

EMPIRICAL CRIMINAL LAW SCHOLARSHIP AND THE SHIFT TO INSTITUTIONS

Robert Weisberg*

INTRODUCTION.....	1371
I. DESIGNING PUNISHMENT.....	1374
II. DESIGNS OF DECISIONMAKING	1383
A. <i>Choosing Decisionmakers</i>	1383
B. <i>Structures of Cost Bearing in Decisionmaking</i>	1385
C. <i>Empirical Opportunism and Institutional Experimentation</i>	1388
III. STAGE PROCESSING AND THE LEGACY OF <i>MCCLESKEY</i>	1390
CONCLUSION.....	1399

INTRODUCTION

The broadly defined field of “criminal law” or “criminal justice” has always provided ripe material for empirical studies, but in assessing the most interesting new research from the Seventh Annual Conference on Empirical Legal Studies (CELS), it is useful to underscore the word “legal” and thereby demarcate a more specific category of research. Legal academics have long benefited from the venerable line of social science research on the causes of crime or changes in crime rates. This work, as recently exemplified by inquiries into the causes of America’s dramatic post-1990 decline in crime,¹ usually involves social or other factors often exogenous to the legal system.² Another tradition of research involves the study of broad penal legislation—such as the

* Edwin E. Huddleson, Jr. Professor of Law, Stanford Law School. My thanks to Samuel Eisenberg and Andrew Van Denover for expert assistance.

1. For discussions of the possible causal factors, see FRANKLIN E. ZIMRING, *THE GREAT AMERICAN CRIME DECLINE* (2007); and Steven D. Levitt, *Understanding Why Crime Fell in the 1990s: Four Factors That Explain the Decline and Six That Do Not*, J. ECON. PERSP., Winter 2004, at 163, 176-83 (contending that key factors include increased hiring of police officers, increased incarceration rates, the demise of the crack cocaine epidemic, and the legalization of abortion in the 1970s).

2. To further stress the artificiality of the legal-social distinction, the increase in abortion rates that Levitt offers as a partial explanation of the crime decline, *see* Levitt, *supra* note 1, at 181-82, is obviously both a social phenomenon and the result of a change in law.

death penalty³ or gun laws⁴—where empiricists test the efficacy of laws in accomplishing goals that are uncontested as a matter of social policy or morality (e.g., protecting victims from violent crime). This research has tended to implicate empirical studies in volatile public and cultural debates at too high a level of generality (often with highly equivocal results) to bear on the operations of criminal justice systems.⁵

While acknowledging the roughness of some of these distinctions,⁶ I will use some of the exemplary papers presented at CELS to describe and assess a trend toward more empirical “legal” research—papers that help demonstrate how statistical findings can inform government in the context of institutional decisions within the formal legal system. This research, focusing on the political economy of criminal justice agencies, examines legal decisionmaking under the constraints of constitutional demands, managerial and fiscal challenges, and, sometimes, the peculiar deontological forces that distinguish criminal law from other forms of regulation. Most notably, these studies examine the unavoidable discretion that the legal system invests in officials at key decision points, with special attention to prosecutors—the most important and unex-

3. Compare Hashem Dezhbakhsh et al., *Does Capital Punishment Have a Deterrent Effect? New Evidence from Postmoratorium Panel Data*, 5 AM. L. & ECON. REV. 344, 361-70 (2003) (finding a murder-preventing effect as high as eighteen lives saved per execution), with John J. Donohue & Justin Wolfers, *Uses and Abuses of Empirical Evidence in the Death Penalty Debate*, 58 STAN. L. REV. 791, 818-35 (2005) (criticizing the statistical accuracy of Dezhbakhsh et al., *supra*).

4. For the fullest elaboration of the notion that concealed carry rights deter criminal victimization, see generally JOHN R. LOTT, JR., *MORE GUNS, LESS CRIME: UNDERSTANDING CRIME AND GUN CONTROL LAWS* (3d ed. 2010). For a major critique of Lott’s work, see Ian Ayres & John J. Donohue III, *Shooting Down the “More Guns, Less Crime” Hypothesis*, 55 STAN. L. REV. 1193, 1285-86 (2003) (reviewing earlier versions of Lott’s work, finding flaws in his regression and techniques, and inferring that there is no solid proof of such a deterrent effect).

5. Of course, a full cost-benefit analysis of such laws would indeed involve consideration of the ground-level operations of criminal justice agencies in enforcing them, but such an analysis may have too many moving parts across too varied a set of jurisdictions to permit useful general conclusions. Proof of that complexity comes from recent discussions of the possible effects of alterations in drug prohibition laws. It is very daunting to assay a study of the efficacy of these laws in reducing either the incidence of drug use or the incidence of whatever might be (perhaps tautologically) defined as the incidence of drug crimes. For speculations about the possible effects of the repeal of marijuana prohibitions, see Robert Weisberg, *Approaches to Assessing the Effects of Marijuana Criminal Law Repeal in California*, 43 MCGEORGE L. REV. 1, 4-10 (2012) (assessing such outcomes as increased marijuana use and possible resulting increase in crime; shift of police resources to increase arrests for other crimes; and greater federal enforcement of federal drug laws to fill the gap).

6. Thus, while I describe a new trend toward very targeted analyses of official discretion, I note that legal academics enriched the study of institutional decisionmaking in criminal justice decades ago through such work as the classic empirical study of juror behavior, HARRY KALVEN, JR. & HANS ZEISEL, *THE AMERICAN JURY* (1966), and through viewing prosecutorial discretion as a species of administrative procedure, see, e.g., Robert L. Rabin, *Agency Criminal Referrals in the Federal System: An Empirical Study of Prosecutorial Discretion*, 24 STAN. L. REV. 1036 (1972).

amined of all categories of officialdom. They identify patterns and outcomes at early and low-level stages of adjudication, especially where choices of both means and ends implicate nonutilitarian values. In so doing, this research frames and informs questions for the officials who delegate, supervise, or exercise discretion, but it ultimately cannot answer the questions it poses.⁷ Empirical research can provide what can be measured, but criminal law consists of more than just the measurable. At some point, research can do no more than arm decisionmakers with a sharpened sense of the scope and nature of the residual factors that empirics cannot measure with the reassurance of precision.

In this Essay, I will review, describe, and assess this trend through some CELS examples sorted into some of the key varieties of institutional empirical research about criminal justice. Part I considers how empiricists can inform the design of penal and sentencing systems through microanalysis. Such work takes advantage of quasi-experiments in the administration of criminal justice that exploit very specific legal changes under the methods of regression discontinuity, testing out marginal instances of crime reduction and efficiency, and cautiously generalizing upward. Part II considers research that evaluates (and can recommend) structural or bureaucratic changes in criminal justice administration, including selection of officials and allocation of cost bearing among

7. These distinctions turn in part on the special pull of moral retributivism and Eighth Amendment principles in substantive crime definition and sentencing, as well as the nonutilitarian values of privacy and autonomy that weigh heavily in criminal procedure doctrine. Thus, in current debates about enhanced stop-and-frisk practices by police, while empiricists can choose to limit their inquiries to the purely utilitarian question of the efficacy of stop and frisk as opposed to other crime-fighting measures, a true cost-benefit analysis must deal with incommensurables, such as the deontological interest in privacy, or the moral and constitutional demands of racial fairness. For an empirical study of the efficacy and possibly discriminatory implementation of stop-and-frisk policies permitted by *Terry v. Ohio*, 392 U.S. 1 (1968) (holding that a stop and frisk based on the reasonable suspicion standard, which is short of probable cause, is consistent with the Fourth Amendment), see Andrew Gelman et al., *An Analysis of the New York City Police Department's "Stop-and-Frisk" Policy in the Context of Claims of Racial Bias*, 102 J. AM. STAT. ASS'N 813, 821-22 (2007) (finding significant racial disparities in street stops and inferring that this disproportionate stopping undermines efficacy because such racially disparate stops are not productive in terms of establishing probable cause for arrests). Another example of the empirical challenge in criminal procedure research is the debate over the rules derived from *Miranda v. Arizona*, 384 U.S. 436 (1966). The research debate has focused on measuring how many otherwise legitimate confessions are lost as a result of the rules, or how many otherwise legitimate convictions are lost as an ultimate result. Compare Paul G. Cassell & Bret S. Hayman, *Police Interrogation in the 1990s: An Empirical Study of the Effects of Miranda*, 43 UCLA L. REV. 839, 904-16 (1996) (finding significant effect of *Miranda* on admissible confessions and conviction rate), with George C. Thomas III, *Plain Talk About the Miranda Empirical Debate: A "Steady-State" Theory of Confessions*, 43 UCLA L. REV. 933, 934-35 (1996) (arguing that Cassell and Hayman's data are inadequate to support their conclusions). But clear answers to these empirical questions still would not provide guidance as to the desirable legal response, because numbers alone cannot tell us which confessions result from police practices that are, by some normative standard, unacceptable regardless of their efficacy.

agencies or levels of authority. Part III turns to the crucial phenomenon of official discretion as it is exercised at distinct and incremental stages of the criminal justice process. Empiricists have turned to analyzing how such discretion has an often hidden, and often cumulatively self-reinforcing, effect on the ultimate criminal law outcomes traditionally studied.

While numbers can teach us some things, in all these types of research—especially the work that focuses on discretion—the nature of official decisions demands attention to immeasurable moral and constitutional restraints. Whether realizing it or not, the researchers in this last area are responding to a great and often unacknowledged challenge to criminal empirical legal studies. I refer to the United States Supreme Court's warning in 1987 in *McCleskey v. Kemp*⁸ that an honest empirical inquiry into racial and other inequities in any key part of a criminal justice system can lead to knowledge that is dangerous because it induces an existential crisis in those who would then reform that system.

I. DESIGNING PUNISHMENT

Along with the great American crime drop of the 1990s, perhaps the most notable general subject of criminal justice commentary in recent years has been the anomalous size of the American prison population, which has been denounced in political and moral commentary under the evocative term “mass incarceration.”⁹ The sharp and sustained rise in the rate of incarceration that was simultaneous with the 1990s crime drop has naturally led to empirical research about its causes and cost effectiveness. This research has necessarily faced daunting questions of endogeneity—that is, questions about the relationship between the crime rate and the size of the prison population, or about whether the increase in incarceration is itself the major cause of the crime drop.¹⁰ The national embarrassment of mass incarceration has raised moral and political concerns that defy any measurement.¹¹ But even in the realm of the theoretically measurable, it has forced us to consider complicated questions of “cost,” including institutional costs on the government side but also the

8. 481 U.S. 279, 314-19 (1987); *see infra* text accompanying notes 89-102.

