serve.

---

* Benjamin N. Berger Professor of Criminal Law, University of Minnesota. I would like to thank Kevin Reitz and Michael Tonry for their helpful comments on an earlier version of this Article.

1. *See* Anthony N. Doob, *The United States Sentencing Commission Guidelines: If You Don't Know Where You Are Going, You Might Not Get There*, *in* THE POLITICS OF SENTENCING REFORM 199, 233-35 (Chris Clarkson & Rod Morgan eds., 1995); Paul J. Hofer & Mark H. Allenbaugh, *The Reason Behind the Rules: Finding and Using the Philosophy of the Federal Sentencing Guidelines*, 40 AM. CRIM. L. REV. 19, 36-37 (2003); Peter Westen, *The Empty Idea of Equality*, 95 HARV. L. REV. 537 (1982).

Sentences can serve many purposes, and these purposes are often in conflict. Some of the most difficult conflicts are between proportionality principles, on the one hand, and case-specific crime-control or restorative-justice purposes, on the other. Proportionality serves both retributive (just deserts) and practical (utilitarian) sentencing purposes. Under a retributive theory, sanctions should be scaled in proportion to each offender's blameworthiness, and equally culpable offenders should receive equally severe sanctions. Sentencing proportionality and uniformity also have practical benefits, such as reinforcing public views of relative crime seriousness and maintaining public respect for criminal laws and the criminal justice system.

But realizing the goal of efficiently preventing future crime sometimes requires unequal or disproportional treatment. For example, if two first-time offenders commit the same crime but one has genuine feelings of remorse, strong family ties, and other indications of amenability to supervision and low risk of reoffending, putting that offender on probation and sending his much riskier counterpart to prison saves scarce correctional resources while still promoting public safety. But doing so produces disparate sentences for equally culpable offenders and undercuts the practical values served by uniformity and proportionality.

The best solution to conflicts such as this is not to adopt a narrow punishment theory (e.g., one based solely on retributive or risk-management goals), but rather to design a hybrid sentencing system that gives appropriate scope to all legitimate sentencing purposes. The hybrid approach adopted by most state guidelines systems is a version of the theory of limiting retributivism. Under this approach, principles of uniformity and proportionality relative to crime seriousness and offender desert set upper and lower limits on sentencing severity. Within the range defined by these limits, other principles provide the necessary fine-tuning of the sentence imposed in a particular case. These other principles include not only traditional crime-control purposes such as deterrence, incapacitation, and rehabilitation, but also a concept known as parsimony—a preference for the least severe alternative that will achieve the purposes of the sentence. The parsimony principle recognizes that severe penalties are expensive and usually harmful to offenders and that the crime-control benefits of such penalties are uncertain and often quite limited. Severe penalties should therefore be used as sparingly as possible.

Part I of this Article provides a brief overview of contemporary sentencing purposes and discusses some of the many ways in which these purposes conflict with each other. Part II describes the theory of limiting retributivism and shows the widespread support that exists for this theory among sentencing philosophers, in model codes and standards, and in contemporary sentencing law and practice, particularly in state guidelines systems. Several illustrative cases are used to show how limiting retributivism reconciles conflicting punishment purposes. The Conclusion emphasizes the importance of clearly defining and reconciling such purposes, both in designing a sentencing system

and in deciding what sanctions are appropriate in a given case. It also offers suggestions for how courts and Congress could reinforce the limiting retributive elements which are implicit in the Federal Guidelines' enabling legislation.

I. OVERVIEW OF CONTEMPORARY SENTENCING PURPOSES AND LIMITATIONS

Unless criminal penalties serve valid purposes, they impose useless costs and hardship. Even when the purposes are valid, punishment may be limited by moral values or practical concerns. Punishment purposes are positive, justifying principles; punishment limitations are negative, restraining principles. It is important to clearly define these positive and negative principles and to guide judges in their application.

In an indeterminate sentencing system,[2] the legislature usually only provides a very general, all-inclusive list of sentencing purposes and limitations, giving little or no guidance to system actors (prosecutors, trial and appellate judges, parole and other correctional officials) as to how these principles should be defined and applied in specific cases. Some state guidelines commissions have provided such definitions and guidance.[3] The United States Sentencing Commission chose not to,[4] and federal trial and appellate judges have not done so either.[5] As a result, federal sentencing is no more coherent and principled today than it was before the Federal Sentencing Guidelines were adopted. The overall structure and specific provisions of the existing Guidelines lack any clear underlying or principled basis, and judges called upon to interpret the Guidelines have thus been left to invoke whatever purposes and limitations they prefer.

Why should violators of criminal laws be punished, and what principles should be recognized to limit the type and degree of punishment?[6] Punishment purposes and limitations are traditionally grouped in two categories: utilitarian and nonutilitarian. Utilitarian purposes and limitations seek to achieve beneficial effects (or a net benefit) and, in particular, lower frequency and/or seriousness of future criminal acts by this offender or others. Nonutilitarian

2. For further discussion of such systems, see *infra* text accompanying note 8.

3. *See, e.g.*, MINN. SENTENCING GUIDELINES COMM'N, REPORT TO THE LEGISLATURE 8-13, at 35 (1980).

