Correcting Criminal Justice through Collective Experience
Rigorously Examined

James S. Liebman
*Columbia Law School*, jliebman@law.columbia.edu

David Mattern

Follow this and additional works at: [https://scholarship.law.columbia.edu/faculty_scholarship](https://scholarship.law.columbia.edu/faculty_scholarship)

Part of the *Courts Commons, Criminal Law Commons, Criminal Procedure Commons, and the Law and Politics Commons*

**Recommended Citation**
Available at: [https://scholarship.law.columbia.edu/faculty_scholarship/471](https://scholarship.law.columbia.edu/faculty_scholarship/471)

This Article is brought to you for free and open access by the Faculty Publications at Scholarship Archive. It has been accepted for inclusion in Faculty Scholarship by an authorized administrator of Scholarship Archive. For more information, please contact scholarshiparchive@law.columbia.edu.
CORRECTING CRIMINAL JUSTICE THROUGH COLLECTIVE EXPERIENCE RIGOROUSLY EXAMINED

JAMES S. LIEBMAN* & DAVID MATTERN†

ABSTRACT

Federal and state law confers broad discretion on courts to administer the criminal laws, impose powerful penalties, and leave serious criminal behavior unpunished. Each time an appellate court reviews a criminal verdict, it performs an important systemic function of regulating the exercise of that power. Trial courts do the same when, for example, they admit or exclude evidence generated by government investigators. For decades, judicial decisions of this sort have been guided by case law made during the Supreme Court's Criminal Procedure Revolution of the 1960s and 1970s. It is becoming increasingly clear, however, that the rule-bound, essentially bureaucratic regulatory regime that emerged in the 1960s does not assure accurate outcomes, especially for poor and minority criminal defendants and victims. As an additional protection, this Essay urges criminal courts to stop resisting and to embrace regulatory innovations in wide use in other domains that foster self-improvement through continuous observation and rigorous analysis of data on the system's own results and error rates. The Supreme Court's own penalty-default and head-counting innovations point the way forward. A certain conception of liberty or individualism—which intersected with the Criminal Procedure Revolution but has not traditionally served the interests of populations most at risk from miscarriages of criminal justice—is the main obstacle to such reforms and should be abandoned.

* Simon H. Rifkind Professor of Law, Columbia Law School.
† J.D. 2013, Columbia Law School. Thanks to Elizabeth Cruikshank for superb research assistance.
I. INTRODUCTION

In the 1960s and 1970s, a funny thing happened to U.S. criminal law. Forces concerned about the welfare of the poor and racial minorities aligned with a longstanding libertarian strain in American life to promote the Criminal Procedure Revolution. The result was to afford judicially enforceable procedural protections against potentially abusive police, prosecutorial, and trial tactics to individuals suspected of or charged with crimes. This alignment was unusual. For most of American history before and since, criminal courts and other actors impelled by this libertarian or individualistic impulse, which is often combined with forms of parochialism and populism, have stymied the interests of the poor and racial minorities through slavery, racial segregation, lynching, capital punishment, and harsh mandatory sentencing. These forces have also


2. See, e.g., Dred Scott v. Sandford, 60 U.S. 393, 450-51 (1856) ("[T]he rights of private property have been guarded with . . . care . . . [a]nd an act of Congress which deprives a citizen of the United States of his liberty or property, merely because he came himself or brought his property into a particular Territory of the United States . . . could hardly be dignified with the name of due process of law. . . . [T]he right of property in a slave is distinctly and expressly affirmed in the Constitution.").

3. See, e.g., Plessy v. Ferguson, 163 U.S. 537, 551-52 (1896) (approving state-mandated racial segregation given the futility of government interference with individuals' "racial instincts").


5. A number of commentators have linked the death penalty's use in the United States to libertarian values. E.g., Garland, supra note 4, at 35-36; James S. Liebman & Peter Clarke, Minority Practice, Majority's Burden: The Death Penalty Today, 9 OHIO ST. J. CRIM. L. 255, 273-80 (2011); Zimring, supra note 4, at 44-48.

