Florida Law Review

Volume 61 | Issue 4 Article 7

September 2009

Tragedy, Skepticism, Empirics, and the MPCS

Robert Weisberg

TRAGEDY, SKEPTICISM, EMPIRICS, AND THE *MPCS*

Robert Weisberg[*]

I. INTRODUCTION: THE *MPCS* AND THE MEASUREMENT OF SUCCESS

I argue here that an important implicit theme of the Model Penal Code: Sentencing (*MPCS*) Draft (Draft) is what I will call modern tragic skepticism about empirical proof in legal reform. By an admittedly somewhat stretched reading, I will infer from the Draft both the negative and positive aspects of this tragic view of the world: (1) a sober recognition of the limited human knowledge about the effects of our laws that is offset (2) by a sensible, chastened, realistic commitment to a spirit of reasonable experimentation in an unknowable world.

Thus, I am taking this occasion to reflect on our ambivalence about our ability to ever judge the success of legal reform. And I am doing so at a time when the empirical predicates of any such judgment are subject to the

* Edwin E. Huddleson, Jr. Professor of Law, Stanford University; Director, Stanford Criminal Justice Center.

daunting demands of modern social science. In the context of the *MPCS*, this ambivalence results in part from the impossibility of coherently formulating the consensus goals for a sentencing system. But it also results from our recognition that even assessing means or the measures for the means' effect may prove quixotic if we aim to satisfy the standards of modern statistics.

Although this is a contestable distinction, the *MPCS* Draft is a law reform document, not a policy design proposal. It rests on the American Law Institute (ALI)-venerated mixture of reliance on legal precedent (in the broad sense of legal tradition—not, of course, narrowly binding case law), consistency with jurisprudential principle, appeal to common sense fairness, respect for relative institutional competence, and general commitment to efficacy and economy. Thus, the Draft does not commit itself to the type of future empirical evaluation associated with "policy" programs designed by social science experts to achieve functionally specific goals within the legal system.

A. MPCS' *Commitment to Empirical Measurement*

Although a reform document, the Draft is substantially committed to empirical measurement in much of its design. For example, it calls for the forecasting of the financial and demographic effects of new sentencing laws and expresses some confidence that this is a rational and realistic exercise to impose on lawmakers.[1] Additionally, in all the micro-level decisions that any penal and sentencing scheme makes, the Draft is deeply committed to that large, yet vague, body of programmatic thought called evidence-based practices.[2] Moreover, the Draft's fundamental emphasis on a version of retributivism[3] does not preclude empirical measurement, since the notion of an empirical evaluation of a retributivist legal program is not oxymoronic. For example, to the extent that retributivist ideas should rest on common moral values, one can use survey data to test consistency with such values, as Paul Robinson has done in his important work on perceptions of ordinal measures of moral salience of crimes.[4] Indeed, as I later suggest regarding the *MPCS*, another much simpler empirical test of the "success" of a retributivist program is its sheer duration or survival, that is, its ability to withstand political change rooted in moral disagreement.[5]

---

1. *See* MODEL PENAL CODE: SENTENCING § 1.02 (Preliminary Draft No. 6, 2008).

2. *Id*. §§ 7.07B, 6B.09 cmt. d; MODEL PENAL CODE: SENTENCING REPORT 30–32 (2003).

3. MODEL PENAL CODE: SENTENCING § 1.02(2)(a) (Preliminary Draft No. 6, 2008); MODEL PENAL CODE: SENTENCING REPORT, *Reporter's Introduction* 35–41 (2003).

4. *See, e.g.*, Paul H. Robinson & Robert Kurzban, *Concordance and Conflict in Intuitions of Justice*, 91 MINN. L. REV. 1829, 1832 & n.6 (2007).

5. *Id.*

Nevertheless, a comprehensive legal reform such as the Draft does not cast itself as a designed experiment, subjected to statistical verification, because it is much too vast and complex an undertaking and is too inflected with deontological value for such a test. Although there have been many critical reappraisals of the substantive criminal law parts of the Model Penal Code (MPC),[6] as far I am aware, there have been no "policy evaluations" of its utilitarian achievements because the MPC was never conceived in terms for which such evaluation would be expected. And while the Draft's promotion of sentencing-effect forecasting and evidence-based decision-making reflects the *MPCS'* noble utilitarian goals, these goals are a far cry from demanding or expecting formal social science verification of the operations of the system.

B. *Examples of the* MPCS' *Treatment of Specific Statistics*

It is precisely against this background that I observe a few important moments in the Commentary to the Draft that invoke fairly technical matters of empirical evaluation—and passionate, even defensive, emphasis. One moment concerns the debate over the efficacy of marginal discretionary parole judgments and whether they provide a marginally useful incentive to good behavior by prisoners.[7] The other, larger one is about whether the contemporary shift towards greater determinacy in sentencing has contributed to the huge and much-derided increase in the American incarceration rate in recent years.[8] I am not concerned with the validity of what the Commentary says on these subjects, although it makes a very convincing case. Rather, in this article I draw out the implications of these moments to ask how, *if at all*, we should evaluate the sentencing reform scheme for success or failure. I extend this inquiry into areas the *MPCS* does not directly address in any detail, but which are implicit in any discussion of sentencing reform—such as fiscal effects, political dynamics, and racial fairness.

1. The Commentary's Treatment of Parole and Statistics

On the first subject, the Commentary is subtly agnostic. It treats this question about parole release decisions against the background of the Draft's major premise—the commitment to a version of determinate sentencing and especially, to the abolition of parole board discretionary

6. In the major collections of essays appraising the MPC, the crime-definition parts of the Code (as opposed to the earlier sentencing provisions) receive normative reconsideration but are not subjected to any empirical study of their effects on crime or criminal justice. Richard G. Singer, *Foreword*, 19 RUTGERS L.J. 519, 519–20 (1988); *see also* Paul H. Robinson, *Structuring Criminal Codes to Perform Their Function*, 4 BUFF. CRIM. L REV. 1, 1–11 (2001).

7. MODEL PENAL CODE: SENTENCING, *Reporter's Introduction* at 3–13 (Preliminary Draft No. 6, 2008).

8. *Id.* at 14–29.

release. The Draft, of course, calls for considerable control over the structure of this determinacy through a commission or guidelines approach, and it permits a small exception for *judges* to reappraise especially long sentences late in the duration of the original sentence.[9] However, the Draft's commitment to determinacy is clear. It concludes that the research offers no convincing evidence of any comparative advantage possessed by parole officials to judge rehabilitation or predict behavior.[10] Those notions have set a baseline assumption against any delegation of authority to parole authorities or any faith in the utility of engineering or measuring rehabilitation late in imprisonment. In effect, the Draft places a huge burden of proof on any empirical argument to the contrary. But such a high bar is set because of what the Draft perceives to be very strong empirical evidence on one side of the question and because, in the Draft's view, such discretion violates more deontological notions of proportionality and equity.[11]

Thus, in the view of the Draft, the deck is stacked against the proposal for a small exception for inducing late-term good behavior and self-improvement, as some otherwise sympathetic to the abolition of discretionary parole have proposed.[12] The Commentary pays heed to the admonition of Norval Morris that predictions of post-release behavior can be pretty much set by pre-incarceration static factors, and that the incentive theory of discretionary release backfires because it simply invites feigning by manipulative prisoners.[13] The Commentary does not explicitly commit itself to Morris' skepticism, but it makes a clear, if implicit, decision about burden of proof.[14]

Citing the strongest new study which supposedly finds some efficacy in these marginal incentive programs,[15] the Commentary purports to refute this study on two key methodological grounds. First, the study's numbers are dominated by the bizarre anomaly of the California parole system. Second, because revocation criteria vary widely among the states, the study cannot draw reliable conclusions about the relationship between parole incentive and good behavior.[16] The Commentary concludes, almost

9. *Id.* at 52–53.

10. *Id.* at 2.

11. *Id.* at 4–6.

12. *See generally* JOAN PETERSILIA, WHEN PRISONERS COME HOME: PAROLE AND PRISONER REENTRY (2003) (discussing the origins and evolution of modern parole).

13. NORVAL MORRIS, MACONOCHIE'S GENTLEMEN: THE STORY OF NORFOLK ISLAND & THE ROOTS OF MODERN PRISON REFORM 183 (2002).

14. MODEL PENAL CODE: SENTENCING, *Reporter's Introduction* 6 (Preliminary Draft No. 6, 2008).

15. Connie Stivers Ireland & JoAnn Prause, *Discretionary Parole Release: Length of Imprisonment, Percent of Sentence Served, and Recidivism*, 28 J. CRIME & JUST. 27, 27–50 (2005).