4. U.S. SENTENCING GUIDELINES MANUAL § 1A1.1 (2004); Doob, *supra* note 1, at 246; see *infra* text accompanying notes 49-55 (discussing support for limiting retributivism in guidelines statute). *But see* Hofer & Allenbaugh, *supra* note 1, at 51-52 (claiming that modified just deserts theory, a form of limiting retributivism, is implicit in and provides the most coherent explanation for the Federal Guidelines).

5. *See* Hofer & Allenbaugh, *supra* note 1 *passim* (finding scant theory in guidelines case law).

6. *See generally* Kent Greenawalt, *Punishment*, *in* ENCYCLOPEDIA OF CRIME AND JUSTICE 1282-94 (Joshua Dressler ed., 2d ed. 2002).

punishment purposes and limitations embody principles of justice and fairness which are viewed as ends in themselves, without regard to whether they produce any particular social or individual benefit.

While these two major categories of punishment purposes and limitations are conceptually separate, in practice they are often closely linked. As discussed more fully below, although proportionality and uniformity in sentencing are often claimed to be ends in themselves, each also has important utilitarian value. Sentences which depart greatly from widely held views of proportionality and uniformity may fail to prevent future crime and may even encourage crime by undermining the public's ability to gauge the relative seriousness of crimes and by lessening respect for criminal laws and the criminal justice system.

A. *Utilitarian Purposes and Limitations*

The most widely adopted utilitarian sentencing principles focus on using criminal penalties to prevent or lessen the seriousness of future criminal acts by the offender being sentenced and/or by other, would-be offenders. Criminal penalties have the potential to achieve these crime-control effects through at least five causal mechanisms: *rehabilitation*, *incapacitation*, *specific deterrence*, *general deterrence*, and *denunciation*. Each of these methods depends on certain critical assumptions and conditions for its effectiveness.

The first three methods seek to prevent future crimes by this particular offender; these methods thus assume both that certain defendants have an elevated risk of reoffending (justifying special measures addressed toward them specifically) and that these offenders and their degree of elevated risk can be identified in advance. Rehabilitation further assumes that the offender has identifiable and treatable problems which cause him to commit crimes; this approach seeks to reduce the offender's future criminality by addressing those causes through education and treatment in prison or in a nonprison program. Incapacitation prevents crime by imprisoning high-risk offenders, thus physically restraining them from committing further crimes against the public. This crime-control method assumes not only that such offenders can be reliably identified but also that they are not made worse by imprisonment, and that—while in custody—they are not replaced by other offenders.[7] Specific deterrence (also known as special or individual deterrence) seeks to discourage the defendant from committing further crimes by instilling fear of receiving the same or a more severe penalty in the future.

Rehabilitation and incapacitation are the most important sentencing purposes underlying traditional indeterminate sentencing systems.[8] Judges are

7. *See* JAMES Q. WILSON, THINKING ABOUT CRIME 146 (rev. ed. 1983).

8. *See* MICHAEL TONRY, SENTENCING MATTERS 4, 6 (1996); Steven L. Chanenson, *The Next Era of Sentencing Reform*, 54 EMORY L.J. 377, 385-86 (2005); Richard S. Frase, *The*

given very broad discretion to assess the degree of risk posed by the offender, diagnose the causes of that risk, assess whether those causes can effectively and safely be treated without incarceration, and, if they cannot, decide the maximum and sometimes the minimum term of incarceration. Within the maximum and any minimum prison term set by the judge, correctional officials and parole boards are given very broad discretion to further assess the causes of the offender's criminality, the best treatment options available to address those causes, and the precise moment at which, due to prison programming and/or other factors (e.g., maturation or aging) the offender can be safely released. To a lesser extent, indeterminate sentencing regimes also promote specific deterrence. Sentencing judges can use their discretion to determine whether the defendant poses an elevated risk of reoffending, and if so, what additional punishment is needed to offset that risk.

Such highly discretionary determinations of risk, treatment needs, and offender deterrability are very difficult to make reliably and consistently. For this reason, indeterminate sentencing began to fall out of favor in the 1970s.[9] Mounting evidence revealed that virtually identical offenders receive very different sentences from different judges. Similarly, broad parole discretion fell out of favor because of the different treatment given to similar offenders and the inherent difficulty of assessing individualized risk and progress toward reform on the basis of behavior in prison (some model prisoners behave much worse in the community than they did in the highly controlled prison environment; other offenders adapt poorly to the prison regimen but behave much better when released).

The fourth and fifth crime-control methods, general deterrence and denunciation, are designed to prevent future crimes by members of the public at large or certain subgroups believed to have an elevated risk of offending. General deterrence seeks to discourage would-be offenders from committing further crimes by instilling a fear of receiving the penalty given to this offender.[10] General deterrent effects depend on a number of factors: the severity of the penalty; the swiftness with which it is imposed; the probability of being caught and punished; the target group's perceptions of the severity, swiftness, and certainty of punishment; the extent to which members of the target group suffer from addiction, mental illness, or other conditions which significantly diminish their capacity to obey the law; and the extent to which these would-be offenders face competing pressures or incentives to commit crime. As a result of the combined impact of these factors, some offenses and offenders are likely to be easily deterred by the threat of criminal penalties; at

---

Uncertain Future of Sentencing Guidelines, 12 LAW & INEQ. 1, 7 (1993) [hereinafter Frase, *Uncertain Future*].