6. See Franklin Zimring, Gordon Hawkins & Sam Kamin, Punishment and Democracy: Three Strikes and You're Out in California 231-32 (2001) (concluding that mistrust of government has led to harsher mandatory-sentencing practices in California). One might expect libertarian mistrust of government to generate opposition to incarceration and execution. In the
resisted the right of labor to organize and of government to protect children, workers, unsophisticated consumers, and other vulnerable populations from the depredations of unregulated private enterprise.

Judicially enforced libertarian opposition to government protection of vulnerable populations peaked at the end of the nineteenth century and in to the early 1930s. During much of the remainder of the twentieth century, the bureaucratization of large-scale business and government activity gradually led to broad agreement on rules negotiated between central experts and newly powerful national interest groups representing workers, low-income individuals, and racial minorities. Through contract, legislation, and administrative regulation, these agreements rationalized business and government practice in ways that, among other things, protected the liberties of workers, the poor, and minority populations. Exemplified by rules enshrined in lengthy collective bargaining agreements protecting employee “due process” rights, these regimes proceeded on the sensible assumption that individuals empowered by such rights were more likely to cooperate in the productive endeavors, regulatory schemes, and United States, however, libertarians' strong belief in criminals' moral desert of punishment, together with their doubts about empowering judges and parole officers to exercise discretion as to the length of sentences and their parochial concerns about the incursion of crime and other unwelcome outside forces on local communities, has led libertarian values to coincide with support for mandatory sentences and executions. For discussions of this phenomenon, see ZIMRING, supra note 4, at 45–48, 98–99, 108–11; Liebman & Clarke, supra note 5, at 273–80; Angela M. Schadt & Matt DeLisi, Is Vigilantism on Your Mind? An Exploratory Study of Nuance and Contradiction in Student Death Penalty Opinion, 20 CRIM. JUST. STUD. 255, 259 (2007).

7. See, e.g., Carter v. Carter Coal Co., 298 U.S. 238, 311 (1936) (overturning New Deal regulation of coal production as an “intolerable and unconstitutional interference with personal liberty and private property”); Adair v. United States, 208 U.S. 161, 175 (1908) (invalidating a law forbidding employers to bar employees from joining unions as an “arbitrary interference with the liberty of contract which no government can legally justify in a free land”).

8. See, e.g., Adkins v. Children’s Hosp., 261 U.S. 525, 545 (1923) (invalidating minimum wage law as “interfer[ing] with the freedom of contract”); Lochner v. New York, 198 U.S. 45, 56 (1905) (invalidating limits on the length of bakers’ work week as an “unreasonable, unnecessary, and arbitrary interference with the right of the individual to his personal liberty or to enter into those contracts in relation to labor which may seem to him appropriate”); Allgeyer v. Louisiana, 165 U.S. 578, 589 (1897) (rejecting state regulation of business practices as violating “the right of the citizen . . . to live and work where he will, to earn his livelihood by any lawful calling, to pursue any livelihood or avocation, and for that purpose to enter into all contracts which may be proper”).

social-welfare objectives of their business and government interlocutors. Starting at about the midpoint of the twentieth century, when faced with business interests or government agencies unprepared to live by such rules and protections, federal and state courts imposed their own complex rule-based regimes through "public law litigation" designed to enforce due process and egalitarian rights.  

The mid-twentieth-century move toward rule-based bureaucratization and a concern for the due process rights of the poor and minorities extended to criminal justice agencies in northern and western states, including urban police forces and prosecuting offices, state police and justice departments, and state courts. In the South, however, particularly in nonurban communities, a combination of libertarian resistance to government and other large institutions, parochial worries about encroaching aspects of modern life thought to threaten traditional values, and resistance to extending equal rights to African-Americans led police, prosecuting, and judicial institutions to resist bureaucratically or judicially enforced rules.