16. MODEL PENAL CODE: SENTENCING, *Reporter's Introduction* 11–14 (Preliminary Draft No. 6, 2008).

ominously, that "we should be wary of building the important components of a sentencing system, especially rules and policies that apply to large numbers of prisoners, upon an absence of knowledge."[17]

2. The Commentary's Treatment of Incarceration Rates and Statistics

As for the larger question, Professor Reitz himself has taken great pains in recent years to argue vigorously against what he sees as a common perception that the guidelines or commission model has generally increased incarceration rates.[18] He has offered studies indicating that most of the modern increase in incarceration has happened under old-fashioned indeterminate or unstructured schemes, and that more contemporary cross-state comparisons do not support any inference that the guidelines schemes lead to incarceration rate increases.[19] The Commentary augments that argument with increased detail and passion. It concedes the superficial logical appeal of the notion that determinacy increases prison populations and ascribes that appeal in part to the anomalous example of the federal system.[20] But the Commentary casts doubt on that logic by noting that parole officials are actually very weak regulators of prison size:[21] they are, after all, subject to political pressure and can only control the back-end, not the front-end.[22]

More important, the Commentary amasses impressive statistics to show that, if anything, states that retained indeterminate sentencing and back-door parole release well into the modern determinate era showed larger than average growth in prison populations.[23] But it is the normative context in which the Commentary places these numbers that is most telling. The Commentary cautiously avoids taking a position in favor of reducing prison populations, and, indeed, even avoids attributing such a goal to the proponents of Sentencing Guidelines Parole Release Abolition (SGPRA) regimes, at least in any categorical sense.[24] Rather, the Commentary suggests that a desire to relieve prison populations may have run *parallel* to the independent goals—presumably the limited retributivist and proportionality goals—of SGPRA, and that any legal system that overtly

17. *Id.* at 14.

18. Kevin R. Reitz, *Don't Blame Determinacy: U.S. Incarceration Growth Has Been Driven by Other Forces*, 84 TEX. L. REV. 1787, 1794 (2006).

19. *Id.* at 1798–1801.

20. MODEL PENAL CODE: SENTENCING, *Reporter's Introduction* 14–15 (Preliminary Draft No. 6, 2008).

21. *Id.* at 14–16.

22. MODEL PENAL CODE: SENTENCING REPORT, *Reporter's Introduction* 64–65 (2003).

23. MODEL PENAL CODE: SENTENCING, *Reporter's Introduction* 16–22 (Preliminary Draft No. 6, 2008).

24. *Id.* at 23–29; MODEL PENAL CODE: SENTENCING REPORT, *Reporter's Introduction* 76–85 (2003).

seeks to change the prison population in any direction can do so quite independently of its stances on determinacy and parole.[25]

The Commentary not only eschews using statistics to prove the success of the move to determinacy—it even denies that any particular stance on prison population is necessarily the goal of any sentencing reform. It implies that anyone objecting to a sentencing reform proposal on the ground that, other things being equal, it increases the prison population cannot carry the necessary burden of proof. The Commentary thereby suggests that critics of SGPRA either wrongly assume increased populations are bad or wrongly ascribe to determinacy-advocates the goal of reducing populations.[26] Therefore, the Commentary uses statistics solely to undermine the strength of that latter claim. There is a kind of deeper uncertainty in the Commentary at this point. It does not so much suggest that careful statistics can prove success, but rather suggests that careful statistics can sometimes *dis*prove or at least cast grave doubt on suspiciously easy or confident claims of some particular effect.

Before I discuss the implications of tragic skepticism for sentencing reform, I provide in Part II a background for this theme by drawing on the work of Franklin Zimring and Gordon Hawkins. I then illustrate in Part III how tragic skepticism has pervaded the analysis of programs at the center of the 1990s crime drop, emphasizing the work of Richard Rosenfield and his fellow scholars. Taking my analysis of tragic skepticism regarding the 1990s crime drop, I connect my discussion back to the *MPCS* and the implications of multiple causation factors in Part IV. In Part V, I discuss the tentative speculation regarding the measurement of sentencing reform success, revisiting the moments of incarceration rates and parole among other issues. Finally, in Part VI, I propose duration as possible measurement for sentencing reform success.

II. THE BACKGROUND OF TRAGIC SKEPTICISM

To elaborate this theme of tragic skepticism, I beg the reader's indulgence as I engage in a two-stage diversion from the immediate focus on the *MPCS*—although, the connection back to the *MPCS* should soon be clear. The first stage involves the Commentary's subtle reference to *The Scale of Imprisonment*, the magisterial 1991 work of Franklin Zimring and Gordon Hawkins.[27] Because I believe this reference is very telling, I will extend my discussion of the work further than the Commentary does. The Commentary cites Zimring and Hawkins chiefly for a factual point about the wide variety of incarceration rates among the U.S. states and how that variety negates any easy inferences about national trends.[28] Yet, the

25. MODEL PENAL CODE: SENTENCING, *Reporter's Introduction* 27–29 (Preliminary Draft No. 6, 2008).

26. *Id.* at 23–29.

27. FRANKLIN E. ZIMRING & GORDON HAWKINS, THE SCALE OF IMPRISONMENT (1991).

28. MODEL PENAL CODE: SENTENCING, *Reporter's Introduction* 18 (Preliminary Draft No. 6,

Commentary thereby teases those who know *The Scale of Imprisonment* about possible approval of the book's major thesis (for which the technical point is relevant as supporting evidence). That major point is the very apex of empirical agnosticism because Zimring and Hawkins find no clear evidence of any causal link between any penal or sentencing policy and the rate of imprisonment in any polity.[29]

Reviewing vast bodies of historical and empirical research, both domestic and international, Zimring and Hawkins draw very bracing inferences. One such inference is that a polity's general jurisprudence of imprisonment rarely expresses anything very precise about desirable numbers of prisoners or the duration of sentences, which may be just as well, because implementations of penal systems seem to have little predictable effect on prison population.[30] One reason for the absence of jurisprudential commitment to such numerical measures is that even in a somewhat controllable system, vast leeway to influence the numbers is delegated to all the actors in the system so that control lies beyond the role of abstract jurisprudence or legislated policy.[31] In the United States, rough similarity of policy among the states is belied by vast differences in imprisonment rates, leading to the default conclusion that "social forces" are the explanation.[32]

Zimring and Hawkins then posit, and strike down, various common hypotheses for explaining imprisonment rates.[33] Perhaps most soberingly, crime rates seem to have little demonstrable relation to prison populations.[34] The disconnects between the two are astonishing and well-documented, as illustrated by the absence of any incarceration increase during the great American index crime spike in the 1960s.[35] As Zimring and Hawkins explain, "[A] fact of overriding importance is that imprisonment is not the statistically usual response to felony crime in the United States."[36] As for other possible explanations or even correlations, there is little solid support for the build-and-fill theory—whereby prisons tend to accept prisoners up to their capacity, nor its obverse—that penal

---

2008); *see also* MODEL PENAL CODE: SENTENCING REPORT, *Reporter's Introduction* 67 (2003).

29. ZIMRING & HAWKINS, *supra* note 27, at xii, 122.

30. *Id.*

31. *Id.* at xii.

32. *Id.* at xii, 119–36.

33. *Id.*

34. *Id.* at 121–24.

35. *Id.* at 122, 126.

36. *Id.* at 124. Among the numerous reasons why this may be is that younger criminals in new crime waves do not always get sent to prison, and that legislative responses to crime waves may affect the theoretical length of sentences but not the likelihood of conviction or post-conviction imprisonment. *Id.* at 122–25. Zimring and Hawkins would probably agree that the latter effects sometimes occur through pressure to plead guilty generated by mandatory minimum sentences or rigid enhancements, but presumably they would also say that such effects are very diffuse and unpredictable.

policy seems to produce no more prisoners than can be accommodated by existing capacity.[37] Nor can we clearly link prison populations to "lock-them-up" politicking, because the public is rarely given an explicit choice to increase prison populations, and the time lag between tough penal policy and effects on prison population is often so long as to diffuse political pressure.[38]

The reasons for skepticism continue: American geographic regions seem to be dominant correlates, but a region is not a jurisdiction with any distinct policies. Furthermore, some links for which a modicum of statistical evidence has emerged seem random, if not counter-intuitive. Thus, even policies that may be motivated specifically to reduce reliance on prison will sometimes increase imprisonment.[39] Regimes of alternative sanctions often end up simply widening the net of criminal supervision and feeding more people into prisons; they serve not as replacements for prisons, but as pressure-release valves to enable the state to retain large prison censuses.[40] And the seemingly neutral and uncontroversial goal of improving the flow of information within the criminal justice system may itself tend to increase incarceration because officials, when presented with more data about criminals, tend to incarcerate out of risk-aversion.[41]

The overall problem is a vast entanglement of variables. For example, as Zimring and Hawkins note, the very political pressures that lead to prison construction may independently increase prosecutions so as to additionally increase prison populations.[42] For Zimring and Hawkins, the modern boom in incarceration rate is a "compound mystery."[43] Having cast doubt on all explanations rooted in purposeful programs, they lament that "there is no simple set of social variables thought to be related to imprisonment that would explain the increase as a consequence of forces external to governmental criminal justice policy."[44] We are left with notions of national "trend" and "mood" that Zimring and Hawkins consider "nebulous" and even "illusory."[45]