9. *See, e.g.*, AM. FRIENDS SERV. COMM., STRUGGLE FOR JUSTICE (1971); TONRY, *supra* note 8, at 7, 9-10; Frase, *Uncertain Future*, *supra* note 8, at 7-9.

10. *See* FRANK E. ZIMRING & GORDON J. HAWKINS, DETERRENCE: THE LEGAL THREAT IN CRIME CONTROL 72-73 (1973).

the other extreme, some offenses and offenders are essentially undeterrable.

Several of the deterrence factors listed above interact. For example, a major increase in penalty severity may cause a decrease in the swiftness and/or certainty of punishment. This is because severe penalties give offenders a stronger incentive to vigorously contest the charges and may make prosecutors, judges, and juries reluctant to consistently impose such penalties. Research has found that offenders are more sensitive to the probability of punishment than to its severity.[11] Thus, increased severity may cause crime rates to remain the same or even increase.

Punishment can also prevent crime by means of more diffuse processes that depend on internalized values rather than fear of punishment. Through denunciation (also referred to as the communicative, educative, or expressive function of punishment, or as positive general prevention) criminal penalties serve to define and reinforce important social norms of law-abiding behavior and relative crime seriousness.[12] Such norms guide and restrain behavior even when the chances of detection and punishment are slight. Given the many difficulties of preventing crime by rehabilitation, incapacitation, or deterrent threats, norm-reinforcement is probably the most important crime-preventive effect of punishment (but also the most difficult to measure, since its effects are so diffuse and long term).

In addition to crime control, sentences may achieve several other important utilitarian purposes: promoting satisfaction, closure, and/or compensation for crime victims and victimized communities; reassuring the public that something is being done about crime; and facilitating the offender's successful reintegration into society. Each of these effects is desirable for its own sake but may also help to prevent future crimes by the defendant or other would-be offenders.

Utilitarian theory also imposes several important limitations on the form or severity of punishment. Criminal penalties should not cost more than the benefits they achieve or cause individual or social harms which outweigh their crime-controlling effects or other benefits.[13] Punishment should also be efficient. Penalties should not be more severe or more costly than necessary; if the same crime-control and other benefits can be achieved with less severe or

---

11. *See, e.g.*, ANDREW VON HIRSCH ET AL., CRIMINAL DETERRENCE AND SENTENCE SEVERITY: AN ANALYSIS OF RECENT RESEARCH 5, 47 (1999). The tradeoffs noted in text also explain why, in practice, mandatory minimum penalties provide little crime benefit and make sentencing outcomes less, not more, uniform. *See* TONRY, *supra* note 8, at 134-64.

12. *See* Greenawalt, *supra* note 6, at 1286-87; Paul Robinson & John M. Darley, *The Utility of Desert*, 91 NW. U. L. REV. 453, 468-77 (1997).

13. *See* Richard S. Frase, *Excessive Prison Sentences, Punishment Goals, and the Eighth Amendment: "Proportionality" Relative to What?*, 89 MINN. L. REV. 571, 593-95 (2005) [hereinafter Frase, *Excessive Prison Sentences*] (discussing utilitarian ends proportionality principle).

less costly methods, those methods should be preferred.[14] In a world of limited resources, punishment must also be prioritized. Prison beds and other scarce correctional resources should be reserved for the most socially harmful offenses and offenders. Prisons must also not be used beyond their effective capacities. Overcrowded prisons are unsafe for prisoners and staff, and reduced security and resources for programming increase the odds that prisoners will leave prison more violent or antisocial than when they entered. (Overcrowded prisons are also likely to violate nonutilitarian principles, in particular the right to humane treatment discussed below.)

B. *Nonutilitarian Purposes and Limitations*

Retribution is the most widely recognized nonutilitarian sentencing principle. Under this theory, offenders should be punished in proportion to their blameworthiness (or desert) in committing the crime being sentenced. Two basic elements determine an offender's degree of blameworthiness: the nature and seriousness of the harm caused or threatened by the crime and the offender's degree of culpability in committing the crime.[15] Culpability depends on several factors: the offender's intent (deliberate wrongdoing is more culpable than criminal negligence); his or her capacity to obey the law (which may be diminished by mental disease or defect, chemical dependency, or situational factors such as threats or other strong inducements to commit the crime); the offender's motives for committing the crime (which may mitigate or aggravate culpability); and, for multi-defendant crimes, the defendant's role in the offense as instigator, leader, follower, primary actor, or minor player.

Retribution can serve both as a purpose (positive justification) for punishment and as a limitation on penalties imposed to achieve other purposes. The purpose theory views retribution as the primary or even exclusive goal of punishment—offenders are punished simply because they deserve to be and the severity of their punishment should be no more and no less than they deserve. The underlying moral arguments supporting this view often involve claims of fairness: fairness to the victim and the victim's family (whose right to seek vengeance is supplanted by the criminal law); fairness to law-abiding persons (who refrained from committing this offense); and fairness to the defendant (who, according to this theory, has a right to be punished in proportion to his blameworthiness).

As discussed more fully in Part II, the limiting (negative) version of retributive theory merely sets outer limits on punishment, defining a range of permissible severity for any given case. In terms of the fairness arguments

14. *Id*. at 595-96 (discussing utilitarian means proportionality principle); *see also infra* text accompanying notes 24-29 (discussing the parsimony principle).