The U.S. Supreme Court was no stranger to the threat these latter attitudes and practices posed to the rights of African-Americans and others outside the white Protestant mainstream, as evidenced by the Court's lynchings and early coerced confession, right to counsel, prosecutorial misconduct, and jury discrimination decisions. Until the 1960s, however, the Court confined itself to requiring the most basic features of Anglo-American criminal process. It was only during the 1960s and 1970s—evidently influenced by the post-New Deal linkage between bureaucratic rules and due process rights and by urban administrative and state judicial efforts to bring police officers and prosecutors under the bureaucratic control of professionalizing agencies—that the Court extended its engagement with criminal justice beyond a focus on the fundamentals of fair process by imposing what amounted to a comprehensive constitutional code of procedure. Over time, these rules encompassed, among other things: search and seizure; interrogations and confessions; procedures [and] incompetent staff in the 1950s and 1960s extended well beyond the South.

13. See, e.g., Frank v. Mangum, 237 U.S. 309, 337–38 (1915) (declining to intervene in criminal proceedings against a Jewish man charged with raping a white woman, despite evidence that proceedings were compromised by threats from angry mobs).


16. See, e.g., Powell v. Alabama, 287 U.S. 45, 71 (1932) (ruled that the state's failure to allow criminal defendants to secure counsel denied due process).


19. See STUNTZ, supra note 11, at 31–39, 194 (linking the Warren Court's Criminal Procedure Revolution to a nationwide trend toward professionalizing and bureaucratizing criminal justice systems).


21. See, e.g., Mapp v. Ohio, 367 U.S. 643, 655 (1961) (prohibiting admission of evidence against the accused that was obtained in violation of his or her Fourth Amendment rights). Each procedural rule discussed here proliferated into a body of detailed judicial rules, as defendants sought to extend the principle, and states' attorneys proposed exceptions to it. See, e.g., One 1958 Plymouth Sedan v. Pennsylvania, 380 U.S. 693, 696 (1965) (extending the Mapp exclusionary rule to forfeiture proceedings).

silence in the face of police accusations;\textsuperscript{23} eyewitness identification;\textsuperscript{24} defense representation;\textsuperscript{25} defense access to information in police files;\textsuperscript{26} competency to stand trial;\textsuperscript{27} guilty pleas;\textsuperscript{28} jury selection\textsuperscript{29} and jury trial;\textsuperscript{30} admission or exclusion of evidence and argument at trial;\textsuperscript{31} distinguishing “elements” of the offense subject to rigorous adjudicatory process from mere sentencing factors;\textsuperscript{32} evidentiary presumptions;\textsuperscript{33} burdens and

\textsuperscript{23} Compare Salinas v. Texas, 133 S. Ct. 2174, 2178, 2180–82 (2013) (allowing a defendant’s selective failure to answer some but not all police questions prior to his arrest and receipt of \textit{Miranda} warnings to be used against him at trial), with Doyle v. Ohio, 426 U.S. 610, 611 (1976) (forbidding prosecutor to use defendant’s silence in the face of police questioning after \textit{Miranda} warnings were given against him at trial as proof of guilt).


\textsuperscript{25} See, e.g., Rothgery v. Gillespie Cnty., 554 U.S. 191, 213 (2008) (concluding that the right to counsel arises once a judicial officer informs the defendant of the charge against him); Alabama v. Shelton, 535 U.S. 654, 657–58 (2002) (discussing prior Supreme Court decisions extending a right to counsel to any prosecution, whether classified as petty, misdemeanor, or felony, that actually leads to imprisonment or to a suspended sentence); Gideon v. Wainwright, 372 U.S. 335, 339–40 (1963) (requiring states to provide counsel to indigent felony defendants).

\textsuperscript{26} See, e.g., Brady v. Maryland, 373 U.S. 83, 87 (1963) (holding that a prosecution’s failure to provide defense counsel with material exculpatory evidence violated due process); United States v. Agurs, 427 U.S. 97, 103–07 (1976) (assessing how the specificity of the defendant’s pretrial request for government-held information affects a defendant’s \textit{Brady} rights).

\textsuperscript{27} See, e.g., Drope v. Missouri, 420 U.S. 162, 180 (1975) (identifying constitutional right not to be tried when mentally incompetent and requiring trial courts sua sponte to inquire about competence upon learning of evidence of incompetence); Cooper v. Oklahoma, 517 U.S. 348, 362 (1996) (invalidating a statute requiring defendants to prove their incompetence to stand trial by clear and convincing evidence).