Zimring and Hawkins, however, at least *purport* to be optimistic about the possibility of research that might yet meet their tough criteria.[46] On the one hand, they call for large, broad, historical, and comparative studies that look at significant changes.[47] Thus, they suggest that if the decades-long

37. *Id.* at 76–78.
38. *Id.* at 125–30.
39. *Id.* at 185–86.
40. *Id.*
41. *Id.* at 173–74.
42. *Id.* at 173.
43. *Id.* at 175.
44. *Id.*
45. *Id.*
46. *Id.* at 176.
47. *Id.*

upward curve in imprisonment in the United States flattens out and reverses itself, we might then have a phenomenon of enough significance to yield insights.[48] They also make more specific suggestions about enriching the factor analysis for such studies, such as including jails and civil commitment facilities in the analysis.[49] Although their heart surely lies in their wish for controlled studies, they recognize that formally designed studies are improbable in the criminal justice setting. Therefore, they urge research that attends in details to "precursory circumstances" and specific program interventions in the hope that the deficiencies of observational studies can be mitigated.[50]

Zimring and Hawkins surely recognize that their rigorous standards of research will challenge the patience of policymakers. Throughout the book, they emphasize the "academic" question of social science proof, but are sensitive to the policymakers' demands and pleas. Writing almost two decades before the current *MPCS* Draft, and somewhat early in the campaign for SGPRA, they note the potential promise of sentencing commissions to control prison population and acknowledge that their study would be depressing for policymakers who hope to build population forecasting into the commission model.[51] Somewhat vaguely, they suggest that the potential for error in forecasting can be somewhat diminished by "paying [some] attention to the social, political, and economic factors that influence imprisonment rates."[52]

Zimring and Hawkins' overall theme is one of a wizened, almost world-weary agnosticism. For them, comparing states with different imprisonment rates to infer differential explanatory causes, such as emphasis on incapacitation, or perceptions of the costs of crime, seems to lie beyond modern empirical powers.[53] Under the social scientist's gaze, the differences in rates do not square well with differences in the overtly proclaimed means or ends of jurisdictions or political leaders. Thus, Zimring and Hawkins explain:

> [I]t is much easier for conservatives to defend a presumptive imprisonment regime that does not exist and for liberals to attack it as if it did than it is for either party to conduct a principled debate within the statistical field of choice that exists for imprisonment policy in the United States.[54]

48. *Id.* at 216–17.
49. *Id.*
50. *Id.* at 177, 216, 218–19.
51. *Id.* at 113.
52. *Id.* at 115.
53. *Id.*
54. *Id.* at 207.

Zimring and Hawkins therefore imply a very serious challenge for anyone hoping to conduct an assessment of the Draft, and they implicitly foretell the risk of political resistance to the *MPCS*, rooted in failure to heed the empirical cautionary tales they offer.

III. THE PARALLEL LESSON OF THE CRIME DROP

Thus, if the Commentary's citation of Zimring and Hawkins suggests that the *MPCS* has implicated the intellectual crisis of modern criminology in understanding the effects of penal policy, an analogy to another, closely related contemporary criminal justice issue might be helpful. It is one linked to the imprisonment question by a similar appearance of Zimring and Hawkins' tragic skepticism. This is the great issue—or mystery—of the 1990s crime drop, part two of my diversion from the *MPCS*. Here I ask the reader's indulgence for a second and more extended digression from the subject of the *MPCS*—before drawing the connection back in some detail.

The downward spike in American crime in the 1990s naturally led criminologists to seek the causes of the downward trend. Criminologists also tried to accommodate the demands and respect of city officials, for whom the rarefied rigors of academic research operated too slowly to fit the officials' needs. The first round of research revealed social scientists' recognition of the limits of empirical evaluation of criminal justice program success.[55]

Even if sentencing reform goals are more contestable and heterogeneous than the goal of a straightforward policing strategy, the problem of lesson-drawing may be very similar, because the challenge of social scientists to help lawmakers draw constructive lessons from the crime drop parallels the potential lessons the social scientists may be asked about sentencing.

As the big crime drop became undeniable, public and political debate began to focus on the potential efficacy of certain new policing and prosecutorial practices, especially those implemented in New York City.[56] But criminologists generally preferred to point to independent, and sometimes mysterious, changes in drug markets or the national economy or to lagged birthrate reductions in certain demographic groups.[57] Certainly, the first round of academic assessments of the crime drop concentrated on

55. *See generally* THE CRIME DROP IN AMERICA (Alfred Blumstein & Joel Wallman eds., 2000) (discussing increased prison populations and the crime drop in America).

56. *See* John E. Eck & Edward R. Maguire, *Have Changes in Policing Reduced Violent Crime?: An Assessment of the Evidence*, *in* THE CRIME DROP IN AMERICA, *supra* note 55, at 207; William Spelman, *The Limited Importance of Prison Expansion*, *in* THE CRIME DROP IN AMERICA, *supra*, at 97.

57. Robert Weisberg, *Meeting Consumer Demand in Modern Criminology*, 4 CRIMINOLOGY & PUB. POL'Y 471, 472 (2005).

such factors, along with mass incarceration and increased police density.[58] These last two factors, while attributable to government design, are nevertheless not very satisfying explanations; the former, which carries many types of costs, may only be an indirect result of other more deliberate programs, while the latter says little about constructive program design or social improvement. Most prominently, econometrician Steven Levitt weighed in with a modestly skeptical take on the numbers.[59] He cast doubt on the relevance of new policing procedures but identified four factors that exhibited some causal potential: (1) the sheer increase in density of police personnel; (2) the demise of the crack-cocaine phenomenon; (3) the increase in incarceration; and, most controversially, (4) the legalization of abortion.[60]

One very specific scholarly encounter illustrates the significance of this statistical debate. A group of criminologists led by Richard Rosenfeld undertook to evaluate the claims of success of the three major urban programs, Boston's *Operation Ceasefire*, New York's *Compstat*, and Richmond's *Project Exile*.[61] Many studies of recent programs had failed to do comparative analysis among sets of cities or sets of programs, nor had they taken account of enough possible determinants within particular cities. In the wake of the apparent success of *Ceasefire*, *Compstat*, and *Exile*, officials in other cities were under great pressure to imitate these programs and were authentically desperate for guidance and reassurance.[62]

Rosenfeld and his fellow criminologists examine the homicide trends in Boston, New York, and Richmond by applying growth-curve analysis to data from the ninety-five largest United States cities, controlling for conditions known to be associated with violent crime rates.[63] They conclude that New York's homicide trend during the 1990s did not differ significantly from that of other large cities and, though Boston exhibited a sharper homicide drop than elsewhere, the small number of incidents in Boston precluded any useful conclusions.[64] On the other hand, they find

58. Steven D. Levitt, *Understanding Why Crime Fell in the 1990s: Four Factors that Explain the Decline and Six that Do Not*, 18 J. ECON. PERSP. 163, 163–64 (2004).

59. *Id.*

60. *Id.*

61. Richard Rosenfeld et al., *Did* Ceasefire, Compstat, *and* Exile *Reduce Homicide?*, 4 CRIMINOLOGY & PUB. POL'Y 419, 419 (2005). *Ceasefire* was Boston's anti-firearm initiative; *Compstat* referred to a set of much-publicized New York police practices, including punishing low-level "quality of life crimes" and using computer technology to quickly allocate police to high-crime neighborhoods; *Exile* involved deploying the high federal sentences for drugs and firearm crimes and the enhanced resources of United States attorneys to attack urban crime. *Id.* at 422–24.

62. *Ceasefire* replications were adopted in a number of cities. Erin Dalton, *Targeted Crime Reduction Efforts in Ten Communities—Lessons for the Project Safe Neighborhoods Initiative*, 50 U.S. ATT'YS BULL. 16, 17 (2002). The federal government has invested millions in attempting widespread replication of *Exile*. *Id.* at 16–25.