15. ANDREW VON HIRSCH, CENSURE AND SANCTIONS 29-33 (1993); Frase, *Excessive Prison Sentences*, *supra* note 13, at 590.

summarized above, retributive limits defining the minimum acceptable penalty reflect concerns about fairness to victims, law-abiding persons, and other offenders; upper-retributive limits, defining the maximum allowable penalty, reflect the widely shared sense that it is fundamentally unfair and an abuse of governmental power to punish an offender more severely than he deserves.

Uniformity is another very important nonutilitarian sentencing principle; similarly situated offenders should receive similar punishments. But uniformity is not a self-sufficient criterion; as was noted at the outset, concepts of uniformity and disparity always require reference to some other principle or principles which render two offenders similarly situated. In recent years retributive values have been the most common reference principle. This combination of retributive and uniformity principles is often called the Theory of Just Deserts; offenders of comparable blameworthiness (in terms of social harm and offender culpability) should receive similar penalties, and offenders differing in blameworthiness should be punished in proportion to their respective degrees of blame.[16]

Like retribution, uniformity is based on concepts of fairness—fairness to other offenders (who could justly complain if this defendant received a lighter penalty for the same conduct), and fairness to the defendant (who could justly complain if he were punished more severely than other equally blameworthy offenders). Uniformity can be a positive sentencing purpose, defining an exact quantum of appropriate punishment, or a limiting principle, defining a range of allowable penalties designed to prevent gross disparities without seeking to impose exact equality.

The civil and human rights of defendants provide another set of nonutilitarian limiting principles. The requirement of humane treatment forbids torture, dismemberment, and other brutal physical or psychological punishments, without regard to whether some offenders might be thought to deserve such penalties or whether the penalties could be deemed necessary and effective to achieve crime control or other utilitarian purposes. Convicts also retain First Amendment and other civil rights which may limit the form or conditions of punishment.

C. *Utilitarian Proportionality and Uniformity*

Sentencing proportionality and uniformity are usually linked to theories of retribution or just deserts, but they also have important utilitarian value.[17] When penalties for different crimes are proportional to the harms caused or threatened by those crimes, offenders have an incentive to stop at the lesser crime. Such proportionality also matches punishment costs with crime-control

---

16. *See generally* VON HIRSCH, *supra* note 15.
17. *See generally* Robinson & Darley, *supra* note 12.

benefits.[18] In addition, more uniform sentencing permits more accurate forecasts of future prison populations and other correctional resource needs.[19]

Denunciation and public respect are additional benefits of sentencing proportionality. As the punishment theorist H.L.A. Hart noted, if "the relative severity of penalties diverges sharply from this rough [proportionality] scale, there is a risk of either confusing common morality or flouting it and bringing the law into contempt."[20] In other words, disproportionate penalties undercut the law's desired norm-reinforcing messages and reduce public respect for the criminal law and criminal justice systems. Such respect can help reduce crime. Research has shown that people are more likely to obey the law if they perceive the law and its processes to be fair.[21] As discussed above, proportionality and uniformity of sentencing are based on widely shared fairness concerns, so highly disparate penalties are likely to reduce the public's willingness to obey the law and cooperate with law enforcement.

D. *Conflicts Within and Across Punishment Principles*

The sentencing principles summarized above are all valid and widely recognized, but they often conflict with each other. In the example cited at the outset of this Article (Case No. 1), uniformity and proportionality principles require that two equally culpable offenders receive equally severe sanctions, despite one's lower risk of reoffending. However, it is difficult to find a sanction that satisfies all relevant sentencing purposes. Sending both offenders to prison uses scarce prison space for a low-risk offender, may make that offender worse, and, if the crime is not serious, contradicts the desired norm-reinforcing message conveyed by punishment. Putting both on probation fails to protect the public from the high-risk offender (at least without substantial increases in probation resources) and, if the crime is serious, fails to provide appropriate general deterrent and norm-reinforcing messages.

Sentencing purposes conflict in many other ways. The following are three more examples:

Case No. 2. Efforts to promote victim or community satisfaction and compensation may result in sanctions which, from a retributive and/or a crime-control and efficiency standpoint, are either too severe (if the victim or community insist on more severity than other purposes require) or not severe enough (if the victim or community are forgiving or want to keep the offender out of prison in order to perform compensatory service or earn the money to

18. *See* Frase, *Excessive Prison Sentences*, *supra* note 13, at 593.

19. *See infra* note 34 and accompanying text (discussing the important benefits of accurate resource-impact assessments in state guidelines systems).

20. H.L.A. HART, PUNISHMENT AND RESPONSIBILITY: ESSAYS IN THE PHILOSOPHY OF LAW 25 (1968).

21. *See* TOM R. TYLER, WHY PEOPLE OBEY THE LAW 31, 64-68 (1990).

pay restitution).

Case No. 3. An offender's mental illness or drug addiction may greatly limit his capacity to obey the law, thus making him less blameworthy and less deserving of severe punishment. But the same offender characteristics make the offender very likely to reoffend, thus justifying a lengthy prison term. Mental health or treatment programs could, in theory, reduce the need for punishment severity in such cases while adequately assuring public safety, but, in the real world, such programs are all too often absent or severely underfunded.