9. For broad critiques of the current state of American incarceration, *see, for example*, TODD R. CLEAR, *IMPRISONING COMMUNITIES: HOW MASS INCARCERATION MAKES DISADVANTAGED NEIGHBORHOODS WORSE* (2007); and MARIE GOTTSCHALK, *THE PRISON AND THE GALLOWS: THE POLITICS OF MASS INCARCERATION IN AMERICA* (2006).

10. *See, e.g.*, RYAN S. KING ET AL., *THE SENTENCING PROJECT, INCARCERATION AND CRIME: A COMPLEX RELATIONSHIP 4* (2005), *available at* http://www.sentencingproject.org/doc/publications/inc_iandc_complex.pdf (finding that increased incarceration may explain roughly 25% of the violent crime decline of the 1990s).

11. *See, e.g.*, MICHELLE ALEXANDER, *THE NEW JIM CROW: MASS INCARCERATION IN THE AGE OF COLORBLINDNESS* (2010) (arguing that mass incarceration is leading to the creation of a de facto racial class system).

devilishly difficult question of cost on the individual's side in terms of the true lifetime consequences of incarceration for offenders and their families.¹²

Thus, much recent empirical research examines the structures of punishment that attend to both quantifiable and nonquantifiable values but also necessarily turn to the traditional litany of purposes of punishment (deterrence, incapacitation, etc.) by which sanctions are to be justified and tested.¹³ Recent decades have seen various phases of this scholarship, including the debate about "selective incapacitation" as a parsimonious design for reducing crime.¹⁴ But over the years, in both abstract economic modeling and ground-level empirical studies, most of the utilitarian attention has been placed on the area of general deterrence.¹⁵ Here, the landmark is the work of economist Gary Becker. Becker disdained explanations of criminal behavior rooted in notions of inherent malevolence, deviance in individuals, or social causes, instead positing a ruthlessly utilitarian model of the encounter between rational self-interest and the price of crime, which he calculated as a function of the probability and severity of sanctions.¹⁶ Becker's work led to the insight that in equilibrium, criminals are more likely to be the risk-preferrers among us—that is, more sensitive to changes in probability, or "certainty," than severity.¹⁷ Those general princi-

12. For measurements of the lifetime economic costs of the experience of imprisonment, see BRUCE WESTERN, *PUNISHMENT AND INEQUALITY IN AMERICA* 125-26 (2006) (finding that there is an "aggregate earnings penalty" of having been incarcerated that results in a significant reduction in lifetime wages and chance of employment). Indeed, I have so far oversimplified the problem of unmeasurable costs and benefits by eliding the argument that the purpose of incarceration is retributive and has nothing to do with crime reduction efficacy in the first place.

13. See Robert Weisberg, *Reality-Challenged Philosophies of Punishment*, 95 MARQ. L. REV. 1203, 1230-51 (2012) (reviewing how different theories of punishment try to account for, or fail to address, changes in imprisonment rates).

14. See JAMES Q. WILSON, *THINKING ABOUT CRIME* 145-58 (rev. ed. 1983) (arguing that selective use of incapacitation is the most efficient mechanism for reducing crime). Selective incapacitation turned out to be one of the more apparently measurable mechanisms of crime reduction to invite empirical studies, with equivocal results because of both empirical uncertainty on the strictly empirical questions of estimating avoided crimes (i.e., the diminishing utility of marginal incarceration and the replacement crime problem) and moral concern about the very phenomenon of selectivity and the constitutional implications of preventive detention. See FRANKLIN E. ZIMRING & GORDON HAWKINS, *INCAPACITATION: PENAL CONFINEMENT AND THE RESTRAINT OF CRIME* 31-38, 83-85 (1995) (casting empirical doubt on selective incapacitation theory).

15. For a comprehensive economic model of deterrence, see generally A. Mitchell Polinsky & Steven Shavell, *On the Disutility and Discounting of Imprisonment and the Theory of Deterrence*, 28 J. LEGAL STUD. 1 (1999).

16. See generally Gary S. Becker, *Crime and Punishment: An Economic Approach*, 76 J. POL. ECON. 169 (1968).

17. Cf. David P. Farrington et al., *Changes in Crime and Punishment in America, England and Sweden Between the 1980s and the 1990s*, 3 STUD. ON CRIME & CRIME PREVENTION 104, 128 (1994) (noting an inverse relationship between crime rate and probability of conviction). One innovative twist on the standard wisdom was explored by one of the papers presented at CELS. See Christoph Engel & Daniel Nagin, *Who Is Afraid of the Stick? Exper-*

ples having been long established, the challenge of researchers has been to identify the specific legal rules and structures that deter most efficiently, and researchers who venture into this area know the sobering results. At least once we recognize the baseline of our very strict penal system, the evidence of marginal deterrent effects is very hard to establish.¹⁸ And since the more weighted certainty side of the equation implicates stages of criminal justice decisionmaking and institutional mechanisms rather than just legislated sentences, general deterrence studies, as discussed below, implicitly force us to study those ground-level mechanisms.

One phase of this scholarship, motivated in part by the worry over mass incarceration, has been the reconsideration of the transition from the pre-1980 mode of unstructured and indeterminate sentencing to the more formulaic and rigid “determinate” schemes,¹⁹ especially in the controversial example of the federal sentencing guidelines enacted almost thirty years ago.²⁰ The most recent focus has been on the reconsideration, by both academics and lawmakers, of the move to more determinacy through the medium of the proposed new Model Penal Code (MPC) sentencing rules.²¹ Debate over reconsidering modern “determinate” structures has been an important test of the relevance of empirical legal studies.²² The new MPC proposal has provoked various analyses

imentally Testing the Deterrent Effect of Sanction Certainty (CELS Version, Nov. 2012), available at <http://ssrn.com/abstract=2097055>. In an admittedly very artificial laboratory experiment, Engel and Nagin address the standard explanations for the standard wisdom: present orientation, stigma, and risk preference. Their elaborate gambling/theft experiment uses anonymity to eliminate the stigma factor and instant penalization to eliminate the present orientation factor. The experimenters thereby nicely isolate the risk preference question and find something intriguingly nonstandard. They conclude that risk-averse participants are more deterred by severity, as predicted by theory. But risk-seeking participants are also more deterred by severity if the offense has a positive expected value, or if a risk-neutral participant would be indifferent between committing and desisting from crime. Thus, in the authors' view, the theory is only confirmed where the sanction is severe and frequent enough that committing the offense has a negative expected value. *Id.* at 2-5.

18. Gary Kleck et al., *The Missing Link in General Deterrence Research*, 43 *CRIMINOLOGY* 623, 653-55 (2005) (finding weak evidence of marginal deterrence effects because of the weak relationship between changes in actual punishment levels and in public perceptions thereof).

19. “Determinacy” is an ambiguous term, but it typically has signaled two different but related phenomena: (1) rigid or even mandatory sentencing formulas that reduce judicial discretion at sentencing, and (2) the abolition of older systems of parole that rested late-stage discretion in executive branch officials paired with the complementary move toward sentences that are relatively fixed at the time of initial sentencing. For a superb taxonomy of these different schemes, see Steven L. Chanenson, *The Next Era of Sentencing Reform*, 54 *EMORY L.J.* 377, 381-86 (2005).

20. For a harsh critique of the federal guidelines, see generally KATE STITH & JOSÉ A. CABRANES, *FEAR OF JUDGING: SENTENCING GUIDELINES IN THE FEDERAL COURTS* (1998).

21. *MODEL PENAL CODE: SENTENCING* (Tentative Draft No. 2, 2011).

22. In addition, contemporary policy analysis increasingly relies on “evidence-based practices” for risk assessment of criminal offenders, especially at the point of sentencing or possible parole release or while on supervision during probation or parole. The idea of evi-

of different sentencing schemes in regard to their effects on prison size. The drafts of, and commentaries on, the new MPC rules refer to empirical studies about the choice between fairly rigid and determinate statutes and guidelines and more flexible structures that place much more discretion in a parole board or other administrative branch officials.²³ Weighing in on this debate and responding to laments about mass incarceration, Kevin Reitz, Reporter for the MPC's revision of its sentencing rules, himself has proffered empirical evidence that more fixed and determinate rules—generally lauded by the draft—are not associated with higher imprisonment rates,²⁴ while at the same time carefully taking no position on whether any particular degree of imprisonment is inherently desirable.²⁵

To squeeze this heterogeneous body of work into a descriptive category, let us think of it as “system design” research that attempts cost-benefit analysis of punishment schemes (both sentencing and correctional policies). It looks to traditional utilitarian rationales for punishment not as philosophical subjects, but as possible mechanisms by which certain means of punishment can lead to the goal of crime reduction. The sample of new CELS papers in this category happens to be very small and skewed. Nevertheless, I will draw from this sample a few modest conclusions relevant to this category, and mainly I infer a need for caution about overly ambitious claims of empirical studies in this area. One reason for caution is the danger that new tools and datasets might tempt us to

dence-based decisionmaking has come to dominate the rhetoric of reform discussions, though sometimes more as a symbolic mantra than as a well-defined principle of research guiding practice. For a general elaboration of the values of evidence-based risk assessment in various stages of adjudication and sentencing, see generally Roger K. Warren, *Evidence-Based Practices and State Sentencing Policy: Ten Policy Initiatives to Reduce Recidivism*, 82 IND. L.J. 1307 (2007).

23. See generally Robert Weisberg, *Tragedy, Skepticism, Empirics, and the MPCs*, 61 FLA. L. REV. 797 (2009).

24. See, e.g., Kevin R. Reitz, *Don't Blame Determinacy: U.S. Incarceration Growth Has Been Driven by Other Forces*, 84 TEX. L. REV. 1787, 1794-1801 (2006). For some empirical confirmation that modern guidelines systems have been reasonably successful in their goal of reducing illegitimate disparities in sentencing, see 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); and John F. Pfaff, *The Continued Vitality of Structured Sentencing Following Blakely: The Effectiveness of Voluntary Guidelines*, 54 UCLA L. REV. 235, 235 (2006). One can add the question of sheer economic efficiency in imprisonment, though beyond the surmountable difficulty of identifying the actual cost per unit of imprisonment, the measure of the output of imprisonment is a highly contestable matter, unless by some indirection it can be linked to the uncontested goal of reduced crime. Thus, ironically, the empirical doubts about any utilitarian conclusions from this research are matched by the normative uncertainty about the goals of the institution being empirically studied.