\textsuperscript{28} See decisions cited infra note 56.

\textsuperscript{29} See, e.g., Morgan v. Illinois, 504 U.S. 719, 729 (1992) (requiring states to exclude jurors who would automatically vote to sentence any defendant convicted of capital murder to death without considering mitigating evidence); Batson v. Kentucky, 476 U.S. 79, 89 (1986) (holding that a prosecutor’s use of peremptory challenges to strike African-Americans from the jury violates equal protection); Witherspoon v. Illinois, 391 U.S. 510, 522 (1968) (forbidding states to exclude prospective jurors for cause in capital cases based on conscientious scruples against capital punishment unless they make clear that they will not impose death under any circumstances).

\textsuperscript{30} See, e.g., Burch v. Louisiana, 441 U.S. 130, 138–39 (1979) (forbidding states to simultaneously limit juries to six members and allow nonunanimous criminal verdicts); Duncan v. Louisiana, 391 U.S. 145, 149 (1968) (extending jury-trial right to state criminal defendants).


\textsuperscript{32} See, e.g., Apprendi v. New Jersey, 530 U.S. 466, 490 (2000) (requiring jury determination
standards of proof;\textsuperscript{34} harmless error;\textsuperscript{35} imposition of capital punishment;\textsuperscript{36} and appellate and postconviction process.\textsuperscript{37} The goal was to bureaucratize and professionalize law enforcement through adherence to predictable rules that protected the rights of previously underprotected individuals.\textsuperscript{38}

As this Symposium illustrates, the Court’s criminal justice code is under severe pressure today.\textsuperscript{39} The political right criticizes the Court for beyond a reasonable doubt of any fact authorizing an increase in the maximum possible criminal penalty); Ring v. Arizona, 536 U.S. 584, 609 (2002) (extending \textit{Apprendi} to require a jury finding of any aggravating circumstance that raises the maximum possible penalty from imprisonment to death).


\textbf{34.} \textit{See}, \textit{e.g.}, \textit{In re Winship}, 397 U.S. 358, 368 (1970) (requiring state to prove every element of a criminal case beyond a reasonable doubt); Mullaney v. Wilbur, 421 U.S. 684, 704 (1975) (concluding that the absence of “heat of passion on sudden provocation” was an element of, not an affirmative defense to, murder under the state law at issue, and thus the prosecution had to prove that element beyond a reasonable doubt); Cage v. Louisiana, 498 U.S. 39, 41 (1990) (per curiam) (ruling unconstitutional a “‘beyond a reasonable doubt’” instruction that the jury might have understood to dilute the \textit{Winship} standard).

\textbf{35.} \textit{See}, \textit{e.g.}, Chapman v. California, 386 U.S. 18, 24 (1967) (allowing court to uphold conviction despite constitutional error by finding the error harmless beyond a reasonable doubt).

\textbf{36.} \textit{See}, \textit{e.g.}, Furman v. Georgia, 408 U.S. 238, 239–40 (1972) (per curiam) (striking down all death-sentencing laws in the nation as then applied as excessively discretionary). Suggesting the intricacy of the rules the Court has developed to regulate capital punishment in the wake of \textit{Furman}, see, for example, Mills v. Maryland, 486 U.S. 367, 384 (1988) (barring states from requiring juror unanimity before treating a mitigating factor as a basis for a sentence less than death).

\textbf{37.} \textit{See}, \textit{e.g.}, Fay v. Noia, 372 U.S. 391, 398–99 (1963) (adjudicating effect of “adequate and independent state law ground” doctrine on availability of federal habeas corpus relief for state prisoner denied state court relief on the same claim for failure to raise the claim properly); Douglas v. California, 372 U.S. 353, 357–58 (1963) (establishing constitutional right to counsel in criminal appeals); Trevino v. Thaler, 133 S. Ct. 1911, 1921 (2013) (extending recent precedent—establishing a constitutional right to effective counsel at an initial postconviction proceeding in states that require prisoners to raise claims of ineffective trial representation in postconviction proceedings rather than on direct appeal—to states that encourage, without requiring, prisoners to raise ineffective-trial-representation claims in state postconviction proceedings).