Case No. 4. Increased rates of imprisonment may heighten general deterrent and denunciation effects, but some of the additional offenders sent to prison will probably be made worse (i.e., more dangerous, less able to cope with freedom) than they were before entering prison. In theory this conflict can be resolved by balancing the positive and negative effects of increased rates of imprisonment, but this approach requires substantial data collection and calculation that courts are ill equipped to handle in day-to-day sentencing practice.

How can these various conflicts be resolved by busy courts? The theory of limiting retributivism, discussed below, has been widely adopted and offers the best solution.

II. RECONCILING CONFLICTING PUNISHMENT PRINCIPLES: LIMITING RETRIBUTIVISM

The most serious conflicts among punishment principles occur between case-specific utilitarian purposes and just deserts principles. According to the latter, all offenders should receive their particular deserts—no more and no less. Such a system leaves very little room for consideration of other punishment purposes, and no jurisdiction in the United States or elsewhere has ever adopted such a one-dimensional approach. Instead, almost every system has adopted some form of what Norval Morris called "limiting retributivism" (also known as modified just deserts).[22] Under this widely endorsed and adopted model, the offender's desert defines a range of morally justified punishments, setting upper and lower limits on the severity of penalties that may fairly be imposed on a given offender. These upper and lower limits also promote the utilitarian benefits of uniformity and proportionality.[23] Within the

22. NORVAL MORRIS, MADNESS AND THE CRIMINAL LAW 161, 182-87, 196-200 (1982) [hereinafter MORRIS, MADNESS]; *cf.* MINN. SENTENCING GUIDELINES COMM'N, REPORT TO THE LEGISLATURE 9 (1980) (adopting modified just deserts approach). *See generally* Richard S. Frase, *Limiting Retributivism*, *in* THE FUTURE OF IMPRISONMENT 83, 90-104 (Michael Tonry ed., 2004) [hereinafter Frase, *Limiting Retributivism*] (discussing widespread support for basic elements of limiting retributivism); Richard S. Frase, *Sentencing Principles in Theory and Practice*, 22 CRIME & JUST. 363, 365-78 (1997) [hereinafter Frase, *Sentencing Principles*] (summarizing Morris's theory of punishment and Minnesota's approach).

23. *See supra* text accompanying notes 17-21.

range of deserved penalties, case-specific incapacitation, rehabilitation, deterrence, and other sentencing goals may be pursued, but only to the extent that they are needed in a given case. Sentences within the desert range should be no more severe than necessary to achieve defined aims, a humane and utilitarian principle of necessity and efficiency which Morris referred to as "parsimony."[24]

Morris argued that desert can only define a range of penalties because the very concept of desert is inherently imprecise.[25] In any given case there will be widespread agreement that certain penalties are clearly undeserved (because they are either excessively severe or excessively lenient). But there may be little political or philosophical consensus on the offender's precise deserts, even relative to other offenders committing the same crime.

But if a range of morally permissible penalties exists, why not sentence all offenders at the top of the range, or at least use that as the starting point in order to maximize crime-control effects? Morris's opposite presumption, in favor of the least severe penalty in the range, is based on both moral and practical grounds.[26] The moral arguments are analogous to those that underlie the requirement that guilt be established by elaborate trial procedures and proof beyond a reasonable doubt: punishment intrudes on physical liberty and other very important rights, and the crime-control benefits of punishment are uncertain. Thus, the burden should be on the state to justify each additional increment of punishment severity.

The practical arguments for preferring sentences less severe than the offender's maximum desert flow not only from efficiency concerns (less severity is often adequate to achieve all utilitarian goals) but also from the pervasive need to encourage and reward cooperation from those accused of crime. Given the state's limited resources and powers, there is a compelling need to obtain guilty pleas, waivers of jury trial and other rights, and cooperation in convicting other defendants. In addition, defendants placed on probation or parole must be encouraged to obtain and hold employment, support their dependents, make restitution, avoid risky places or behaviors, and accept treatment and supervision. Incarcerated defendants sent to prison or jail must have an incentive to cooperate with institutional rules and programs. Sentences must also leave room for backup sanctions—subsequent tightening of control (e.g., by revocation of probation or parole) if the defendant fails to cooperate—even if that failure is not, in itself, very blameworthy. Thus, in practice, modern systems of law enforcement and punishment always function

24. NORVAL MORRIS, THE FUTURE OF IMPRISONMENT 59-62 (1974) [hereinafter MORRIS, FUTURE OF IMPRISONMENT]; Frase, *Excessive Prison Sentences*, *supra* note 13, at 595-97.

25. MORRIS, MADNESS, *supra* note 22, at 198; *see also* NORVAL MORRIS & MICHAEL TONRY, BETWEEN PRISON AND PROBATION: INTERMEDIATE PUNISHMENTS IN A RATIONAL SENTENCING SYSTEM 84-89, 104-05 (1990).

26. MORRIS, FUTURE OF IMPRISONMENT, *supra* note 24, at 61.

according to a limiting retributive model under which most offenders, in return for their cooperation, receive less severe sanctions than the maximum they deserve.