25. The goals were not exactly purely utilitarian in the cruder sense of affecting crime or prison size, and not exactly deontological, but rather some hybrid. See Kevin R. Reitz, *American Law Institute, Model Penal Code: Sentencing, Plan for Revision*, 6 BUFF. CRIM. L. REV. 525, 555-57 (2002) (stating that the rationale for proposed sentencing rules rests on tempering the utilitarian vision of the original MPC with the retributive concept of proportionality).

overrate the likelihood that questions of strict utilitarian efficacy can be resolved. The other is the reminder that, in this arena, the efficacy questions are embedded not only in a complicated mix of utilitarian goals, but also involve an interaction among utilitarian goals, deontological values, and institutional complexities. Thus, I suggest, the most promising research in this category might look to smaller, more manageable questions.

At the most avowedly ambitious end of the spectrum is *The Prisoner's Dilemma: A Cost Benefit Approach to Incarceration* by David Abrams.²⁶ In a sense, Abrams reflects the trend in scholarship I am describing because the topic of the paper is not something as general as whether imprisonment reduces crime. Rather, its subject is how government can most sensibly respond to economic or judicial demands to reduce prison population; indeed, his paper is particularly motivated by the current situation in California. There, the state prison system has been under federal court injunctive orders to reduce overcrowding,²⁷ and so the state must identify the most cost-effective and safety-protecting remedy. If prisons are to be downsized, choices need to be made by legislators and officials as to the most cost-beneficial ways to do so, with special heed to the threats any changes pose to public safety. One possible mechanism is a marginal change in the average sentence length for all crimes (merely theoretical, Abrams says, though arguably not as theoretical as he seems to readily concede). Another is a crime-by-crime reduction in sentences by downwardly reclassifying certain crimes. The third is a one-time prisoner release.²⁸

Abrams's ultimate conclusion is that, in terms of balancing all the costs and benefits (with a special emphasis on public safety because of its political salience), the one-time prisoner release is the way to go, especially because it leads to the smallest reduction in general deterrence.²⁹ Let me stress that this conclusion is intuitively quite plausible. After all, whatever general deterrent effect we get from the criminal justice system comes far more from generalized knowledge of crime definitions and penalties than from awareness of or sensitivity to prisoner releases; any individual who relaxes his self-restraint in light of such a one-time event is likely someone who is already willing to contemplate spending time in prison in the first place.

But what is the research significance of the way Abrams gets to his conclusions? He applauds the contemporary empirical tools now being applied to

26. David S. Abrams, *The Prisoner's Dilemma: A Cost Benefit Approach to Incarceration* (CELS Version, Nov. 2012), available at <http://ssrn.com/abstract=2109703>.

27. See *Brown v. Plata*, 131 S. Ct. 1910, 1923 (2011) (affirming the order of a three-judge district court that California's prison population must be reduced to resolve unconstitutional conditions).

28. See Abrams, *supra* note 26, at 5-6.

29. See *id.* at 6. Abrams concedes that crime reduction is not the only purpose of incarceration, noting that retribution also motivates punishment, but he disclaims any ability to account for retribution in empirical policy analysis. See *id.*

criminal justice studies, including the ones he draws on for his own cost-benefit analysis. These tools include:

natural experiments, instrumental variables (IV), difference-in-difference, and other approaches that allow for causal inference even when randomized experiments are unavailable. These techniques, along with the increased availability of large digitized datasets of criminal justice data make a cost-benefit analysis of incarceration feasible for the first time. This is the first article to use sophisticated empirical estimates to arrive at precise valuations of specific policy changes. In so doing this article will hopefully be the first of a new type of analysis, a scientific analysis of crime and decision-making about criminal justice.³⁰

Abrams's ambition is admirable, and in the course of his paper he often does a great service in elegantly summarizing many bodies of empirical research in this area. But in my view, Abrams's approach has the virtues and vices of being encyclopedic. He helpfully summarizes earlier context-specific research, but he may be generalizing too much when he extrapolates from context-specific studies to conduct a contextless cost-benefit analysis. To forecast utilitarian effects of various policy changes, Abrams plugs in estimates of the elasticities of various mechanisms of punishment effects from analogous situations analyzed in other studies.³¹ In the impressive effort to be comprehensive, Abrams further takes on the daunting task of putting dollar values on the harm caused by particular crimes (and hence the benefit of prevented harm),³² deploying all the theories for doing so, especially focusing on various tort measurements and economic theories of contingent valuation. He also amasses a large number of state studies of governmental costs per prisoner per year and synthesizes them into his calculus.³³

My reservations are twofold. First, while the shift toward natural experiments has allowed empiricists to draw more careful, robust, and credible inferences about policy effects, it is not clear whether such effects are the same across different settings. High internal validity can come at the risk of low external validity. For instance, in measuring the power of general deterrence, Abrams reports statistical inferences from studies of the deterrent effect of

30. *Id.* at 5 (footnotes omitted).

31. *See id.* at 22-23. For example, Abrams draws on two studies involving differential punishments of juveniles and adults—Steven D. Levitt, *Juvenile Crime and Punishment*, 106 J. POL. ECON. 1156 (1998); and David S. Lee & Justin McCrary, *The Deterrence Effect of Prison: Dynamic Theory and Evidence* (Ctr. for Econ. Policy Studies, Working Paper No. 189, 2009), available at <http://www.princeton.edu/ceps/workingpapers/189lee.pdf>. *See* Abrams, *supra* note 26, at 17.

32. He briefly acknowledges the challenge: “While it may seem unusual to put a dollar value on things such as pain or fear from crime, this is exactly what judges and jurors do in calculating damages.” Abrams, *supra* note 26, at 4.

33. *See id.* at 37-39.

recidivist laws³⁴ and the significance of the juvenile/adult discontinuity in criminal sanctions,³⁵ as well as his own study of gun enhancement laws.³⁶ Then, surveying the disparate results, he settles on a figure roughly in line with earlier studies.³⁷ He similarly reviews a variety of studies of the marginal crime-reducing effect of incapacitation and chooses estimates “in the middle of the range.”³⁸ It is something like predicting an electoral race by averaging the polls. But when the analyses being averaged (or summarized) are themselves very divergent, the estimates may become quite iffy.³⁹

34. See Abrams, *supra* note 26, at 23 & n.76 (citing Eric Helland & Alexander Tabarrok, *Does Three Strikes Deter?: A Nonparametric Estimation*, 42 J. HUM. RESOURCES 309 (2007)).

35. See Abrams, *supra* note 26, at 17 & nn.50-55 (citing Levitt, *supra* note 31; Randi Hjalmarrsson, *Crime and Expected Punishment: Changes in Perceptions at the Age of Criminal Majority*, 11 AM. L. & ECON. REV. 209, 231-33; Lee & McCrary, *supra* note 31).

36. See Abrams, *supra* note 26, at 18 & n.62 (citing David S. Abrams, *Estimating the Deterrent Effect of Incarceration Using Sentencing Enhancements*, 4 AM. ECON. J.: APPLIED ECON. 32 (2012)).

37. See Abrams, *supra* note 26, at 22-23, 51 n.198.

38. *Id.* at 31-32.

39. Let me add one narrower concern. Abrams works hard to disentangle general deterrence from incapacitation, because recent studies of the role of increased incarceration in explaining the 1990s crime drop lead to the question of what effect of incarceration is most operative in that explanation. In doing so, Abrams arguably falls prey to a common analytic error that—while perhaps more theoretical than material—needs to be addressed in criminal law empirics: the phantom of “specific deterrence.” See *id.* at 23-26. Specific deterrence is enumerated in the ritual recital of purposes of punishment, but in fact it is easy to decompose its meaning. First, if we include rehabilitation in the litany, specific deterrence cannot really be distinguished from it, unless we can establish a firm distinction between moral improvement and the scared-straight consequences of the experience of incarceration. But more importantly, recidivism is simply one form of crime commission, and if the purpose of specific deterrence is to reduce recidivism then there is nothing unique about potential recidivists in the pool of potential offenders. Further, recidivism itself has no settled meaning. For example, it could be limited to new criminal convictions or it could also apply to acts that, whether or not they are crimes, are penalized as violations of probation or parole conditions and send the offender back to incarceration. As Franklin Zimring and Gordon Hawkins elucidated many years ago, because any general deterrent message is about finding a pool of people who will respond to a legislated threat of punishment, and because the population of potential responders is wildly heterogeneous in terms of sensitivity, people who have already suffered criminal punishment are simply one group in this heterogeneous mix with their own form of sensitivity. FRANKLIN E. ZIMRING & GORDON J. HAWKINS, *DETERRENCE: THE LEGAL THREAT IN CRIME CONTROL* 72-74 (1973). Finally, specific deterrence is very hard to disentangle from incapacitation because the recidivist tendencies of any imprisoned criminal will automatically be affected by the degree to which he is older when he gets out. Whether this question is “merely” theoretical is hard to say. It may be material if a faulty separation of specific from general deterrence leads to a kind of double counting of inputs into the cost-benefit equation. The overall point here is that if empirical studies are to help us understand how well the purposes of punishment achieve their goals, we need great analytic care in defining and distinguishing these purposes—and to do so requires very fine attention not only to such questions as institutional processes for implementing punishment, but also the sociological and psychological questions of how deterrent signals are received.

Second, the analysis provides somewhat less clarity about which studies are included and how findings are aggregated. For example, the cost-benefit analysis relies on an elasticity of crime of 10% with respect to sentence length, but some studies report an elasticity of negative 74% or 0%. The reader is left guessing how and why the paper arrived at one estimate, and the results may be quite sensitive to these choices. A meta-analysis, which would formalize the process of aggregating research findings, may have been helpful to make transparent the precise steps by which the cost-benefit analysis was conducted.

Abrams displays both creativity⁴⁰ and sharp institutional knowledge in his eclectic estimates of relevant factors.⁴¹ But because Abrams is also rigorous in acknowledging the roughness of all these estimates and the level of generality at which he amasses them, he leaves us more with tempting possibilities of practical lessons than with concrete guidance. When he synthesizes all these estimates into a complex set of cost-benefit equations, he concludes that a prisoner release can save “millions of dollars” and that the “small increase in crime costs” still leaves us with a “substantially positive net benefit as well as a large benefit to cost ratio for each crime.”⁴² To put this nevertheless praiseworthy effort in a different perspective, Abrams’s paper might have been rhetorically recast as a “what-if” exercise: if we could come up with reliable estimates for all the numbers he tries to generate, we might well be able to endorse the one-time release over other schemes. But to reverse the emphasis (the glass is half-empty), the difficulty of achieving confidence in these estimates demonstrates the need for caution in such endorsements. In fact, to switch the perspective again, with California’s current downsizing efforts as the major example of a situation for which Abrams hopes to provide the solution,⁴³ the better next step in research is to attend to detailed empirical assessments of what is now happening in California—which happens to be a complicated mix of the different schemes Abrams is analyzing.