\textbf{38.} \textit{See}, \textit{e.g.}, Cox, supra note 1, at 129 (crediting the Warren Court with creating “a vastly more humane society, one freer, more equal, and more respectful of the human dignity”); Stevens, supra note 1 (crediting the Warren Court with transforming mid-century police agencies “riddled with incompetent, lazy, and brutal behavior into a law enforcement profession that is both more effective and also commands widespread public admiration for its dedicated public service”).

interfering with communities’ and individuals’ ability to govern and protect themselves. The left is troubled by evidence that the rules continue to allow the conviction of innocent defendants and systematically undermine defendants’ rights to trial and proportionate punishment. Those in between worry that the rules are ill-suited to modern investigations and evidence. Perhaps the rules’ worst enemy is the
evidence, and assistance of counsel, among others, for failing to address the potential for inaccuracy in new forms of forensic evidence that were not contemplated when the rights were initially recognized); Christopher Slobogin, Lessons from Inquisitorialism, 87 S. CAL. L. REV. 699, 702–09 (2014) (criticizing the Fifth Amendment privilege against self-incrimination as inconsistent with a search for the truth); Dan Simon, Professor of Law and Psychology, Address at the University of Southern California Law Review Symposium: Goal Displacement: The Diminished Role of Accuracy in the Criminal Justice Process (June 7, 2013) (describing tensions between procedures designed to ensure due process and the objective of accurate verdicts).


41. See, e.g., Jim Dwyer, Peter Neufeld & Barry Scheck, Actual Innocence: When Justice Goes Wrong and How to Make it Right xviii (2003) (crediting DNA testing with “expos[ing] a system of law that has been far too complacent about its fairness and accuracy”); Brandon L. Garrett, Convicting the Innocent: Where Criminal Prosecutions Go Wrong 7–8 (2011) (“[A]t a trial, few criminal procedure rules . . . . regulate accuracy rather than procedure. [Accuracy] matters are typically committed to the discretion of the trial judge.”); Dan Simon, In Doubt: The Psychology of the Criminal Justice Process 206–15 (2012) (providing estimates of the rate of wrongful conviction in the United States); Simon, supra note 39 (citing evidence that false convictions do occur); Slobogin, supra note 39, at 702–09 (describing conditions under which the American criminal justice system is prone to convicting the innocent).

42. Professor Stuntz has argued, for example, that by taking criminal process out of the hands of state legislatures and making criminal trials risky and unpredictable for law enforcement, Warren Court rules incentivized legislators to lengthen sentences dramatically and incentivized prosecutors to use abusive practices to induce trial-avoiding pleas, which has led to overcriminalization, overcharging, oversentencing, racially disparate enforcement of drug laws, a shift in budget dollars from preventing crime to punishing it, and subordination of substantive to procedural fairness. Stuntz, supra note 11, at 58, 139, 235–36, 257–64. For contrary views, see Corinna Barrett Lain, Countermajoritarian Hero or Zero? Rethinking the Warren Court’s Role in the Criminal Procedure Revolution, 152 U. PA. L. REV. 1361, 1365–66 (2004) (arguing that the Warren Court’s criminal procedure jurisprudence was less counter-majoritarian than conventionally thought); Dripps, supra note 11, at 470–75 (criticizing Stuntz’s argument); Schulhofer, supra note 12, at 1056–79 (criticizing Stuntz’s historical analysis and providing alternative explanations for trends Stuntz attributes to the Warren Court).