For all of these moral and practical reasons, limiting retributivism has been widely endorsed by scholars, model code drafters, legislators, sentencing commissions, judges, and practitioners.[27] Desert-based limits on maximum sanction severity have received particularly broad support.[28] Morris's concept of parsimony—that sentences within the deserved range should be no more severe than needed to achieve all relevant utilitarian purposes—has been strongly promoted by utilitarian philosophers as far back as Beccaria and Bentham, and finds support in the Model Penal Code (the original version and recent proposed revisions), all three editions of the American Bar Association sentencing standards, and many fields of American law.[29]

Some version of limiting retributivism has also been the basis for most contemporary sentencing laws.[30] Indeterminate systems (still the most common sentencing regime) reflect a very loose version of this approach; most American sentencing guidelines systems embody a more precise and structured form of limiting retributivism.[31] A range of allowable sanctions is provided for each case. The limits of sanction severity (tops and bottoms of the ranges) are scaled according to offense severity and prior conviction record, with the former usually having greater weight. Unlike the Federal Guidelines, state guidelines systems retain considerable sentencing flexibility, and mandatory minimum sentences are less prevalent and severe than in the federal system.[32]

The Minnesota Sentencing Guidelines, in effect since 1980, represent a good example of limiting retributivism in operation.[33] The modified just

---

27. Frase, *Limiting Retributivism*, *supra* note 22, at 90-104; *see also* CONSTITUTION PROJECT, PRINCIPLES FOR THE DESIGN AND REFORM OF SENTENCING SYSTEMS (2005) (noting that crime-control purposes should operate within upper and lower proportionality limits), http://www.constitutionproject.org/si/Principles.doc (last visited Sept. 13, 2005) .

28. Frase, *Limiting Retributivism*, *supra* note 22, at 92-93 (noting the support of influential writers such as H.L.A. Hart and the drafters of the Model Penal Code, recent proposals to revise the Code, and the two most recent editions of the American Bar Association's sentencing standards).

29. *Id*. at 94-95; *see also Frase*, *Excessive Prison Sentences*, *supra* note 13, at 595-627 (discussing many examples of principles akin to parsimony in United States constitutional law and in foreign and international law).

30. Frase, *Limiting Retributivism*, *supra* note 22, at 95-104.

31. *Id*. at 97-104; Frase, *Sentencing Principles*, *supra* note 22, at 407-30 (discussing the limiting retributive features of Minnesota's guidelines); Hofer & Allenbaugh, *supra* note 1 (arguing that the Federal Guidelines implicitly embody the modified just deserts model).

32. *See* Richard S. Frase, *State Sentencing Guidelines: Diversity, Consensus, and Unresolved Policy Issues*, 105 COLUM. L. REV. 1190, 1198-1201, 1208 (2005) [hereinafter Frase, *State Sentencing Guidelines*]; *see also* Richard S. Frase, *Sentencing Guidelines in Minnesota, Other States, and the Federal Courts: A Twenty-Year Retrospective*, 12 FED. SENT'G. REP. 69, 69, 76 (1999).

33. MINN. SENTENCING GUIDELINES AND COMMENTARY (rev. Aug. 1, 2005), http://www

deserts sentencing philosophy adopted in Minnesota sets desert-based upper and lower limits on sanction severity, recognizes the principle of parsimony, leaves substantial scope for the application of non-desert sentencing purposes, retains substantial case-specific flexibility, and tailors overall sanction severity levels to available resources.[34] Minnesota's guidelines have achieved a stable and workable balance between the conflicting purposes and limitations of punishment.

A. *The Limits of the Criminal Law as an Instrument of Crime Control*

For most offenses and offenders, the crime-control effects of punishment probably are quite limited.[35] One of the most important factors determining the deterrent effect of sanctions is the probability of punishment.[36] But the detection rates for most crimes are very low, and the probability of an offender receiving a custody sentence is often less than one out of every one hundred crimes committed.[37] Low detection, conviction, and incarceration rates also severely limit the effectiveness of criminal sanctions designed to achieve rehabilitation or incapacitation. Moreover, each of these crime-control mechanisms relies on numerous additional assumptions which often do not prove true.[38]

The crime-control effects are likely to be particularly limited for many federal crimes. Although separate data on federal crimes are not available, there is no reason to believe that crime-detection rates are higher for federal crimes than for other crimes. Moreover, police and prosecutors are very selective in deciding what to make a federal case.[39] Federal criminal law often overlaps

---

.msgc.state.mn.us/Guidelines/guide05.DOC (last visited Sept. 13, 2005). *See generally* Frase, *Sentencing Principles*, *supra* note 22.

34. Frase, *Sentencing Principles*, *supra* note 22, at 388-430. *See generally* Richard S. Frase, *Sentencing Guidelines in Minnesota, 1978-2003*, 32 CRIME & JUST. 131 (2005). Although Minnesota's guidelines retain substantial case-level flexibility, sentencing practices are sufficiently uniform to permit the Minnesota Guidelines Commission to make accurate forecasts of the impact of changes in guidelines or statutes on future prison populations and other correctional resources. Such resource-impact assessments have allowed Minnesota to set priorities in prison use and avoid prison overcrowding. *Id*. at 146-47, 204-05. Several other guidelines states have taken a similar approach, but resource-impact assessments have never played any significant role in the federal system. Frase, *State Sentencing Guidelines*, *supra* note 32, at 1198, 1216-19.

35. For further discussion, Michael Tonry, *The Functions of Sentencing and Sentencing Reform*, 58 STAN. L. REV. 37 (2005) (in this Issue).