A less-is-more, smaller-is-better approach to these questions may be too timid. But at least for the sake of argument, I will match the Abrams study with an example of the opposite approach—a very delimited study that produces some interesting results in very narrow datasets but which might permit the timid to assay at least potential generalizations upward.

40. See, e.g., Abrams, *supra* note 26, at 40 (citing David S. Abrams & Chris Rohlfs, *Optimal Bail and the Value of Freedom: Evidence from the Philadelphia Bail Experiment*, 49 *ECON. INQUIRY* 750, 750 (2011) (using defendant’s willingness to post bail as a measure of lost freedom or wages)).

41. See, e.g., *id.* at 11 (observing that prisoner releases tend to disproportionately include those with minimal time remaining on their sentences, the medically infirm, or those convicted of less serious crimes).

42. *Id.* at 54.

43. See *id.* at 3-4.

The example is Benjamin Hansen's *Punishment and Recidivism in Drunk Driving*.⁴⁴ It is a model of the carefully delineated regression discontinuity study. Hansen relies on a narrowly circumscribed but data-rich set of drunk driving arrests and convictions in Washington State. He exploits an interesting oddity in Washington state law: a blood alcohol content (BAC) of 0.08% makes one guilty of an ordinary DUI, while a BAC over 0.15% makes one guilty of aggravated DUI, which carries a much higher punishment (a mix of fines, license restrictions, and possible incarceration).⁴⁵ Hansen is attracted to the DUI dataset because of its cleanness. For one thing, this DUI law offers a great example of a regression discontinuity because the arbitrary cutoffs for these BAC numbers sit along a very smooth continuum of alcohol effects that cannot be differentiated by the drunk driver near the line dividing the penalty regimes.⁴⁶ Further, as a limiting virtue of the study, the largely reactive mechanism for stopping and busting drunk drivers with breathalyzers removes many possible confounding variables in the decision to arrest.⁴⁷ Also, as a virtue of this particular dataset, Hansen is able to manipulate time periods of differing lengths after an initial arrest to address the temporal effects of any earlier offense.⁴⁸

Perhaps the key to this study is an odd fact of Washington statutory law—that any repeat offense gets a much-enhanced penalty, but the enhancement is the same whether the prior is ordinary or aggravated.⁴⁹ And on this score, the results are intriguing. First, holding lots of demographic and personal factors equal, the recidivist penalty definitely reduces drunk driving on its own by at least a few percentage points.⁵⁰ But Hansen qualifies this happy inference by noting that the rationality that offenders thereby exhibit is of a very “bounded” kind, because those whose first offense is an aggravated DUI seem to be deterred from a next offense more than those with an ordinary DUI prior, even though the enhancement is the same.⁵¹ As Hansen puts it, the aggravated prior offender is thereby showing a heightened updating effect—she revises her fear of future arrest by the nature of her last punishment, but is thereby acting irrationally.⁵²

44. Benjamin Hansen, *Punishment and Recidivism in Drunk Driving* (CELS Version, Nov. 2012).

45. *See id.* at 1.

46. *See id.* at 8.

47. *See id.* at 6.

48. *See id.* at 16-17.

49. *See id.* at 12.

50. *See id.* at 12-13. Alas, Hansen uses the term “specific deterrence,” *id.* at 5, without acknowledging its conceptual instability—but the possible error is harmless here because it is simply a name given to a mechanism he otherwise accurately measures.

51. *See id.* at 2, 16.

52. Of course, other explanations may be available for this apparent irrationality, including the greater shaming effect of the aggravated penalty for the first offense as compared to a low likelihood of awareness of the oddity of the recidivist law.

Moreover, Hansen demonstrates an acute eye for conceptually and institutionally different notions of punishment by incorporating incapacitation in a subtle way. He nicely notes that various licensing and driving restrictions following a DUI conviction can count as incapacitation every bit as much as incarceration.⁵³ And he takes operational advantage of this broader concept of incapacitation when he considers the strength of the deterrent effect here over time. While he finds a significant recidivism-reduction effect from a conviction, he finds that after conviction the effect drops off notably after two years.⁵⁴ This is psychologically plausible, but Hansen refines his numbers by taking into account the contribution of incapacitation immediately following the conviction, thereby reducing the rate of diminution of the deterrent effect.

The Hansen study serves as a fine bookend to the Abrams study, as a rigorously limited study opportunistically exploiting an interestingly arbitrary twist in a habitual offender law of a very specific sort. DUI crimes are very different from other crimes in many ways, and the generalizing power of the Hansen study should be viewed in this regard. Hansen guides legislators toward possible experiments with sharply delineated changes in drunk driving (and perhaps somewhat analogous) laws and shows how a fair amount of internal validation in conducting such experiments can be achieved. Its invitation to conduct small experiments in parsimonious and finely calculated punishment schemes—especially those involving discrepancies between actual and perceived punishment penalties—will likely prove more salutary than the broader strategic conclusions that might be extrapolated from more ambitious studies.

II. DESIGNS OF DECISIONMAKING

A. *Choosing Decisionmakers*

Another fruitful line of institutional research concerns the selection of the officials who will make discretionary decisions—most obviously prosecutors and judges. This research is especially interesting because it involves exogenous factors—voters' political preferences—as they bear on criminal justice. In *Prosecutor Elections, Mistakes, and Appeals*,⁵⁵ Bryan McCannon finds an ingenious way of testing the effect of popular elections on prosecutorial decisions. His innovation lies in identifying not only the particular sensitivity points in the politics, but also the subtleties by which we can measure distortions in prosecutorial decisionmaking under the constraints of formal legal doctrine and institutional dynamics. McCannon delineates a six-month period before an election, sorts the candidates out in terms of whether there is an incumbent, and

53. *See id.* at 17.

54. *See id.* at 18.

55. Bryan C. McCannon, *Prosecutor Elections, Mistakes, and Appeals* (CELS Version, Nov. 2012), available at <http://ssrn.com/abstract=2099730>.

examines whether a conviction obtained within that six-month window is reversed on appeal. Put bluntly, his question is whether “hawkish behavior” of prosecutors nearing elections leads to inaccuracies in the criminal justice system.⁵⁶ McCannon is obviously aware that appellate reversal is neither rare nor an uncontested measure of the validity of a prosecutor’s decision to charge and try. Thus, he refines his study by controlling for various “innocent” factors in an appeal, such as the complexity or numerosity of the issues. Relying on a rich panel dataset of appellate decisions in western New York, he finds that if an initial felony conviction takes place in the six months before the election date in a race with an incumbent, the probability of reversal increases by roughly 6%.⁵⁷ In considering the relationship between measurable outcomes and normative goals, this study does not present huge problems. If we believe that, at least to some degree, an appellate court reversal signifies an injustice at the trial stage, then McCannon has given us a measure of how a particular governmental structural scheme creates injustice. Whether injustice is the necessary meaning of reversal would be a vast and difficult inquiry implicating normative consideration of much legal doctrine. But in his carefully circumscribed study and attention to the institutional nuances of legal decisionmaking, McCannon shows that conventional regression methods under highly restricted boundaries offer us an honest way of at least tentatively identifying injustice-producing vectors in the system.

That feature makes the McCannon study an interesting counterpart to *Judicial Selection and Death Penalty Decisions* by Brandice Canes-Wrone, Tom Clark, and Jason Kelly.⁵⁸ Many recent studies have examined judicial elections outside the realm of criminal law,⁵⁹ but the authors here focus on the relationship between the election of judges and the morally and politically volatile area of death penalty cases.⁶⁰ Most importantly, they distinguish partisan from nonpartisan elections (and retention schemes from purely competitive schemes), testing the intuition that if we are to have judicial elections at all, it is better to keep them “less political” by favoring nonpartisan elections and retention schemes. They amass plenty of data from many states, carefully refining their datasets by grouping together roughly similar elections schemes, focusing on elections for the states’ highest appellate courts and using regressions to compare the election data with votes to affirm or reverse death sentences. And they discover the nice irony that it is in the “less political” elections that the voters

56. *See id.* at 3-4.

57. *See id.* at 15.

58. Brandice Canes-Wrone et al., *Judicial Selection and Death Penalty Decisions* (CELS Version, Nov. 2012).

59. *See, e.g.*, JAMES SAMPLE ET AL., *THE NEW POLITICS OF JUDICIAL ELECTIONS 2000-2009: DECADE OF CHANGE 1-4* (2010) (finding that the consequences of judicial elections include a surge in campaign spending, a rise in attack ads, and the loss of public faith in the judiciary).

60. *See* Canes-Wrone et al., *supra* note 58, at 3-4.

rely more on stated, or at least apparently demonstrable, positions on hot-button issues like the death penalty (as compared to more reliance on party label), and that this phenomenon clearly tilts the judges towards majority sentiment favoring or disfavoring capital punishment when they decide individual death penalty appeals.⁶¹

My point of contrast with the McCannon study here is that this effect cannot be called “unjust” in the same way as McCannon’s discovered effect can be. Aside from obvious cases of purposeful discrimination or violations of clear substantive criteria or criminal procedure rules, it is difficult to distinguish correct from incorrect individual death penalty decisions when the vague principles of the Eighth Amendment defer so much to popular morality and jury sentiment. Of course, an extension of the judicial election study could test the fairness of death sentences upheld under the nonpartisan/retention schemes by looking at later stages of appellate review—federal habeas corpus review or direct review by the United States Supreme Court—or by a nonjudicial statistical study of whether a sentence is inequitable in comparison to those for similar offenders or offenses in that jurisdiction. There may be a broader measure of injustice here as well. In light of the focus of Eighth Amendment law on avoiding arbitrary and capricious outcomes (wholly independent of any right/wrong answers in individual death penalty decisions), if a factor like selection of judges can materially influence death penalty outcomes, that effect itself might serve as proof of injustice. But such large speculations aside, this study at least isolates a vector in decisionmaking of measurable degree and then forces us toward a normative evaluation of its meaning.