63. Rosenfeld et al., *supra* note 61, at 430.

64. *Id.* at 434–35.

that Richmond's homicide reduction was indeed significant, thus suggesting that the tough policies under *Project Exile* may have had some effect.[65] Concerned that criminologists had given police and public officials something of a free ride as they claimed credit for the 1990s crime drop, Rosenfield and his fellow criminologists proffer this paper "as a point of departure for ongoing evaluation" and therefore as an admonition to be skeptical of claims that could not at least pass the modest test they implemented here.[66] In technical terms, they suggest that, as minimal basis for such claims, scholars should employ the kind of tests they use here, including piecewise strategy for estimating linear trends for successive subperiods and hierarchical generalized linear models for comparisons within and across units.[67]

Rosenfeld's group may have hoped to find evidence of the efficacy of these programs, but they expect local officials to be satisfied with, even grateful for, a study which casts doubt on that efficacy, and which finds some positive evidence only in a program that relies on aggressive federal intervention in what is really just a new packaging of old-fashioned punitive enforcement. In fact, the group seems somewhat ambivalent about the message they are delivering to the police chiefs and mayors of the world. Their confidence that they have significantly corrected the deficits of earlier studies is balanced with remarkable concessions about the modesty of this study. They *dis*claim having either the data or the methods to study "the three interventions *per se*," and they admit that their own "inferences cannot be as strong as many program proponents or critics would like, but they should be preferred over the baseless success claims and counterclaims and handful of disparate empirical investigations that have characterized evaluations of the three policing initiatives to date."[68]

In effect, Rosenfeld's group speaks, at one point, of burdens of proof.[69] They say that, at least where their research turns up no statistically significant evidence of incremental homicide reduction due to an intervention, burden of proof shifts to proponents of those interventions.[70]

The Rosenfeld study immediately met vigorous opposition from sociologist Richard Berk.[71] Berk bluntly argues that any such efforts to

65. *Id.* at 436–38.
66. *Id.* at 440.
67. *Id.* at 430.
68. *Id.* at 422.
69. *Id.* at 438.
70. *Id.*
71. Richard A. Berk, *Knowing When to Fold 'Em: An Essay on Evaluating the Impact of* Ceasefire, Compstat, *and* Exile, 4 CRIMINOLOGY & PUB. POL'Y 451, 451 (2005).

Although in this paper I do not discuss the general political science of risk analysis and cost-benefit analysis, I note that my mention of burden of proof implicates the vast subject of these principles in regulatory policy. *See generally* Cass R. Sunstein, LAWS OF FEAR (2005) (reviewing the various forms of burden of proof placed on the use of scientific studies in guiding risk-

draw conclusions from non-designed, observational studies are futile.[72] In his view, even with a discrete variable like homicide to study, there are simply too many other variables to control, and even more fundamentally, it is too hard to determine what a stringent causal model can accomplish.[73] To illustrate, Berk lays out what he considers minimal criteria for testing the effect of the New York-style order-maintenance policing on homicide rates.[74] He deliberately sets a simple model with a single intervention—number of police officers assigned to order-maintenance policing—and a single response—homicide rate—that is a linear function of the intervention variable and a perturbation.[75] He then dauntingly lays out what he argues as the absolutely necessary conditions for this experiment—that the perturbation be generated independently of the value or the level of the intervention; that all units in the study be affected in the same way by the intervention; and that the intervention value assigned to some other unit not affect how the subject unit responds.[76] In addition to these technical requirements, Berk emphasizes the need for considerable institutional knowledge to ensure that the intervention is truly manipulable—i.e., that the number of officers assigned to order-maintenance policing is truly subject to direct control.[77]

These factors are a mere sketch of the methodological demands Berk insists must apply to observational studies, and he ultimately concludes that available statistical tools under the current state of knowledge are simply inadequate to the task assumed by the Rosenfeld group.[78] Berk admonishes his social science colleagues that the only good alternatives are carefully pre-designed, true randomized experiments.[79] At most, he shows some tolerance for "quasi-experiments," for example using "generalized regression-discontinuity design" relying on a "known deterministic

---

regulation under versions of the so-called "Precautionary Principle"). Academics have generally been loath to apply the principles of risk-regulation to criminal law because these principles are difficult to adapt to the retributive basis for the criminal sanction. *See* Kenneth Mann, *Punitive Civil Sanctions: The Middleground Between Criminal and Civil Law*, 101 YALE L. J. 1795, 1863 (1992) (urging caution in treating criminal law as a regulatory tool); Nancy Frank, *From Criminal to Civil Penalties*, 30 SOC. PROBS. 532 (1983) (framing criminal law as risk-regulation raises concerns about strict liability crimes). Nevertheless, some have argued that criminal justice policy is always at least implicitly involved in risk regulation and that risk analysis and cost-benefit analysis doctrines are an unavoidable component of criminal justice policy. Mariano-Florentino Cuéllar, *The Institutional Logic of Preventive Crime* (Stan. Pub. Law Working Paper, No. 1272235, 2008), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1272235.

72. Berk, *supra* note 71, at 451.
73. *Id.* at 452–56.
74. *Id.*
75. *Id.* at 452.
76. *Id.* at 452–56.
77. *Id.* at 454.
78. *Id.* at 455, 58.
79. *Id.* at 459–61.

mechanism," or, more weakly, estimating treatment effects without conventional causal models but treating confounders as nuisance variables.[80]

In surrebuttal, Rosenfeld's group argues that they needed to address only criterion of consistency, not causality.[81] They insist they have drawn no causal inferences, and that even where they find consistency between their statistical analyses and the officials' claims, it could have been produced by other factors.[82] Thus, in effect, "the answers to the question[s] posed in [their] title . . . are, respectively, 'hard to tell,' 'probably not,' and 'maybe.'"[83]

This research was published in the context of a lively debate among criminologists about the debt they owe to policymakers,[84] a debate that can be situated in a larger set of cultural questions about the role of social science in American policymaking. Do academics impose overly rigorous and complex standards on government because they have an intellectual investment in grandiosity to disdain concrete solutions to evident social problems? Are academics invested in the persistence of high crime rates because the idea of American exceptionalism is itself an endlessly fertile subject for scholarly inquiry,[85] or do academics suffer a political bias, a leftist refusal to credit American virtue to American power and know-how? Or is the problem something much more substantive or technical—a desire by social scientists to correct Americans' failure to appreciate that in a criminal justice system already so severe, marginal changes are futile? Or, to return to the work of elite empiricists, might academics on the whole be right in adopting an attitude of tragic skepticism, as evidenced by Zimring's new book on the subject?[86]

In the *Great American Crime Decline*, Zimring offers the following lessons: The crime drop was anomalously large enough to be historically significant, and it was "real" in the sense that it was not an artifact of changes in reporting.[87] Further, the crime drop took place without any change in the social fabric, "the population composition, economy, or ecology" of the city.[88] Thus, policy change may have played a major causal

80. *Id.*

81. Richard Rosenfeld et al., *The Straw Man Bluff: Reply to Berk*, 4 CRIMINOLOGY & PUB. POL'Y 467, 467–68 (2005).

82. *Id.* at 468.

83. *Id*. at 469.

84. Weisberg, *supra* note 57, at 477.

85. Joan Petersilia, *Policy Relevance and the Future of Criminology—The American Society of Criminology: 1990 Presidential Address*, 29 CRIMINOLOGY 1, 1–3 (1991); Robert Weisberg, *Values, Violence, and the Second Amendment: American Character, Constitutionalism, and Crime*, 39 HOUS. L. REV. 1, 9 (2002).

86. *See* FRANKLIN E. ZIMRING, THE GREAT AMERICAN CRIME DECLINE 195–207 (2007).

87. *Id.* at 196.

88. *Id.* at 143.

role.[89] But in terms of both social economic factors and policy changes, the drop was clearly the result of a multiplicity of causes that cannot be disentangled with our current statistical tools. For example, says Zimring, Levitt's inference about the role of abortion may be flawed because it does not account for the much wider sets of contexts in which the "wantedness" of children might be relevant to their future crime.[90] As for policing practices, New York's crime reduction was indeed anomalously large enough to suggest that *Comstat*-style policing probably played some role, but how much cannot be determined once we acknowledge the many components of police logistics.[91] Thus, it remains a vexing and disappointing task to attempt to forecast the effect of any policy intervention because of the underlying problem of predicting or retroactively inferring anything at all. Indeed, Zimring notes that many of the factors that probably contributed to the mid-late 1990s drop were present and quite visible to the best criminologists a few years before the drop, but none predicted it.[92] And finally, for Zimring, there is the sobering and chastening significance of Canada. Canada exhibited none of the policy changes undertaken in the United States and none of the social factors, save for one (the drop in the proportion of youth in the population), yet it experienced virtually the same crime drop in the same period.[93] Canada is a benign but challenging reminder of how little we know about the world.[94]

IV. SOME IMPLICATIONS OF TRAGIC SKEPTICISM FOR SENTENCING REFORM

If the elites in the world of criminology lament how invisible the world of causation is, where does that leave law reformers called on to enact reforms like the *MPCS*, and academics who might be asked to opine on its chances of success? As I have suggested, in terms of *policing* practices, one possible inference from criminology is a very specific burden or standard of proof. By parallelism, perhaps the goal of lawyers and social scientists in assisting in sentencing reform should be to identify, encourage, and even laud programs that show *any* prospect of effectiveness and exhibit no measurable downside because that is the standard of proof under which police officials legitimately need to act.[95] Perhaps the governing principle should be a kind of Hippocratic oath: it is worth taking

89. *Id.* at 201.
90. *Id.* at 85–86.
91. *Id.* at 150–51.
92. *Id.* at 21–24.
93. *Id.* at 132–34.
94. *Id.* at 107–34.
95. *See generally* Weisberg, *supra* note 57 (discussing the studies performed by Rosenfeld and his fellow criminologists).

up a policy that does no measurable and little conceivable harm, and that is within the limits of econometric prediction, might be rebuttably presumed to do some good.[96]

So should such a relaxed burden of proof apply to sentencing? Sentencing and corrections may be a harder and bigger matter than short-term crime control, depending upon the level of generality at which a program is evaluated. Other things being equal, taking a chance on street-level police practices is cheaper and more reversible than undertaking a more structural change in penal policy (although the latter category is far too heterogeneous for this generalization to take us very far). But should the burden of proof be different or should it be the same modest one, applied in a different context? If the burden of proof should be different, the role of social science in sentencing reform would be to admonish lawmakers about which programs—predicted or claimed—*are at least not inconsistent with the data*. Social science's role would also encompass discerning which observed schemes are obviously doing worse than those in otherwise roughly similar jurisdictions and urging lawmakers to be reluctant to rely on forecasting which is not derived from designed or quasi-experimental studies.