36. *See supra* note 11 and accompanying text.

37. *See, e.g.*, CROSS-NATIONAL STUDIES OF CRIME AND JUSTICE 57 (David Farrington et al. eds., 2004) (showing U.S. custody rates per 1000 crimes committed for various crimes).

38. *See supra* text accompanying notes 6-11. In contrast, early childhood intervention and other crime-preventive measures outside the criminal justice system have been found to be quite effective. *See* Frase, *State Sentencing Guidelines*, *supra* note 32, at 1219 n.138.

39. *See* Richard S. Frase, *The Decision To File Federal Criminal Charges: A*

with state law, so federal law officers can defer to state authorities. Even where federal criminal jurisdiction is exclusive (for example, in many areas of regulatory crime and crime committed against the federal government), criminal prosecution is usually reserved for aggravated or persistent violations.[40] In recent years, a large proportion of suspects prosecuted in federal courts were charged with drug violations,[41] but about ten times as many drug offenders were prosecuted in state courts.[42] Given the clandestine nature of drug use and drug trafficking, it seems likely that detection rates are at least as low for these offenses as for most other crimes. Furthermore, drug crimes are very hard to deter, since they are often motivated by addiction, high profits (maintained by supply-side enforcement), and/or lack of attractive lawful activities. White-collar and regulatory offenders are more likely to be deterred, even by selective enforcement and modest penalties; such offenders have many lawful alternatives and much to lose from being convicted, regardless of the penalty. But the highly selective use of criminal penalties in such cases raises very serious uniformity problems. When offenders appear to have been unfairly singled out, respect for the law and law enforcement suffers.[43]

B. *Illustrative Cases*

The four examples of conflicting sentencing purposes described above can be used to illustrate how sentencing guidelines based on the limiting retributive model can help provide resolution. The first step is to determine whether the top and bottom of the recommended guidelines range need to be adjusted to account for factors which increase or decrease the seriousness of the conviction offense. Applying the parsimony principle, the sentencing judge then begins at

---

Quantitative Study of Prosecutorial Discretion, 47 U. CHI. L. REV. 246, 252 tbl.1 (1980) [hereinafter Frase, *Decision To File*] (finding that from 1974 to 1978, approximately 22% of criminal matters were prosecuted). Recent published reports show higher federal prosecution rates. *See, e.g.*, BUREAU OF JUSTICE STATISTICS, FEDERAL CRIMINAL CASE PROCESSING 2002, at 9 tbl.3 (2005) (noting that 73% of suspects investigated were prosecuted). However, these data exclude matters on which attorneys spent less than one hour of time. *Id.* at 17. In earlier years Justice Department data included all matters received from investigating agencies, producing much lower prosecution rates. *See* Frase, *Decision To File*, *supra* (citing Justice Department statistical reports); *see also id.* at 254, 321 (finding that in the Northern District of Illinois, half of matters were immediately declined, and only 14% of these were declined because no federal crime appeared to have been committed).

40. *See* BUREAU OF JUSTICE STATISTICS, *supra* note 39, at 9 (finding a 62% declination rate for regulatory crime). As discussed in note 39, *supra*, more complete data from earlier years suggests that true declination rates are much higher than recent published rates.

41. *Id.* at 10 (noting 30,673 drug defendants in 2002, comprising 35% of defendants prosecuted in federal courts).

42. BUREAU OF JUSTICE STATISTICS, FELONY SENTENCES IN STATE COURTS, 2002, at 2 (2004) (finding that an estimated 340,330 defendants were charged with drug crimes in state courts).

43. *See supra* text accompanying note 21.

the low end of the adjusted range and increases sentencing severity until all applicable case-specific utilitarian sentencing purposes have been satisfied.

Case No. 1.[44] In this example, two offenders are identical in everything but the risk of reoffending. This case is easily resolved by the state guidelines limiting retributive model. Within the adjusted range, offenders may receive unequal punishments based on risk assessments. The lower and upper boundaries limit unfairly lenient or severe sentences and satisfy utilitarian (general deterrent and denunciation) needs for uniformity and proportionality.

Case No. 2.[45] This example highlights a conflict between the needs of victims and the community, on the one hand, and just desert values and/or crime-control goals, on the other. Again, the upper and lower adjusted range boundaries address uniformity and proportionality concerns. Within the range, case-specific crime control and victim or community needs may all be considered; whatever purpose calls for the greatest sanction will prevail unless the legislature provides that a particular purpose should receive priority and should trump the others.

Case No. 3. This example involved a high-risk offender with substantially diminished capacity to obey the law due to mental illness or addiction. Fairness to the offender, as well as the utilitarian advantages of uniformity and proportionality, require a desert-based upper limit on allowable punishment severity. If this limit appears inadequate to ensure public safety, mental health or other noncriminal alternatives can be invoked; public safety imperatives, combined with upper desert limits, give the legislature an incentive to provide the necessary alternatives.

Case No. 4. This example posited a conflict between the need for increased deterrence (for instance, because of a recent increase in frequency of the crime) and the risk of damaging vulnerable offenders. The latter risk is lessened by the upper range limits and by the court's discretion within the range to give a shorter prison or jail term (or, where authorized, a noncustodial sentence) to offenders who appear particularly vulnerable.[46]

CONCLUSION

Many legitimate purposes and limitations apply to punishment decisions. These purposes and limits require both discretion to tailor sentences to particular offense and offender circumstances and substantial limitations on

44. This case is described *supra* in the Introduction of this Article and is further discussed *supra* at the end of Part I.