B. *Structures of Cost Bearing in Decisionmaking*

As empirical studies constructively turn to the details of institutional management in criminal justice, one important emerging theme is cost externalization. One CELS paper examines this phenomenon at a fairly abstract level using experimental game methods. In *When Punishment Doesn’t Pay: “Cold Glow” and Decisions to Punish*, Aurélie Ouss and Alexander Peysakhovich help confirm the intuition that any party with power to impose punishment will be more inclined to do so when that party need not internalize the cost.⁶² But a

61. *See id.* at 4.

62. Aurélie Ouss & Alexander Peysakhovich, *When Punishment Doesn’t Pay: “Cold Glow” and Decisions to Punish* (CELS Version, Nov. 2012). The authors posit a general theory of “cold glow,” that is, a positive personal benefit to an individual from assigning punishment to wrongdoers, as a corollary to the “warm glow” benefits of altruistic behavior. *Id.* at 2. They attempt to tease out the impact of this phenomenon in a series of clever games that allow observation of the effects of varying the structure of the costs of punishment.

For example, in one game, group members can decide to punish other members who steal from each other by making them sit out future rounds of the game. *Id.* at 17-20. Punishers tended to assign additional penalty rounds, beyond the number rationally necessary for deterrence, when the cost of punishment was taken not out of the individual’s point total, but

more concrete test of internalization in a specific institutional setting comes from Itai Ater, Yehonatan Givati, and Oren Rigbi in their paper *Organizational Structure, Police Activity and Crime*.⁶³ The authors examine how a peculiar structure in Israel, and an abrupt change to that structure, produced a fabulous quasi-experimental opportunity on this very theme. If the police determine whom to arrest for certain crimes, as is virtually universal, but if the identity of the payer of the cost of housing arrestees also changes—in this case, from the police themselves to some other entity—one might expect a change in the number of arrests. In the spirit of more circumscribed microresearch, the paper benefits from its opportunistic cleanness: the major reason for the change in cost bearing has to do with the very exogenous event of a scandal of an escaped rapist, so we have no endogeneity here.⁶⁴ Further, the implementation was staged incrementally by region, and scheduled according to the exogenous factor of the capacity of national facilities, so that we get a clean way of testing how the change occurred over time. These features of the quasi-experiment also provide a check against spatial displacement of crime as a hidden explanation for apparent crime or arrest rate changes. Moreover, the study benefited from large numbers and controls for demographic and environmental variables.⁶⁵

The direction of the change here is hardly surprising—when the police no longer bear the costs of jailing arrestees, the arrest rate goes up.⁶⁶ But the degree of the change here and the specificity of the results are remarkable. And a further payoff is a delightfully logical set of effects on arrest rates for different categories of crime, crime rates, and “arrest quality.”⁶⁷ On the one hand, freed from bearing the full costs of arrest, the police arrest more people, and, largely as a result of the incapacitating effects of longer detentions, the crime rate goes down.⁶⁸ Moreover, it goes down mainly for lower-level crimes⁶⁹—the ones most likely to have been at the margins of police priority before the change. On the other hand, the indictment rate per arrest goes down⁷⁰—a logically

rather out of a collective pool of points. *See id.* at 22-26. In a second game, consisting of only one round, the probability of capture is varied. *Id.* at 27-28. While the behavior of the takers varies with the probability of being captured, the behavior of those assigning penalty points does not. *Id.* at 28-30. Ex post assignment of punishment has no deterrent effect in this game, but occurs even though there is a personal monetary cost to the punisher. *See id.* at 27-28. The authors discuss several pathways by which cold glow benefits and punishment cost structures could influence the real-world behavior of voters, juries, law enforcement, and judges. *Id.* at 14-17.

63. Itai Ater et al., *Organizational Structure, Police Activity and Crime* (CELS Version, Nov. 2012).

64. *Id.* at 2, 6-7.

65. *Id.* at 7-8.

66. *Id.* at 9.

67. *See id.* at 10-12.

68. *Id.* at 9, 12-13.

69. *Id.* at 11.

70. *Id.* at 10.

consistent outcome if we view prosecutorial discretion as a screen for “quality” of arrests.

This Israeli experiment is so neat as to not have exact analogies to American criminal justice, but it has plentiful partial analogies that represent a major new subject for empirical analysis. Its relevance to American law is especially striking in regard to one consequence of the mass incarceration problem. I refer to changes in state, county, and local government structure and financing that shift the cost of incarceration, especially where motivated by exogenous pressure to reduce the size of prison populations. The most dramatic recent example is currently taking place in California, in its controversial innovation called Realignment. In large part because of a court injunction mandating a reduction in the amount of overcrowding in the state prison system, the state has shifted huge numbers of prisoners to county jails.⁷¹ Under the new scheme, large categories of lower-level felons will serve time in county jails rather than in state prisons. Equally dramatic, whereas parole violators had for many years been sent right back to prison for low-level crimes and even noncriminal violations, revocations from supervision after release from either prison or jail will now be served in jail. These changes move responsibility for the costs of punishment far closer to the level at which prosecutorial, judicial, and law enforcement decisions are made that create the need to incarcerate in the first place. Of course, the state has to reimburse the counties for many of these costs, partially out of the savings to the state prison system, but the very question of how to reimburse itself implicates the question of internalization. As David Ball has shown, the first cut at a reimbursement formula was superficially logical but, in fact, economically perverse: it made reimbursement proportionate to the counties’ previous “contributions” to the prison system, thus reinforcing externalization.⁷²

Doubtless the cost allocations are far more complicated in the American system than the Israeli system, and the indirectness of cost bearing is far more

71. For a useful summary of the new law, see BARRY KRISBERG & ELEANOR TAYLOR-NICHOLSON, *THE CHIEF JUSTICE EARL WARREN INST. ON LAW & SOC. POLICY, REALIGNMENT: A BOLD NEW ERA IN CALIFORNIA CORRECTIONS* (2011), available at http://www.law.berkeley.edu/files/REALIGNMENT_FINAL9.28.11.pdf.

72. See W. David Ball, *Tough on Crime (on the State’s Dime): How Violent Crime Does Not Drive California Counties’ Incarceration Rates—And Why It Should*, 28 GA. ST. U. L. REV. 987, 1074 & n.146 (2012). The state is now modifying that formula. At the same time, at least along one dimension, Realignment is turning up something akin to the Israeli experiment. Under the new law, municipalities are not getting much, if any, share of the reimbursement funds given to counties. Typically, municipal police officers pass costs upstream, because their decisions to arrest require counties to pay for any subsequent prosecution and both counties and states to pay for subsequent incarceration. California’s Realignment may provide an example of costs moving in the opposite direction; after supervision of paroled prison inmates shifts back to the counties, local police may bear much of the cost of any necessary arrests for parole violations. Of course, any such cost shifts happen against the baseline of original structural funding formulas among levels of government, but abrupt marginal changes often provide useful quasi-experiments to study cost shifting.

complex. Nevertheless, Ater and his coauthors do not purport to take any normative position on the degree of undesirability of the effects they discern here, as they take no position on the respective optimal numbers of property, public order, and violent crimes. Some of the decreases in crime may be traceable to the incapacitation following arrests that then were proved to lack probable cause, so we get a Fourth Amendment-type principle at war with crime prevention.⁷³ On the other hand, the authors draw strong institutional inferences here, finding that the change in arrest numbers and externalization can clearly be imputed more to top managers than to ground-level police,⁷⁴ especially because low-level police officials cannot influence the duration of detention.

C. *Empirical Opportunism and Institutional Experimentation*

In *Does the "Community Prosecution" Strategy Reduce Crime? A Test of Chicago's Experience*,⁷⁵ Thomas Miles performs a wonderful exercise in empirical research by taking a very general new principle in criminal justice and then finding a way to examine its operation in a concrete institutional way. The term "community" in criminal justice studies is a temptingly dangerous one, its sentimental value often obscuring very troubling line-drawing problems within demographic and political entities.⁷⁶ In the 1990s, the term "community policing" came to be blurred with a variety of trends in modern policing associated with "broken windows," "quality-of-life," or "order-maintenance" law enforcement begun in New York City in the early 1990s.⁷⁷ For some, the term "community policing" signifies the idea of a nonadversarial mode of interaction between police and citizens. But the seductiveness of the term has often obscured the truer meaning of New York-style policing, which involved highly technical targeting of neighborhoods where crime was concentrated and arrests for minor crimes, where the goal was not so much reducing the psychological degradation of neighborhoods as catching felony fugitives.⁷⁸ One important

73. See Ater et al., *supra* note 63, at 12-13.

74. *Id.* at 15.

75. Thomas J. Miles, *Does the "Community Prosecution" Strategy Reduce Crime? A Test of Chicago's Experience* (CELS Version, Nov. 2012).

76. See Robert Weisberg, *Restorative Justice and the Danger of "Community,"* 2003 UTAH L. REV. 343, 370-71.

77. See George L. Kelling & William J. Bratton, *Taking Back the Streets*, CITY J., Summer 1994, available at <http://www.city-journal.org/article01.php?aid=1428> (explaining the animating principles of New York City's policing innovations). William Bratton is the originator of the term "broken windows" policing and later served as New York City Police Commissioner.

78. Two pieces unpacking the multiple meanings of these new notions of policing and critically examining their measurable success are John E. Eck & Edward R. Maguire, *Have Changes in Policing Reduced Violent Crime? An Assessment of the Evidence*, in *THE CRIME DROP IN AMERICA* 207, 224-28 (Alfred Blumstein & Joel Wallman eds., 2000) (finding that evidence is mixed on efficacy of generic zero-tolerance strategies in reducing violent crime while questions remain about effects on police-community relations); and Bernard E. Har-

derivation of the community-policing concept has been the notion of “community prosecution.” As usefully delineated by Anthony Thompson, the promise of this new mode of prosecution is for assistant district attorneys and investigators to be dispersed into neighborhoods where they can participate in setting local priorities concerning the enforcement and charging of criminal offenses.⁷⁹ The key premise of this style of prosecution is that partnerships between prosecutors and local citizens will enhance mutual trust and thereby help reduce the most destructive forms of neighborhood crime, thus encouraging citizens to report crime and participate in its prevention and investigation.