Before proceeding further about the social scientist's relationship with lawmakers, a note about branches and levels of government is in order. This is a matter somewhat important in the policing studies, but hugely important in a sentencing reform like the *MPCS*, where questions of reallocation of power among legislators, judges, and quasi-executive branch officials (i.e., sentencing commissioners) play a major role. Disentangling the various factors that affect the performance of a sentencing system is of course the heart of the problem, but it is an especially interesting challenge when the possible causal forces are different government actors or branches of government, or different mixes thereof, and hence different potential audiences for research. If we return to the inferences made by Rosenfeld's group about the programs in Boston, New York, and Richmond, [97] we see that small *police*-authorized practices do not work much to reduce homicide, whereas the one possible success story is driven at a higher level—legislative or high-level executive. But if we then compare crime-stopping to sentencing, where the latter is at least initially implemented at a higher or more visibly political level, the difference in the audience for research is important.

It may make more sense to offer a more generous burden of proof to police officials than to legislators or high-level executives. Legislators, above all, find self-aggrandizing policy demagoguery all too attractive and the externalization of costs all too tempting. So perhaps even if researchers are inclined to encourage local police in sending their research messages

96. *Id.* at 474.

97. Rosenfeld et al., *supra* note 81, at 424.

they should, in the face of legislative demand for research, adopt a more dourly debunking stance. At the same time, in helping to both evaluate and promote the *MPCS*, academics will face political and moral resistance from various lawmakers about the proper goals of reform, and even more confounded resistance when it comes to sufficiency of proof of outcomes relevant to those goals.

The problem of multiple causation will always haunt such work. Indeed, in their crime-drop paper, Rosenfeld's group concedes a limit of their study in that each of the three campaigns under review contains sub-components whose differential effects cannot readily be tested. Might better sorting then improve the cost-benefit analysis? Even if so, some research should address the clusterings of components we see in these cities, although that research may have to come from political scientists rather than econometricians.

But if de-clustering can be done in the context of the kinds of sentencing changes proposed by the *MPCS*, what costs and benefits might be noted? Can criminologists help prison officials perform productivity analyses, to deploy financial accountants to help cities do cost-attribution for sub-components? And can political science or sociology or other disciplines help us determine the risks of not-so-material costs? Do academics have the kind of intellectual imagination to help us identify truly risk-free campaigns? Of course these amorphous inquiries cannot produce precise answers, but they may be useful. Perhaps they can reassure policymakers that some experiments will be at least harmless and might yield somewhat measurable answers to the question of how to improve sentencing systems.

V. TENTATIVE SPECULATIONS ABOUT MEASURING SENTENCING REFORM SUCCESS

With those admonitions in mind, what might happen if we gently extended the *MPCS* Draft's commitment to empirical evaluation a few suggestive realms beyond the question of whether determinacy necessarily raises incarceration rates? The most obvious result might be even more modesty about what can be done in facing the unbelievable mazes of multiple and unknowable causations in sentencing schemes. But another result might be that we could imagine some lines of micro-research that will someday bear fruit, especially if we relax burdens of proof in the face of public exigency. Below I sketch some possibilities and some rough empirical observations about how we can thereby go through the "stages of grief" of tragic skepticism and emerge somewhat more optimistic about our ability to generate modestly reliable measures of evaluation.

A. *Incarceration Rates Redux*

The *MPCS* does not directly address the question of prison overcrowding and inhumane conditions as a problem in itself. Nevertheless, its deep commitment to cost-efficacy, its call for mechanisms to predict the incarceration effects of legal changes, and its clear concern about unnecessary prison expansion all obviously imply a concern for unjustified increases in incarceration rates. First, a bit more reflection on the incarceration rate is in order. To be blunt, the perceived hyper-incarceration crisis in America has obviously been the major social and economic factor provoking calls for reform, especially where state budgets have hit the wall on prison expenditures, or where federal court rulings on the constitutionality of prison conditions leave states no choice. To put things as delicately as possible, no proponent of sentencing reform can deny that it is important to consider the implications of the reform for incarceration rates. But a more modest predicate point requires emphasis. Regardless of the goal of any sentencing reform in terms of incarceration rates, the possibility of even measuring success of some ends or means will be quixotic if empirical science cannot tease out causality with any confidence. This is the juncture at which Zimring and Hawkins' book obviously demands that we exercise some harsh realism on this subject.[98]

However, to return to the cautious claims in the Draft about the effects of parole release mechanisms, we might say that the adoption of SGPRA systems might well tend to have a downward pressure effect on incarceration rates.[99] If we look at existing schemes in some states, SGPRA was indeed associated with an actual reduction rate in a few states in the late 1990s and a slowing of the rate of increase in others that had exhibited especially high growth rates.[100] However, Professor Reitz himself is very

98. *See generally* ZIMRING & HAWKINS, *supra* note 27 (finding no clear links between incarceration rate and such factors as crime rate, public attitudes toward imprisonment, and national economic trends).

For example, experience has hardly validated the utterly plausible projection that a strong new "Three Strikes" law would greatly increase the incarceration rate. *See* BRIAN BROWN & GREG JOLIVETTE, CAL. LEGISLATIVE ANALYST'S OFFICE, A PRIMER: THREE STRIKES: THE IMPACT AFTER MORE THAN A DECADE 15–16 (2005) (noting 1994 prediction that the law would add 100,000 inmates to state prison in a decade proved wrong, in part because of prosecutorial discretion not to exercise the "Three Strikes" option). One speculation might be that the true effect of laws like "Three Strikes" and "Truth in Sentencing" may not be noticeable for several years, materializing only after the extended time served exceeds that which offenders would have served in the absence of such laws.

99. Thomas B. Marvell, *Sentencing Guidelines and Prison Population Growth*, 85 J. CRIM. L. & CRIMINOLOGY 696, 707 (1995); Sean Nicholson-Crotty, *The Impact of Sentencing Guidelines on State-Level Sanctions: An Analysis Over Time*, 50 CRIME & DELINQ. 395, 396 (2004); Jon Sorensen & Don Stemen, *The Effect of State Sentencing Policies on Incarceration Rates*, 48 CRIME & DELINQ. 456, 469 (2002).

100. *See* Reitz, *supra* note 18, at 1802.

agnostic about the causal direction of this inference, particularly in terms of the proportional significance of the choice of sentencing system in the mix of factors influencing incarceration rates.[101]

A useful and suggestive way to ponder future possible inferences is a snapshot look at the Bureau of Justice Statistics (BJS) annual report on incarceration figures, most recently *Prisoners in 2007*.[102] Examination of the 2007 report reveals that within the last decade—an arbitrary time unit, but a decent stretch of time for a realistic assessment of change—the United States imprisonment rate increased at an average rate of about 2% per year.[103] That means, overall, a 15% size increase against a 6.4% population increase.[104] In 2007, the jump was 1.8%,[105] so in some crude national sense, the country is "doing better" if we assume reduction in the incarceration rate is an uncontroversial good thing. Hewing closer to our modest burden of proof, at least we see no sign that the rate of increase in the prison population is itself increasing. Moreover, the states as a group grew at half the rate of the federal system, and given the bizarrely unusual legal situation of the federal system, that means that the real comparison is in the state figures—between 1.7% for the years 2000–2006 and 1.5% for 2007.[106]

Although these aggregated figures tell us nothing at all about the success of the great variety of sentencing schemes, I offer them to give a flavor of the statistics available. Naturally, we should look at individual states, and if we do so, we see that in terms of increase in incarceration rates in the last almost-decade:

> —The "worst" states, in descending order from highest rate on down, were: Kentucky, West Virginia, Alaska, Indiana, Florida, Virginia, Alabama, Louisiana, Colorado, and North Dakota.[107]
>
> —The "best" states, in ascending order from lowest on up were: New York, Texas, New Jersey, Delaware, Maryland, Illinois, Oklahoma, Utah, Nevada, South Carolina, and Massachusetts, with California in twelfth place.[108]
>
> —In terms of regions, the "worst" to "best" order was: West,

101. *Id.* at 1794–97.