45. Cases 2, 3, and 4 are described *supra* at the end of Part I.

46. *See, e.g.*, State v. Wright, 310 N.W.2d 461, 462-63 (Minn. 1981) (approving probation, jail, and treatment, in lieu of a presumptive twenty-three to twenty-five-month prison term, where the defendant was very amenable to treatment on probation and his immaturity would cause him to be easily victimized or led into crime if he were sent to prison).

that discretion. It is intolerable to allow three identical offenders to receive very different sanction severity because one judge believes in community-based treatment, one believes in proportionality limits, and one believes in using very lengthy prison terms for deterrence. It is likewise intolerable to force judges to impose identical sentences when different penalties are warranted. Sentencing judges must have guided discretion, and they must be clear about which purposes they are pursuing and with what priority. When defining the limits of the sentencing judge's discretion, legislatures, sentencing commissions, and appellate courts must likewise be clear about purposes and priorities.

The principles of sentencing uniformity and proportionality reflect widely shared fairness concerns and have great practical value. But these principles are often in conflict with case-specific crime-control and restorative justice goals, and the two latter purposes often conflict with each other. The theory of limiting retributivism has broad support and provides a workable means of harmonizing all of the important purposes and limitations of punishment. State guidelines, such as those in Minnesota, provide good examples of successful limiting retributive systems in operation.[47]

The Federal Guidelines are not based on limiting retributivism,[48] but courts or Congress could easily adapt the Guidelines to that model. Justice Breyer's remedy opinion in *Booker* directs courts to give particular emphasis to the statement of sentencing purposes and other factors in section 3553(a) of the Guidelines-enabling statute.[49] That section clearly shows that Congress preferred a hybrid theory of sentencing purposes, and it strongly suggests that Congress had something like limiting retributivism in mind.

Section 3553(a) begins with a statement of the parsimony principle,[50] and paragraph 1 specifies that sentences should be based on both offense and offender factors. Paragraph 2 then lists the principal sentencing purposes courts should consider—paragraph 2A appears to endorse proportionality values,[51] and paragraphs 2B to 2D recognize deterrence, incapacitation, and rehabilitation.[52] Later paragraphs endorse sentencing uniformity and restitution.[53] As shown in this Article, the purposes listed in section 3553(a) are

---

47. Frase, *State Sentencing Guidelines*, *supra* note 32, at 1211; Hofer & Allenbaugh, *supra* note 1, at 24.

48. *But see* Hofer & Allenbaugh, *supra* note 1 (arguing that a form of limiting retributivism is implicit in the Guidelines).

49. United States v. Booker, 125 S. Ct. 738, 764-66 (2005).

50. 18 U.S.C. § 3553(a) (2000) (stating that the sentence should be sufficient, but not greater than necessary, to comply with the purposes set forth in paragraph (2) of this subsection).

51. *Id.* § 3553(a)(2)(A) (stating that courts shall consider the need for the sentence to reflect the seriousness of the offense, to promote respect for the law, and to provide just punishment for the offense).

52. *Id.* § 3553(a), ¶¶ 2B-2D.

53. Paragraphs 3 through 5 of § 3553(a) direct courts to consider available penalties and the Sentencing Commission's Guidelines and policy statements. Paragraphs 6 and 7 cite

in conflict, but they can be harmonized under a limiting retributive model. Indeed, the structure of section 3553(a) strongly implies this model: the parsimony and proportionality principles are stated first, suggesting that they set overall limits on the crime-control and other purposes which follow.

The limiting retributive model would bring much-needed coherence to federal sentencing law. In the absence of further congressional action, federal trial and appellate courts should interpret section 3553(a) consistently with this model, using the Guidelines ranges, case-specific aggravating and mitigating circumstances, and other considerations[54] to define proportionality limits within which all other sentencing purposes operate, subject to the overall requirement of parsimony. Of course, it would be preferable for Congress to explicitly endorse the limiting retributive model. Section 3553(a) could easily be rewritten to achieve this result, separately stating the three critical elements of this model: outer proportionality limits, crime-control and other case-specific considerations operating within those limits, and the principle of parsimony.[55]

Whether in federal or state court, sentencing must accommodate a variety of punishment purposes and limitations. The theory of limiting retributivism provides the best means to reconcile and give appropriate weight to all of these purposes and limitations. This approach is the basis for most contemporary sentencing laws, and its essential elements have been widely endorsed by scholars, model code drafters, sentencing commissions, and practitioners. As implemented in Minnesota and other guidelines states during the past twenty-five years, limiting retributivism has proven to be theoretically sound, well-balanced, flexible, and practically viable.

---

the need to avoid unwarranted disparities and to provide restitution to victims.

54. *Cf.* Frase, *Excessive Prison Sentences*, *supra* note 13, at 651 (suggesting that the Eighth Amendment and other constitutional proportionality principles could be adapted to help federal courts define subconstitutional proportionality standards under § 3553(a)).

55. For a clear statement of these three elements, see MODEL PENAL CODE: SENTENCING § 1.02(2)(a)(i)-(iii) (Preliminary Draft No. 3, 2004).