Miles offers a superb example of what might be called “empirical opportunism.” He takes on the implementation of community prosecution in Cook County, Illinois (mostly Chicago), by testing the effects in a directly spatial way: if local satellite offices of the state’s attorney’s office are strategically located in certain corners of the city, what are the effects in terms of improved police-citizen relations, and how do such improvements bear on the crime rate?⁸⁰ Miles recognizes that while the very plausible promise of this concept has been articulated in generic terms and described with some local detail, the concept has not been empirically tested, especially in terms of public safety.⁸¹

The precise mapping of the dispersion of the offices provides just the concrete data Miles needs. Moreover, Miles’ paper offers another CELS example of exploiting a discontinuity in the data to enable cleaner analysis: in Cook County, purely exogenous budget cuts led to an “on/off/on again” variation in the application of community prosecutions in some neighborhoods.⁸² This discontinuity thereby allows Miles to perform a differences-in-differences treatment for numerous offense categories estimated from a panel of police beats observed monthly over a period of nearly seven years.⁸³

The results offer some cause for optimism. Miles infers that office dispersion has significantly reduced certain very serious categories of crime, especially aggravated assaults, robberies, and burglaries, while leaving lesser property crime rates relatively unchanged.⁸⁴ Moreover, in an admirably circumscribed effort at cost-benefit analysis, Miles calculates some realistic estimates of the

court, *Reflecting on the Subject: A Critique of the Social Influence Conception of Deterrence, the Broken Windows Theory, and Order-Maintenance Policing New York Style*, 97 MICH. L. REV. 291, 332-39 (1998) (challenging broken windows policing by arguing that the alleged correlation between disorder and serious crime fails to account for a variety of factors influencing the deterioration of a neighborhood).

79. See Anthony C. Thompson, *It Takes a Community to Prosecute*, 77 NOTRE DAME L. REV. 321, 344-46 (2002).

80. Miles, *supra* note 75, at 2.

81. *Id.*

82. *Id.* at 2-3 (internal quotation marks omitted).

83. *Id.*

84. *Id.* at 24-26. Ironically, the least affected crimes include the lower-level “quality-of-life” crimes of which order-maintenance policing has been viewed as a cure. *Id.* at 13-14, 25-27.

incremental costs of the dispersion of offices in terms of staff salaries and other concrete measures.⁸⁵ Making no broad claims for the feasibility of measuring the cost of a crime (or the benefit of preventing a crime), he relies on some established metrics (direct cost and willingness to pay) and very tentatively suggests that, if done thoughtfully, community prosecution can be cost beneficial.⁸⁶ Miles makes perfectly clear that his study cannot tease out which of the submechanisms of the Cook County approach were the most influential, and he is thus especially cautious in suggesting that this approach will work elsewhere.⁸⁷ But this example of empirical opportunism is a fine model for empiricists seeking to help criminal justice officials design and test small experiments inspired by evocative but often vague ideas in public discourse.

III. STAGE PROCESSING AND THE LEGACY OF *MCCLESKEY*

The last and perhaps most important subsampling of CELS papers I assess falls under a category that I will, perhaps clumsily, call “stage processing.” At one level, the idea is that each particular step in criminal justice investigation and adjudication is worth attending to. At another, it is all about recognizing that discretion, especially by executive branch officials, is the most important and least examined driver of criminal justice outcomes.⁸⁸ Stage-by-stage discretion poses special demands for empirical analysis, because of both data limitations and a dearth of models with which to study these data. But the attractions of such studies are multiple: they can tell us things that end-state outcome patterns cannot, and they can better test the ability of empirical studies to inform the actors and institutions of criminal justice by addressing the key moments at which decisions under legal, constitutional, and moral constraints are made. Let me begin though by observing that there is a deep background to these stage processing studies. It has to do with the great visibility in law and empirics of the issue of racially disparate and discriminatory outcomes in conviction and sentencing, and the haunting shadow of one major case, of which these new studies are the unacknowledged progeny.

85. *Id.* at 27-28. The costs include extra staff salaries and local rental rates for office space. *Id.* at 27.

86. *Id.* at 28.

87. *Id.*

88. An excellent example of recent research into this phenomenon comes from John Pfaff, who examines the various intermediating explanations for the dramatic spike in imprisonment in the last thirty years. In *The Causes of Growth in Prison Admissions and Populations*, Pfaff draws on a vast body of newly available data to conclude that the growth in prison populations has been driven almost entirely by increases in felony filings per arrest, as opposed to rates of arrest, prison admissions per filing, or even time served per admission. John F. Pfaff, *The Causes of Growth in Prison Admissions and Populations* 5-7, 19-20, 35 (July 12, 2011) (unpublished manuscript), available at <http://ssrn.com/abstract=1884674>.

The parent of these studies is *McCleskey v. Kemp*,⁸⁹ the most vexing of the early cases testing the constitutionality of the modern death penalty laws passed in the 1970s. The defendant, a black man convicted of killing a white man, presented in the lower courts the famous Baldus study, a comprehensive statistical analysis that identified a key form of discrimination in the implementation of the modern death penalty.⁹⁰ All else held equal (through numerous controls for severity of the homicidal conduct and criminal priors), the chance of a death sentence was many times higher for the killing of a white person than for the killing of a nonwhite person.⁹¹ The Eleventh Circuit, in its consideration of the case, assumed the validity of the Baldus study but nonetheless found the statistics to be “insufficient to demonstrate” unconstitutional discriminatory intent or arbitrariness.⁹² Justice Powell’s opinion for the Supreme Court similarly finessed the empirical question, assuming the Baldus study was methodologically sound and going straight to the constitutional implications of the evidence.⁹³ In regard to the defendant’s Equal Protection Clause claim, the Court held that whatever the evidence showed in terms of disparate outcomes generally, there was no way of identifying purposefully discriminatory decisionmaking by the particular officials (or jurors) with power to exercise discretion over the outcome in *McCleskey*’s case, and hence the purposeful discrimination requirement of an equal protection claim could not be met.⁹⁴ The more interesting claim was grounded in the Eighth Amendment, where the irrationality or racially disparate pattern of the decisions could create a constitutional problem regardless of intent if it established that the implementation death penalty had become “cruel and unusual.”⁹⁵ The Court’s rejection of this claim is the heart of the matter. Justice Powell made one of the all-time great slippery slope arguments: Racially disparate outcomes were likely the result of many possible layers of discretionary and unintentional (though perhaps negligent or reckless) decisionmaking at all stages of the criminal process.⁹⁶ And to ferret out these

89. 481 U.S. 279 (1987).

90. *Id.* at 283, 286.

91. *Id.* at 286-87.

92. *Id.* at 289-90 (internal quotation mark omitted) (citing *McCleskey v. Kemp*, 753 F.2d 877, 895 (11th Cir. 1985) (en banc)).

93. *See id.* at 291 & n.7.

94. *See id.* at 294-95. One of the CELS papers actually continues the Baldus-type study of the death penalty itself. In *Unconvincing Protestations: The Persistent Role of Race in Capital Charging and Sentencing in North Carolina, 1990-2009*, Barbara O’Brien, Catherine M. Grosso, and George Woodworth assess the North Carolina Racial Justice Act, the instrument the state devised to address *McCleskey*-type disparities. They conclude that even after controlling for multiple measures of culpability, murders of white victims are 1.6 times as likely as others to lead to a death sentence, and the authors attribute the effect largely to charging decisions of prosecutors. Barbara O’Brien et al., *Unconvincing Protestations: The Persistent Role of Race in Capital Charging and Sentencing in North Carolina, 1990-2009*, at 3, 22-23 (CELS Version, Nov. 2012).

95. *See* U.S. CONST. amend. VIII; *McCleskey*, 481 U.S. at 301.

96. *McCleskey*, 481 U.S. at 311-14.

microdecisions and find them unconstitutional would not simply lead to the straightforward remedy of abolishing the death penalty. Rather, according to Justice Powell's haunting admonition, "McCleskey's claim, taken to its logical conclusion, throws into serious question the principles that underlie our entire criminal justice system."⁹⁷

In part the Court was worried that "irrelevant" factors beyond race could be the basis of new claims: "The Constitution does not require that a State eliminate any demonstrable disparity that correlates with a potentially irrelevant factor in order to operate a criminal justice system that includes capital punishment."⁹⁸ But the more important concern was that even were courts to focus solely on race, the implications of accepting the Baldus study for noncapital crimes, and for early, discretionary stages of criminal adjudication, would have been staggering:

In the absence of a current, Baldus-type study focused particularly on the community in which the crime was committed, where would [the prosecutor] find a standard? Would the prosecutor have to review the prior decisions of community prosecutors and determine the types of cases in which juries in his jurisdiction "consistently" had imposed the death penalty when the victim was white and the defendant was of a different race? And must he rely solely on statistics? Even if such a study were feasible, would it be unlawful for the prosecutor, in making his final decision in a particular case, to consider the evidence of guilt and the presence of aggravating and mitigating factors? However conscientiously a prosecutor might attempt to identify death-eligible defendants under the dissent's suggestion, it would be a wholly speculative task at best, likely to result in less rather than more fairness and consistency in the imposition of the death penalty.⁹⁹

Thus, according to the dissent, the Court preferred "a presumption that actors in the criminal justice system exercise their discretion in responsible fashion, and we do not automatically infer that sentencing patterns that do not comport with ideal rationality are suspect."¹⁰⁰ The majority argued that it simply adopted a position of wise realism about the possibility of any formal legal regulation: "Prosecutorial decisions necessarily involve both judgmental and factual decisions that vary from case to case. Thus, it is difficult to imagine guidelines that would produce the predictability sought by the dissent without sacrificing the discretion essential to a humane and fair system of criminal justice."¹⁰¹

So the *McCleskey* decision offers us the great irony that capital defendants had to be denied the right to greater scrutiny of prosecutorial discretion because the legal system could not survive such a level of scrutiny applied to prosecu-

97. *Id.* at 314-15.

98. *Id.* at 319.

99. *Id.* at 318 n.45.

100. *Id.* at 337 (Brennan, J., dissenting).