102. HEATHER C. WEST & WILLIAM J. SABOL, U.S. DEP'T OF JUSTICE, BUREAU OF JUSTICE STATISTICS BULLETIN: PRISONERS IN 2007 (2008), *available at* http://ojp.usdoj.gov/bjs/pub/pdf/p07.pdf.

103. *Id.* at 1.

104. *Id.*

105. *Id.*

106. *Id.* at 2.

107. *Id.* at 5.

108. *Id.*

South, Northwest, and Northeast,[109] but a look at the states listed above suggests that any regional differences are crude.

Watching which states move up and down on these lists may prove helpful, and a fair inference might be that the states in the second list are not currently doing anything especially counterproductive in terms of sentencing, while those in the first list should be in a "watch carefully" category. But, the crudity of these numbers and the vast mélange of variables that affect them obviously require us to withhold judgment, especially when we can only be very tepid in even suggesting that decarceration is a clear social good.

A far less controversial goal of any sentencing scheme is to avoid or reduce the degree of overcrowding. As noted above, the MPCS does not directly consider the question of prison crowding and inhumane conditions, but its obvious worries about unwarranted increases in prison expansion and its calls for cost-efficacy measures are surely consistent with a concern over crowding issues. The numbers here are complex and arbitrary. States usually measure prison capacity by several parameters—the inmates assigned by rated capacity (by governmental directive), the operational capacity (by managerial judgment), and design capacity (by the architectural planners).[110] Then, both BJS and the states select whatever is the highest and lowest among these three measures for each state. From these measures, the BJS rates state prison systems by the ratio of actual prisoners to the lowest and highest capacities.[111] On this score, the BJS has only supplied aggregate national figures, but the story is not too dismal: In 1995 the state prison population as a percent of highest and lowest capacity was 114%, and 125%, respectively, and in 2007 the percent of highest and lowest capacity was 96% and 113%, respectively.[112] As for the distribution of these mild improvements, the BJS tells us that nineteen states in 2007 were above their highest, but notably, that the federal system was an egregious 36% above capacity.[113] Needless to say, even if the goal of reducing overcrowding is uncontroversial, these numbers are too broad and rough to tell us much about success, especially when another possible, but highly immeasurable test of this criterion looms—the possibility of federal court injunctive control of prisons when the overcrowding proves unconstitutional. Tracking federal litigation is not a very good measure of anything, but for at least one empirical datum we might note that the existence of a hugely interventionist injunction in California is not exactly a compliment to its penal system.[114]

109. *Id.* at 18.
110. *Id.* at 6.
111. *Id.*
112. *Id.* at 7.
113. *Id.*
114. On the state of the California litigation, see Robert Weisberg, *How Sentencing*

B. *The Crime-Rate Effects*

More ambitiously, if we treat sentencing and corrections as parts of a holistic criminal justice system, and if we assume that the purpose of criminal justice is to reduce crime, then the crime rates "produced" by the sentencing system would be the best success measure. The *MPCS* Draft stays well clear of any commitment to reduce crime rates through sentencing reform, but the drafters surely would consider time-correlated reductions in crime a potential validation of any change in sentencing law, and a time-correlated *increase* in crime a very bad sign of the social utility of the *MPCS*. But any hope of drawing sound *causal* inferences in this area widely outpaces our ability. For one thing, as discussed above, there is the uncertain condition of our econometric arts in explaining changes in crime rates, evidenced by the great division among authority in explaining the great crime rate decline of the 1990s.[115] Second, to the extent that changes in sentencing affect crime rates, a key mediator of those changes (although far from the only one) will be the marginal incapacitative effects of any alteration or change in the size or demographics of the prison population.

The *MPCS* model's premises about smarter evidence-based incarceration decisions suggest that more rationally selective incapacitation can reduce crime and possibly also reduce incarceration. But the deeper problem is that the more intuitively plausible notion—that *increasing* incarceration and therefore incapacitation reduces crime—is itself an empirically uncertain matter.[116] To return to the 1990s crime drop, prolific recent research has produced mixed and equivocal findings and speculations about the up-spiking incarceration rate's role in the down-spiking crime rate.[117]

Commissions Turned Out To Be a Good Idea, 12 BERKELEY J. CRIM. L. 179, 222–29 (2007).

115. *See supra* Part III.

116. Empirical conclusions about the relationship between prison population increases and crime rates are difficult to draw. As pointed out by Judith Greene and Vincent Schiraldi:

> [D]uring the 1990's, Texas added more prisoners to its prison system (98,081) than New York's entire prison population (73,233) by some 24,848 prisoners. . . . While Texas had the fastest growing prison system in the country during the 1990's, New York had the third slowest growing prison population in the U.S. Over all, during the 1990's, Texas added five times as many prisoners as New York did (18,001). Yet from 1990 to 1998, the decline in New York's crime rate was 26 percent *greater* than the drop in crime in Texas. Texas' 1999 incarceration rate (1,014 per 100,000) was 77 percent higher than New York's (574 per 100,000), yet Texas' 1998 crime rate (5,111 per 100,000) was 42 percent higher than New York's (3,588 per 100,000). In 1998, Texas' murder rate was 25 percent higher than New York State's rate.

JUDITH GREENE & VINCENT SCHIRALDI, JUST. POL'Y INST., CUTTING CORRECTLY: NEW PRISON POLICIES FOR TIMES OF FISCAL CRISIS 3 (2002).

117. For an excellent review of this research, see RYAN S. KING ET AL., THE SENTENCING

C. *The Mysteries of Recidivism*

But of course, no one argues that the sentencing system is the sole, or even necessarily the most important, component of criminal justice in preventing crime. Rather, we should identify the more particularized role that sentencing can play in reducing the crime rates of those *released* from prisons. And to narrow things more, let us eliminate the prospect of legislated prison terms as a mechanism of general deterrence. That means we are left with the phenomenon of imprisonment (and parole) in terms of specific deterrence, incapacitation, and rehabilitation. Thus, we should focus on the more specific goal of reducing crime by recidivists. But this exercise only underscores the preliminary difficulty of determining what set of crimes is in fact recidivist, because we would first have to define the types and numbers of prior offenses, convictions, and incarcerations that help define the recidivism category.

A first cut at this issue is to look at the published BJS statistics on comparative recidivism rates among the states according to the criterion used by the states and BJS—which turns out to be the number of judgments the states denominate as "parole violations" leading to re-imprisonment. The 2007 BJS report allows us a quick look at these numbers, and amid their complexity and the diversity of systemic contexts for them, one very non-subtle thing stands out: California's parole violation rate is the outlier among outliers. It is more than three times that of Texas, and is anomalous no matter how one adjusts for general population or even prison population.[118]

However, it takes little analysis to realize how weak a basis for comparative assessment of recidivism these numbers provide. Even, or especially, with California, of course, the question of how many of its returnees should count as "recidivists" is a very complex matter of definition. The great majority of California returnees are people who were paroled under the formulaic mandatory parole system, not the old-fashioned discretionary parole. Under that scheme, many inmates return to prison for minor parole violations that are not statutory crimes, while a great number commit new crimes that, for prosecutorial and administrative convenience, are classified as parole violations.[119]

PROJECT, INCARCERATION AND CRIME: A COMPLEX RELATIONSHIP 2–6 (2005), *available at* http://www.sentencingproject.org/Admin/Documents/publications/inc_iandc_complex.pdf. The authors note serious deficiencies in the statistical record needed to clarify the incarceration-crime relationship, including inconsistencies of definitions and measures across jurisdictions and differing time frames for analysis. *Id.* at 2–4. But, they also posit some reasons why marginal increases in incarceration might not reduce crime. For example, we would at least expect diminishing returns on crime reduction at some point if the highest rate offenders are captured first. *Id.* at 6. Also, especially with regard to drug crimes, incapacitation of one offender may simply invite a new substitute offender to replace him in a fixed-sum market of drug-dealing opportunities. *Id.* at 6.