101. *Id.* at 314 n.37 (majority opinion) (citation omitted).

tors more generally. The even greater irony, however, may be that, in his dissent, Justice Brennan tried to offer the odd reassurance that extending any ruling in favor of McCleskey to noncapital defendants would be legally and statistically impossible anyway:

Furthermore, the Court's fear of the expansive ramifications of a holding for McCleskey in this case is unfounded because it fails to recognize the uniquely sophisticated nature of the Baldus study. McCleskey presents evidence that is far and away the most refined data ever assembled on any system of punishment, data not readily replicated through casual effort. Moreover, that evidence depicts not merely arguable tendencies, but striking correlations, all the more powerful because nonracial explanations have been eliminated. Acceptance of petitioner's evidence would therefore establish a remarkably stringent standard of statistical evidence unlikely to be satisfied with any frequency.¹⁰²

A suite of new papers from CELS implicitly takes up the implications of the *McCleskey* admonition. In the death penalty context, however agonizing the constitutional choice of remedy would be, it is relatively bounded. One could abolish this unique penalty altogether.¹⁰³ These newer papers raise the question of how we might conduct *McCleskey*-type analysis of the earlier stages of decisionmaking for the vast universe of noncapital cases, where no such surgical remedy is conceivable. Since Baldus-type findings at early stages of ordinary criminal cases cannot possibly lead to abolition of the penal code, the studies sharply confront us with questions about how legal decisionmakers can absorb and apply this empirical information when the responses will have to be more modest, complicated, or subtle.

Sonja Starr, in her paper *Estimating Gender Disparities in Federal Criminal Cases*,¹⁰⁴ advances the inquiry in several key ways. Gender disparity in sentencing, especially federal sentencing, seems at first a usefully delimited area for refining empirical tools. It is hardly the most controversial or constitutionally worrisome form that prejudice in criminal law outcomes takes. The common intuition that women are treated more leniently turns out to be accurate and, to many, morally understandable. Starr seeks, however, to tease out the true magnitude and source of this disparity and to deploy her findings to refine some of the common intuitions about why it happens. But in this circumscribed area, she helps advance our understanding of stage processing. One key step in her paper is to find a technical means to bring nonprison sentences into the denominator and estimate where they would fit on the continuum of factors

102. *Id.* at 341-42 (Brennan, J., dissenting).

103. Or there is the alternative that some states have adopted (though in *Pulley v. Harris*, 465 U.S. 37, 50-51 (1984), the Court said this was not required): a statistical protocol for constantly testing new death sentences for horizontal equity and proportionality, with the obvious limitation that testing is done for capital sentences only.

104. Sonja B. Starr, *Estimating Gender Disparities in Federal Criminal Cases* (CELS Version, Nov. 2012), available at <http://ssrn.com/abstract=2144002>.

that affect the length of prison sentences.¹⁰⁵ This maneuver hugely improves the accuracy of her tracking of the most obvious indicator of disparity—the actual length of prison sentences. Next, and perhaps most crucially, she finds a way to decontaminate the measure of the factual severity of the crime from all the endogeneity of judgments about it that follow in later stages.

Starr’s innovation is to focus on the recorded arrest offense, rather than the formal charge or any later manifestation, and to code that arrest offense according to a formal scheme helpfully provided by the United States Marshals Service.¹⁰⁶ Those codes are more nuanced, pragmatic, and fact sensitive than the penal code definitions. This is crucial when analyzing federal sentencing because it sets a more accurate baseline than the indictment or conviction offense, or even the guidelines offense level under the United States Sentencing Commission’s rules. It also isolates a more accurate depiction of the true facts of the crime, which may affect later decisionmakers, than the ironically named “real offense” facts used at the sentencing stage under the federal guidelines protocol, which, despite the name, are usually the result of a lot of bargaining and winnowing.¹⁰⁷ Further, Starr augments the preprocess predicate of the offense with such fixed factors as race, ethnicity, level of education, citizenship status, and criminal record.

Another key feature of Starr’s approach is to maintain sharp doctrinal sensitivity in deploying statutory law to capture up front the severity of the true offense facts. A good example of this is her decision to treat drug charges differently from others, because of the mazelike overlap and ambiguity of the drug statutes.¹⁰⁸ Her least bad and eminently sensible solution is to use the mandatory minimum sentence for the relevant amount of drug as the best measure for those crimes.¹⁰⁹

Using sequential decomposition methods borrowed from labor economics and inverse propensity score weighting,¹¹⁰ Starr is then able to isolate the unexplained disparities at each of several stages that lead up to the ultimate disparities in sentencing.¹¹¹ Her comments on this method make a very insightful observation about its usefulness for criminal justice stage processing: Normally “[p]ath-dependence” is a drawback to sequential decomposition, “because in

105. *See id.* at 5-6.

106. *See id.* at 3, app. at i-ii.

107. Under U.S. SENTENCING GUIDELINES MANUAL § 1B1.3 & cmt. n.1 (2012), federal sentencing judges may take into account “relevant conduct” facts about foreseeable acts or harms associated with the charged offense, even if those facts are not alleged in the indictment, determined at trial, or admitted by the defendant at a guilty plea colloquy. For a general review of how the federal guidelines expanded sentencing facts to include such “real offense” facts, see David Yellen, *Reforming the Federal Sentencing Guidelines’ Misguided Approach to Real-Offense Sentencing*, 58 STAN. L. REV. 267, 271-72 (2005).

108. *See Starr, supra* note 104, at 10.

109. *Id.*

110. *See id.* at 10-11, 19.

111. *Id.* at 10-11.

many contexts, when multiple correlated covariates together explain a certain portion of an outcome gap, there is no theoretical reason to ‘blame’ one over the others.”¹¹² But here path-dependence is desirable because “the justice process is itself path-dependent: earlier decisions constrain later ones.”¹¹³ And of course, Starr addresses the matter of omitted variables. In finding reassurance that they do not taint her findings too much, she makes some institutionally savvy points about the under-the-radar thinking of officials. Thus, she infers that unobserved divergences between arrest offense and actual criminal conduct may distort bias disparity estimates.¹¹⁴ For example, if police tend to treat men more harshly by overstating a man’s culpability relative to a woman’s in the arrest record, this disparate treatment may not be observed in the disparity estimates, and the estimate would be understated.¹¹⁵ Conversely, if a male offender’s conduct is in fact more violent but that greater violence is not captured by the coding in the arrest report, the result might be an overstatement of the disparity. The results of all this are findings of significant disparity at all the postarrest stages of the process, establishing that the intuition about leniency for females is true and that it sets in very early. In Starr’s key findings, various measures tell us about the contribution of various portions of the ultimate sentencing disparity to the later stages. While the largest contribution shows up at the “[g]uidelines fact-finding” stage before the judge, her method allows us to make a truer measure of that divergence, and the divergence in the judge’s ultimate sentencing calculation, by using the arrest coding as a predicate for the nonprejudiced part of the decision.¹¹⁶ In that regard, we can get a sense of the points along the decisionmaking process at which the leniency is active, before we get to the more visible decisions such as the setting of offense levels or ultimate sentence calculation.¹¹⁷

Starr is then free at the end to speculate, without the expectation of great rigor, on the nature of the motivation, though by this point in the study the unrigorous speculations turn out to be enhanced by some of the data she has turned up. Two speculations—that female parents get more leniency than non-parents, and that defendants seen as “girlfriends” of male defendants are often viewed as less culpable because of the intuition that they are mere “follow-

112. *Id.* at 11 (citing Nicole Fortin et al., *Decomposition Methods in Economics*, in 4A HANDBOOK OF LABOR ECONOMICS 1 (Orley Ashenfelter & David Card eds., 2011); John DiNardo et al., *Labour Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach*, 64 *ECONOMETRICA* 1001 (1996)).

113. *Id.*

114. *Id.* at 12.

115. *Id.*

116. *See id.* at 11.

117. *See id.* at 10-11. Of course, the challenge of causal inference here remains very severe. Even with the complexity of the arrest code analysis, there may be subtle differences between the way men and women commit these crimes—differences that prosecutors may have rational but intuitive ways of evaluating. *See id.* at 12-13. With respect to the latter, women tend to win more leniency in multiple defendant mixed-gender cases.

ers”—find some support in her study.¹¹⁸ But another speculation—that women are more prone to cooperate with the authorities and thereby win sentence deductions—finds little support.¹¹⁹

A comparison between Starr’s paper and two others helps us refine just what we mean by stage processing analysis under the *McCleskey* heading. Starr’s approach teases out the degrees of disparity contributed at various levels, so we can be better informed about the stages most susceptible to distortion by illegitimate factors. Another angle on stage processing comes from *Evaluating Racial and Ethnic Disparities in Criminal Process: Evidence from North Carolina*.¹²⁰ If Starr’s analysis relies on a carefully delimited situation—gender disparity just in federal cases—this study, by Christopher Griffin, Frank Sloan, and Lindsey Chepke, similarly limits itself by staying within one state and, more importantly, staying within the parameters of a single and relatively simple crime: drunk driving.¹²¹ Within those parameters, the authors are able to evaluate data on a number of stages and decision points, especially bail decisions (pretrial detention or release and size and form of cash bail requirements) and choice of type of lawyer (retained, public defender, or pro se representation).¹²² They are also able to examine prosecutorial declinations, the alternate paths of guilty pleas and trials, and sentencing outcomes, with unusual and interesting attention not just to length of sentence, but to the choice between county jail and state prison.¹²³ And one specific development in empirical criminal studies (also present in the Starr and other new CELS studies) is dramatically evident in the North Carolina study: until recently, “racist” disparities were tested almost exclusively in terms of black/white, but that distinction really meant black/nonblack, because Hispanics were classified as white. The obvious question of disparity with respect to other groups, most obviously Hispanics, has haunted such studies, but only recently has sufficient data on Hispanics within the various stages of the criminal justice system been available, and the North Carolina study, as I note below, exploits this in some dramatic ways.¹²⁴

The detailed findings of this study are complicated, but the authors summarize a few of them thus:

118. *See id.* at 13-15.

119. *Id.* at 15.

120. Christopher L. Griffin, Jr. et al., *Evaluating Racial and Ethnic Disparities in Criminal Process: Evidence from North Carolina* (CELS Version, Nov. 2012) [hereinafter Griffin et al., CELS Version]. This paper has since been accepted for publication. Christopher L. Griffin, Jr. et al., *Corrections for Racial Disparities in Law Enforcement*, 55 WM. & MARY L. REV. (forthcoming 2014).

121. Griffin et al., CELS Version, *supra* note 120, at 4.

122. *Id.* at 14-15.

123. *Id.*

124. *See id.* at 2, 9.

We find significant gaps in pretrial release conditions (harsher for black and Hispanic defendants), prosecutorial declinations (more often for Hispanic men), and decisions to enter a not guilty plea (less often for both Hispanic and black men). Although few cases pass through the sieve of plea bargaining, conviction rates conditional on not guilty pleas are lower for black men. In terms of sentencing, we find substantial disparities in the likelihood that Hispanics (almost never) and black defendants (very often) are sent to prison. Taken together, our analysis suggests that less harsh sentences partially offset higher arrest rates and more strenuous pretrial release conditions for Hispanics, who also are likely to face the threat of deportation.¹²⁵

The findings may not seem overly dramatic (nor need they be for this to be a useful model), but a few things stand out. First, the introduction of data on Hispanics has some surprising consequences. There is no simple way in which Hispanics and blacks differ from whites in these stage outcomes. Different degrees and directions of divergence occur at the various stages and decision points. Second, some of the findings—especially that Hispanics enjoy a very high rate of declinations compared to blacks and, most importantly, much more frequently receive jail rather than prison sentences—have one striking explanation: state officials are happy to invite federal immigration authorities to step in to “handle” cases of Hispanic offenders for them.¹²⁶ This inference is, as with Starr’s analysis of penal definitions, a good example of how overall contextual understanding of legal doctrine can greatly enhance the power of legal analysis. But here is the key refinement of stage processing that Griffin and his coauthors offer: whereas Starr’s paper identifies the mix of legitimate and illegitimate factors at various stages, the North Carolina study sees an interactive effect. Some surprising findings of (relative) pro-minority leniency at the later stages might, in the authors’ view, reveal a corrective effect in response to earlier stage disparities.¹²⁷ Thus, the paper tempts us with the possibility of some kind of cultural and institutional “guilty conscience” phenomenon whereby disparities at one point are mitigated at another. The evidence for this phenomenon is equivocal, but the suggestion is provocative for future study.

A final paper in this subset is in some ways the most ambitious. John Sutton’s *Structural Bias in the Sentencing of Felony Defendants* takes on a far bigger and more heterogeneous dataset than the others: it draws on many thousands of cases of several categories of felony crimes from forty counties, with special emphasis on the pretrial detention and guilty plea stages.¹²⁸ It is also the most ambitious in its hypothesis about the meaning of stage processing. While the North Carolina study found some interactive effect in the form of self-correction, Sutton focuses on the phenomenon of *cumulative* disadvantage.

125. *Id.* at 1 (italics omitted).

126. *See id.* at 21-22.

127. *See id.* at 26.

128. John R. Sutton, *Structural Bias in the Sentencing of Felony Defendants* 8-9 (CELS Version, Nov. 2012), available at <http://ssrn.com/abstract=2095594>.

This concept is borrowed, with a nice improbability, from studies about professional mobility, but more intuitively from criminal “life course” studies whereby early labeling and stigma have a cumulative effect.¹²⁹ Focusing in particular on two preliminary outcomes, pretrial detention and guilty pleas, Sutton hypothesizes that “bias arises endogenously from the sequential structure of decisions that shape the flow of cases through the courts.”¹³⁰ He infers from earlier research that minority defendants are more likely to be detained pretrial, and that incentives and opportunities to plead guilty also differ by race.¹³¹

Sutton acknowledges that these sequential effects can follow complicated causal chains. For example, as a general matter, because prosecutors want to induce defendants to plead guilty, the expected sentence accompanying a plea offer is normally lower than the sentence that the defendant faces after conviction at trial. Yet defendants who are detained pretrial are both more likely to plead guilty and be sentenced more harshly than those who are freed, so teasing out the effects of a higher pretrial detention rate for minority defendants is quite an empirical challenge.¹³² Sutton notes that case-level sentencing studies usually control for the expected positive effects of detention and negative effects of guilty pleas, but do not account for racial disparities at earlier stages.¹³³ Sutton’s key step is to posit a snowball effect. People who are detained pretrial are more likely, say, to receive higher charges or longer sentences because of deference by the later decisionmaker to the authority of the earlier one, even if no new racial disparity enters in at the later stage.¹³⁴ Through a daunting set of statistical maneuvers, Sutton traces white, black, and Hispanic defendants through various permutations of stage processing (pretrial detained/guilty plea, pretrial detained/trial, pretrial release/guilty plea, and so forth).¹³⁵ The results are far too complicated for easy summary here, and arguably not dramatic, but Sutton finds a fair amount of support for his broad hypothesis.¹³⁶ And depending on how much ultimate confirmation the hypothesis wins, it may turn out to be the most haunting and worrisome proof of Justice Powell’s admonition in *McCleskey*.

Sutton ultimately infers that sentencing disparity by race/ethnicity is clear and significant. If we take the sentence the average defendant gets, minority defendants are almost one and a third times as likely to get the next highest available sentence than are nonminority defendants. Sutton is then able to trace back these outcomes to measure the cumulative effect of earlier disparities.¹³⁷ And

129. *Id.* at 4.

130. *Id.* at 3.

131. *Id.*

132. *Id.* at 3-4.

133. *Id.* at 3.

134. *Id.* at 5-6.

135. *Id.* at 38 tbl.5.

136. *See id.* at 24-25.

137. *Id.* at 19-20.

he can do so with a matrix of various permeations of stages, with such alternate intermediate points as bail/pretrial detention.¹³⁸ He finds that, depending on the route defendants take through these stages of the system, there are significant racial and ethnic disparities in the probability of a prison sentence and the availability of bargained sentencing discounts.¹³⁹ Sutton concludes that these cumulative effects are “systematic and striking.”¹⁴⁰

Sutton ends with some crucial warnings. He notes that this study cannot tell us why the cumulative disadvantage occurs more at some stages than at others, though he provocatively suggests that the most “formalized” stages, especially judicial sentencing, are the least likely to be influenced by labeling effects. He therefore urges more research into the murkiest and most interesting feature of the criminal justice system: the prosecutor’s office.¹⁴¹ Second, he notes that the brand new abundance of data on Latino defendants will allow for and require highly nuanced analyses of how subcategories of minorities are treated at various stages of adjudication, and how socioeconomic conditions in particular jurisdictions bear on comparative stage processing analysis across these places.¹⁴² And of course, stage processing can be done further to address more detailed decision-point links in adjudication, such as whether the defendant is eligible for release under any circumstances, whether bail is required, and, if so, at what amount bail is set both in absolute terms and relative to the defendant’s resources.¹⁴³

CONCLUSION

When the Court in *McCleskey* declared it would not follow the implications of the defendant’s claim and thereby intervene in prosecutorial decisions throughout the criminal justice system, it offered several reasons for this avoidance. For one thing, given the great decentralization of the prosecutorial function across a state, the Court doubted whether statistics could really help us understand the nuances of the discretion exercised by the various officials involved in this system—especially prosecutors.¹⁴⁴ The Court viewed discretion as so deeply and subtly embedded in the prosecutorial role that no menu of guidelines could ever successfully govern this discretion without destroying the

138. *See id.* at 20.

139. *Id.* at 24 (finding that mean black/Anglo and Latino/Anglo disparities range from 10% for Anglos and Latinos who are detained and plead guilty to 100% for Latinos and Anglos who are not detained and go to trial).

140. *Id.* Perhaps surprisingly, Sutton found little or no support for his hypotheses about what he calls “contextual” effects on sentencing—factors such as indicators of poverty or racial inequality in the relevant jurisdiction or venue. *Id.* at 25.

141. *Id.* at 26.

142. *Id.* at 26-27.

143. *Id.*

144. *McCleskey v. Kemp*, 481 U.S. 279, 294 n.15 (1987).

humane value of individualized decisions.¹⁴⁵ The Court also worried about putting unfair burdens on prosecutors presented with such statistical information. It observed that in the context of employment discrimination claims and charges of bias in jury selection, there are litigation procedures whereby state officials can respond to a *prima facie* case of illegal disparity with rebuttal evidence. By contrast, in regard to claims like McCleskey's, the Court cast the question solely in terms of how a prosecutor might respond to statistics about disparity proffered on appeal, often years after the trial.¹⁴⁶

The institutional turn in empirical legal studies reflects a variety of responses to those concerns. Criminal justice officials surely can benefit from statistical insights about their practices even when no constitutional challenge is posed or litigated. They can absorb and apply the insights as they wish. Most obviously, they can engage in self-examination that might reduce possible biases or distortions in their decisions to charge, to move for pretrial detention, to engage in plea bargains, or to recommend sentences, even if not in any formulaic way. Officials can also generate new administrative guidelines or refine their current ones in light of new evidence. One lesson from sophisticated new empirical studies is that we still face a universe of unobserved variables, especially those related to pre-adjudication decisions by police or prosecutors. Better facts about those variables might either support or disprove possible bias or arbitrariness in those decisions. In that sense, the value of these studies will be to inform officials and academics about what refinements in research are needed to make studies more useful and ensure that partnerships and dialogues between officials and empiricists will continue to be of great value to both.¹⁴⁷ Equally important, where the statistical analysis is very thorough, it induces the official to recognize that empirical inquiry has reached the limits of its wisdom,

145. *See id.* at 311-12.

146. *See id.* at 296.

147. A bold new foray in this direction comes from researchers at the Vera Institute. *See* BRUCE FREDERICK & DON STEMEN, *THE ANATOMY OF DISCRETION: AN ANALYSIS OF PROSECUTORIAL DECISION MAKING* (2012), available at <http://www.vera.org/sites/default/files/resources/downloads/anatomy-of-discretion-summary-report.pdf>. Using data from two populous counties, the researchers examined initial case screening, charging decisions, plea offers, sentencing recommendations, and postfiling dismissals for a variety of offenses. *Id.* at 2. Compiling detailed descriptive statistics on the patterns of these decisions, the researchers conducted interviews with prosecutors to augment and get feedback on the initial findings, and also conducted questionnaire surveys whereby prosecutors identified the factors most central to their charging decisions. *Id.* at 3. The conclusions themselves are not surprising. The self-reported data from the prosecutors indicate that strength of evidence and defendant and victim characteristics are key factors influencing decisions throughout the processing of a case. *Id.* at 3-4. But the data also suggest that "contextual" factors also play a role, including quality of trust and relationships with other criminal justice officials—law enforcement, judges, and defense attorneys—as well as concern about shortages of courtrooms and personnel and poor quality of information from resource-strapped police. *Id.* at 4. The report admonishes that officials "should be alert to the potential for contextual factors to influence and possibly distort the exercise of prosecutorial discretion." *Id.* at 5.

and then the official will clearly face the residual questions that numbers cannot answer. How does she value the moral desert of a defendant, the effects of charges or sentences on a defendant's family or neighborhood, the weight of the privacy interest embedded in a Fourth Amendment rule, or the wider social harms caused by disparities that might otherwise satisfy a cost-benefit analysis of efficiency? Those questions can pose existential challenges to the people who administer criminal justice, and institutionally sensitive empirical research can serve to clarify just when officials must confront these questions.