118. WEST & SABOL, *supra* note 102, at 17.

119. JOAN PETERSILIA, CAL. POL'Y RESEARCH CTR., UNDERSTANDING CALIFORNIA CORRECTIONS

To appreciate the complexity of the recidivism measure, one can turn to James A. Wilson's important recent study on a decade of data from the state of Tennessee.[120] Wilson demonstrates in this one jurisdiction how arbitrary and contingent the definition of a parole violation is, and not just in terms of whether it is itself a new crime.[121] The official criteria for revocability changed during the study period, as did the degree of tolerance exercised by parole officers in applying those criteria, and, concomitantly, the degree of supervision they exercise on parolees.[122] When Wilson examined changes in the parole release and revocation patterns, he found no clear correlations with any demographic criteria or severity of crime or conviction.[123] Rather, the main predictor of the number of parole revocations turned out to be the original number of parole releases.[124] And, Wilson tentatively concludes that *a* major, if not *the* major, predictor of the parole release rate is the population pressure on the prison system—the ratio of those sitting in jail awaiting assignment to prison to the number of those currently in prison.[125] And then there is the paradox—perhaps not a very bizarre one, on reflection—that increases in parole releases lead to increases in parole revocations even where the increase in parole releases was originally implemented in order to relieve population pressure.[126]

In a response essay, Sheila Royo Maxwell further unravels the heterogeneous category of recidivism, as she enumerates the variety of legal and political vectors that determine revocation.[127] Greater supervision may increase revocations because of more stringent discipline of parolees, but relaxed supervision may have the same effect to the extent that foregone supervision might have prevented revocations.[128] In a related essay, Joel Wallman notes, with strong implications for California, that the complexity of analysis would increase further if we tried to sort out these numbers in terms of whether the original release was mandatory or discretionary under state law.[129]

But we may consider one more possible set of measurements of recidivism having to do with ratios of various inputs and outputs to prison systems.

71–73 (2006), *available at* http://www.ucop.edu/cprc/documents/understand_ca_corrections.pdf.

120. James A. Wilson, *Bad Behavior or Bad Policy? An Examination of Tennessee Release Cohorts, 1993–2001*, 4 CRIMINOLOGY & PUB. POL'Y 485, 485 (2005).

121. *Id.* at 494–95.

122. *Id.* at 504.

123. *Id.*

124. *Id.* at 505.

125. *Id.* at 507.

126. *Id.* at 509.

127. Shelia Royo Maxwell, *Rethinking the Broad Sweep of Recidivism: A Task for Evaluators*, 4 CRIMINOLOGY & PUB. POL'Y 519, 519 (2005).

128. *See* Anne Morrison Piehl & Stefan F. LoBuglio, *Does Supervision Matter?*, *in* PRISONER REENTRY AND CRIME IN AMERICA 105, 108–09 (Jeremy Travis & Christy Visher eds., 2005).

129. Joel Wallman, *Unpacking Recidivism*, 4 CRIMINOLOGY & PUB. POL'Y 479, 481–82 (2005).

The first possible measure is the ratio of new court commitments to parole violators. Crudely put, this can be seen as a comparison between the crime prevention success of the criminal justice or law enforcement system generally and the performance of the prisons. In terms of the relative performance of the prison system, the higher the ratio, the better the performance.

If we semi-randomly take eight states from the 2007 BJS report, the numbers for the eight states are: Pennsylvania—1.77; Michigan—2.13; Ohio—6.99; Alabama—6.44; Texas—1.70; California—0.51, Georgia—1.37; New York—1.69. Ohio's "good" outlier status may be due to very restrictive parole rights there, but there is nothing subtle, yet again, about California's opposite outlier status.

The second possible measure is the ratio of prisoner *releases* to parole violators returned. This might seem to be closer to an internal review of the recidivism-reducing powers of the prison, one at least worth observing over time. Here, the numbers for the same states are: Pennsylvania—2.84; Michigan—3.59; Ohio—7.64; Alabama—7.72; Texas—2.79; California—1.47, Georgia—2.15; New York—2.78. Again, "congratulations" to California, although no glory goes to Georgia.

The third possible measure is the ratio of conditional releases to parole violators returned. This might show how well the system is choosing whom to release or how well it is supervising those on whom it took a chance. Here, the numbers are: Pennsylvania—1.95; Michigan—3.04; Ohio—3.99; Alabama—4.82; Texas—2.12; California—1.44, Georgia—0.39; New York—2.44. Georgia clearly merits some scrutiny.

Finally, the fourth possible measure is the ratio of *unconditional* releases from prison to parole violators returned. Might the numbers on this scale show that some states are abject failures? Or that conditional release may be a good idea after all? Here, the numbers are: Pennsylvania—0.54; Michigan—0.42; Ohio—3.61; Alabama—2.79; Texas—0.53; California—0.02; Georgia—1.75; New York—0.31. The extreme crudity of these numbers means that they may tell us nothing across states. California's numbers simply tell us that there is little unconditional release. But again, tracking a state over time may prove helpful.

Either way, recidivism is a double target as a success measure: It invokes our sense of tragedy because it is so unknowable, yet it is tragic because if we did learn something measurable, it might be that prison does not rehabilitate in the sense of making people less crime-prone then they were when they entered. The crime-inducing effects of prison may be built into our society. But prisons may at least differ in the degree to which they mitigate their own damage.

D. *Disparity Reduction*

An unequivocal goal of the *MPCS* is to reduce unwarranted disparity in sentencing outcomes, especially when that disparity has racial or ethnic implications.[130] Since the turn of the new century, the percentage of American prisoners who are African-American has declined, and the Black-White ratio has declined from 8:1 to 7:1.[131] This is not the place to take on the infinitely complex issue of the role and manifestation of race in the criminal justice system, but it is worth noting that for many, this reduction, however small, is an incontestable good. Still more modestly, it is a (roughly) measurable matter and will support claims of success or failure by many chosen standards rooted in various political philosophies.

On the other hand, from 2000 to 2007, the percentage of prisoners who are Latino increased from 18.2% to 19.7%, but the rate of imprisonment per population for Latinos went down from 1,419 per 100,000 to 1,259.[132] So the increase in the percentage of prisoners who were Latino seems to be the effect of the increase of the percentage of Latinos in the general population.

Of course, disparity-reducing effects can be measured in part by whatever analytic schemes are normally used to tease out impermissible factors, especially race and ethnicity, from those factors that could legitimately affect sentencing. In important new research, Professor John Pfaff has inferred that both presumptive and voluntary guidelines systems have decreased sentencing disparities for those convicted of a particular offense and also have reduced the statistically measurable role of race or gender in sentencing.[133] Professor Pfaff carefully acknowledges that by this definition, reduction in disparity is not an uncontestable virtue because it may allow less individuation for offender characteristics than normatively preferable.[134] But if we accept the widespread lament that the pre-guidelines systems exhibited irrational or prejudicial disparity, the changes Professor Pfaff observes do not look very normatively contestable after all.

---

130. MODEL PENAL CODE: SENTENCING REPORT § (b)(iii)–(vi) (Preliminary Draft No. 6, 2008).

131. WEST & SABOL, *supra* note 102, at 4.

132. *Id.*

133. John F. Pfaff, *The Vitality of Voluntary Guidelines in the Wake of* Blakely v. Washington*: An Empirical Assessment*, 19 FED. SENT'G REP. 202, 202–03 (2007); John F. Pfaff, *The Continued Vitality of Structured Sentencing Following* Blakely: *The Effectiveness of Voluntary Guidelines*, 54 UCLA L. REV. 235, 235 (2006) [hereinafter Pfaff, *Continued Vitality*].

134. Pfaff, *Continued Vitality*, *supra* note 133, at 281. Perhaps too carefully, he allows for the possibility that gender neutrality also may be a contestable virtue, since sentencers might want to mitigate to parents who are especially crucial to their children's care and who, he says, may be disproportionately female. *Id.* at 281–82.

E. *Showing Us the Money*

Even if we settled on normative goals for our sentencing system, we would have to recognize that sentencing is just one component of criminal justice and even a smaller component of government. Thus, the question of resource allocation is unavoidable, and therefore equally unavoidable are complex measures of *efficiency* in reducing crime or whatever the goal may be. The *MPCS*, especially through its promotion of the commission model, is clearly committed to cost-efficacy as a component of reform, and in any event, political resistance to adoption of the *MPCS* is likely to raise cost issues—legitimately or not. Finessing for the moment the impossible question of what success outcome we are trying to achieve as cheaply as possible, the cost-efficacy of a system, in turn, might be captured by any number of possible metrics—i.e., a measurement of reduced cost per prisoner or some other unit, or some measure of the cost per unit of reduction in recidivism, or some measure of the relationship between prison expenditures and the crime rate (although we may expect a notable lag between cause and effect in this linkage), or some measure of the relationship between overall prison expenditures and the recidivism rate (where we might expect less of a lag, since any effect of the prison experience on recidivism is likely to show up right after release). These measures are all individually complex and in turn, any one of them will bear only an elusive relationship to incarceration rates. More modest measures would include some combination of reduced prison costs as a portion of the state budget and the absence of any disturbing increase in the crime rate or recidivism rate.[135]

135. A good example of the difficulty of success measurement is privatization research, which aims to determine whether private prison contractors can incarcerate more efficiently than the public system. *See* Kathryne TafollaYoung, Note, *The Privatization of California Correctional Facilities: A Population-Based Approach*, 18 STAN. L. & POL'Y REV. 438, 438–39 (2007). Analyzing the relative efficacy of private and public systems is daunting even if we have good data, because determining the right frames to sort the data is such a contestable enterprise. *Id.* at 439. For example, in the world of privatization, we often distinguish "avoidable" costs, those that can be imposed on the contractor, from "unavoidable" ones, those that the state must continue to bear. However, drawing the boundary between the two is a complex matter that will obviously influence the measure of the contractors' success. *Id.* at 439. In addition, if we look at efficacy in terms of cost per prisoner housed, we would have to control for the varying categories of prisoners housed in private facilities. *Id.* at 440. If the contactors get the less dangerous inmates, the apparent success advantages will suffer from severe sampling biases. *Id.* at 447–54. If they get older prisoners, their security costs will go down but their medical costs will go up, with the relative effects being wildly unpredictable. *Id.* Further, cross-vendor and cross-state comparisons are difficult because some contractors subject themselves to the industry-standard criteria of the American Correctional Association but others do not. *Id.* at 464–65. Moreover, if vendors follow the common public model of using remote rural locations for facilities, the lower land and related costs might seem to improve efficacy, but these savings will be at the cost of greater access to reentry programs and social services in more urban areas. *Id.* at 465–66.

The vast question of measuring cost becomes even vaster when we set it in the context of modern styles of prison financing—and such styles have gotten more creative precisely when, and as a result of, the crowding crises now facing the United States.[136] Until the 1980s, most prison construction was financed either immediately though draws on general revenue or through general revenue bonds on a longer term.[137] These general revenue bonds were highly marketable, because they were backed by the full faith and credit of the state.[138] But since then, the states have mostly (and, in the last decade, entirely) financed prison space through lease or revenue bonds which are payable out of specified assets or income streams and that are sometimes defeasible by the legislature.[139] This modern form of "back-door" financing is an attractive political choice for legislatures, but regardless of its political or economic feasibility, it makes the whole notion of measuring cost-effectiveness even more daunting. This is because the actual or threatened failure of such bonds can have metastatic effects on the creditworthiness of the state in general, so costing out the liability the state takes on is highly speculative. Indeed, the financing has now become triangulated between the state, the creditor entities, and third-party private entities that build or help operate prisons.[140] Under many creative financing arrangements, third-party vendors play a big and sometimes conflicted role in the state's efforts to get financing. And sometimes the financial interests of the third-party lead the state to build a bigger facility than it otherwise needs, often with the expectation that the federal correctional system will rent back some of the space. Whether the result of all this leads to a melt-downable speculative bubble is an intriguing question these days, but in the short run it is clear that this new era of financing makes the cost-efficacy of prisons a great mystery.

In any event, those who promote the *MPCS* to state legislatures will clearly need to be prepared to discuss these financial factors, and equally important, future critiques of adopted *MPCS* schemes will surely include concerns about whether its deontological and non-financial utilitarian benefits are worth the cost.

F. *Some Political or Civic "Measures"*

Finally, consider some "soft factors" that, in an ironic but helpful way, are easy to measure empirically. These are "measures" of civic satisfaction that manifest themselves in superficial but therefore transparent ways.

---

136. Kevin Pranis, *Doing Borrowed Time: The High Cost of Back-Door Prison Finance*, PRISON LEGAL NEWS, Nov. 2008, at 1, 1. The following discussion above is drawn from this superb article.

137. *Id.*

138. *Id.*

139. *Id.* at 1, 3.

140. *Id.* at 3–4.

One is "transparency"—a vague, contestable term, but one with much political salience. In the category of transparency, there is the much-invoked goal of reform called "Truth-in-Sentencing" (TIS).[141] The *MPCS* does not invoke this popular shibboleth, and indeed its mild compromise on parole release would suggest that the *MPCS* would not fully satisfy the vocal adherents of TIS. But at some point, appraisers of the *MPCS* might well have to determine whether the *MPCS* model somewhat disserves the goal of transparency, and *MPCS* promoters need to be prepared to talk in transparency terms.

A number of states now proclaim their success with respect to transparency through the very clear evidence of the TIS gap narrowing because of the elimination of discretionary parole or other under-the-radar release mechanisms.[142] Of course, this is very disputable as both a goal and an indicator of success. On the one hand, TIS enhances general deterrence if certain assumptions are made about the extent to which potential offenders rely on fairly precise warnings of likely incarceration time. It may also help officials who manage prison population flows and costs. But there remains in our jurisprudence a very strong legacy of belief in the rehabilitative value of indeterminate sentencing. For that goal, TIS may be irrelevant. And still worse, if the old indeterminacy model was at least somewhat right (and the *MPCS* is somewhat wrong) about the tempting incentive of earlier release for inmates, TIS might even be an obstacle. Nevertheless, if we accept factual transparency, narrowly viewed, as an inherently good thing, then in that regard sentencing systems have an easy measure of success.

Perhaps more constructively, though also very narrowly, another version of transparency would be captured by the forecasting protocols imagined by the *MPCS.* Any improvement, however incomplete or inexact, in the publicizing of the cost or likely demographic effects of any change in criminal law would seem to be a benefit. And here, the evidence-based and cost-efficacy measures incorporated by the *MPCS* may serve the *MPCS* well politically, however uncertain both the theory and the data are concerning these measures.

Another version of transparency the *MPCS* indirectly endorses is any emphasis on so-called evidence-based risk assessments at all stages of prosecution and sentencing.[143] If these are not only widely used but also

141. *See, e.g.*, Weisberg, *supra* note 114, at 221 (stating that TIS is a "much-invoked goal of reform" which is "the close tie between the sentence decreed in legislation or announced at trial and the sentence the offender actually serves").

142. For a review of the claims of Virginia in this regard, see *id.* at 215–16.

143. MODEL PENAL CODE: SENTENCING, *Reporter's Introduction* 21 (Preliminary Draft No. 6, 2008). *See generally* Roger K. Warren, *Evidence-Based Practices and State Sentencing Policy: Ten Policy Initiatives to Reduce Recidivism*, 82 IND. L.J. 1307 (2007) (discussing how greater reliance on evidence-based practices would allow state courts to improve effectiveness of state sentencing outcomes and reduce recidivism while reducing over-reliance on incarceration and promoting use of

widely publicized, the public will have a better sense of how government decides whom to incarcerate, and this would also seem incontenstable as a contribution to civic understanding in relation to policymaking—in the abstract. But, alas, no benefit is unalloyed: Imagine that an effect of this admired form of risk-assessment, and the resulting redistribution of punishment toward the most violent and dangerous, would be that the prison population would become much younger—or take on some other distinct demographic cast. The pervasive social effects of recharacterizing the model prisoner as a young predator or as some other kind of iconic character may prove culturally troublesome. It might re-invoke images of criminals that sparked the harsh populism that much of sentencing reform has aimed to quell.

VI. A CONCLUDING OBSERVATION

Regarding the dangers of populism for fair sentencing, let me end on a lighter note with the most absurdly modest of all measurements of success: duration and political quietude. By some very extreme version of "other-things-being-unequalness," we might consider the ability of any sentencing scheme, especially one created after the end of the indeterminate era, to endure without significant change or populist demands for change. The long-term endurance of a sentencing scheme might be a sign that the scheme has won popular acceptance—or, more modest yet, that it has not provoked the demagogic populism that has such wider insidious effects on criminal justice. As well, duration without change is surely easy to measure, without multiple regressions. As I have suggested, in its very partial and tentative nod in the direction of empirical assessment of sentencing reform, the *MPCS* partakes, at least implicitly, in the sober, somber recognition of how empirically unknowable the consequences of our law reform actions are. In finessing this issue, the *MPCS* rightly (again implicitly) accepts the second-best idea: we should do the right thing while imposing at least a modest burden of proof to establish that what we do, as measured however imperfectly by our best available social science, is consistent with our goals. That said, the biggest blows to criminal justice, in terms of both fairness and utility, often come from that very political demagoguery. Thus, once we accept that the *MPCS* is at least a very enlightened law reform, even if it is not universally preferred, its very adoption and political survival will be reasonable good indicia of its "success."

---

community-based alternatives for appropriate offenders).

APPENDIX

Table 1: Parole Rates in Comparison[144]

State	New Court Commitments per Parole Violator Returned	Prisoner Releases per Parole Violator Returned	Conditional Releases per Parole Violator Returned	Unconditional Releases per Parole Violator Returned
Pennsylvania	1.77	2.84	1.95	0.54
Michigan	2.13	3.59	3.04	0.42
Ohio	6.99	7.64	3.99	3.61
Alabama	6.44	7.72	4.82	2.79
Texas	1.70	2.79	2.12	0.53
California	0.51	1.47	1.44	0.02
Georgia	1.37	2.15	0.39	1.75
New York	1.69	2.78	2.44	0.31

144. For derivation of all figures, see WEST & SABOL, *supra* note 102, at 17.