43. See, e.g., Erin Murphy, The New Forensics: Criminal Justice, False Certainty, and the Second Generation of Scientific Evidence, 95 CALIF. L. REV. 721, 743, 756–77 (2007) (documenting courts’ “reluctance to approve the means necessary to effectively inspect second-generation [i.e., technologically sophisticated forensic] evidence” for accuracy); Murphy, supra note 39, at 639–45
Supreme Court itself, which repeatedly signals its ambivalence through proliferating exceptions that complicate and weaken the law,\(^4\) harmless-error findings,\(^5\) and blind acceptance of government and lower court explanations for problematic actions.\(^6\) The Court has doubts about how well its inflexible rules "fit" particular cases and about responding to violation of the rules with "nuclear" remedies that exclude powerful evidence of guilt. These doubts evidently lead the Court to blink when asked to affirm that a rule means what it seems to say or to impose the penalty the rule's violation supposedly triggers.

These concerns are hardly surprising. They reprise the powerful critique of rule-bound bureaucracy that has led actors in other public- and private-sector domains to seek alternatives.\(^7\) Because central experts

\(44.\) See, e.g., Kansas v. Ventris, 556 U.S. 586, 594 (2009) (allowing states to impeach the accused with statements made to police informants in violation of the accused’s Sixth Amendment right to counsel); Pennsylvania v. Muniz, 496 U.S. 582, 601 (1990) (declaring exception to Miranda requirements for statements made in response to "routine booking question[s]"); Illinois v. Perkins, 496 U.S. 292, 295–96 (1990) (exempting statements from Miranda rule if made outside the "coercive atmosphere" characterizing custodial interrogations); New York v. Quarles, 467 U.S. 649, 655–56 (1984) (recognizing exception to Miranda requirements for interrogation seeking information about possible ongoing public harms); United States v. Havens, 446 U.S. 620, 627–28 (1980) (declaring exception to Fourth Amendment exclusionary rule for improperly seized evidence used to impeach the accused); Stone v. Powell, 428 U.S. 465, 494 (1976) (holding that federal habeas corpus relief is not available to enforce a defendant's right to exclusion of evidence obtained in violation of the Fourth Amendment); Michigan v. Mosley, 423 U.S. 96, 104 (1975) (allowing state to use inculminating statement made after defendant was informed of his Miranda rights, refused to talk to police, and was interrogated some hours later after new warnings had been given).

\(45.\) See, e.g., DWYER, NEUFELD & SCHECK, supra note 41, at 222–35 (describing several convictions involving prosecutorial and police misconduct that were upheld under the harmless error doctrine, though defendants were later exonerated); GARRETT, supra note 41, at 201 (finding that in 30 percent of the cases of prisoners subsequently exonerated by DNA, appellate judges found that "mistakes were made" at trial but "relied on harmless error in refusing to reverse [the] conviction").

\(46.\) See, e.g., Nevada v. Jackson, 133 S. Ct. 1990, 1992–94 (2013) (per curiam) (requiring federal courts to exercise extreme deference to state courts in applying a statutory limitation on federal habeas relief to state-court applications of federal constitutional law that not only are legally erroneous but also are "unreasonable"); Rice v. Collins, 546 U.S. 333, 341–42 (2006) (requiring deference to trial judges' acceptance of prosecutors' questionable explanations for using peremptory challenges to strike minority jurors); Morgan Cloud, The Dirty Little Secret, 43 EMORY L.J. 1311, 1311–12, 1314 (1994) (claiming that "[i]njudges, prosecutors, defense lawyers, and repeat offenders all know that police officers lie under oath... most often to avoid suppression of evidence and to fabricate probable cause" and blaming current Fourth Amendment rules and practice for "provid[ing] incentives for police officers to fabricate testimony"); Peter Keane, Why Cops Lie, S.F. CHRON. (Mar. 15, 2011), http://www.sfgate.com/opinion/openforum/article/Why-cops-lie-2388737.php ("The first reason [police lie under oath] is because they... know that in a swearing match between a [criminal] defendant and a police officer, the judge always rules in favor of the officer [even when] it is embarrassingly clear to everyone... that the officer is lying under oath.").

\(47.\) See, e.g., CHRISTOPHER K. ANSELL, PRAGMATIST DEMOCRACY: EVOLUTIONARY LEARNING
cannot predict and specify rules for all imaginable situations, "street-level bureaucrats"—whether production-line workers, teachers, police officers, or trial judges—will encounter many situations for which no rule has been prescribed or in which the prescribed response is risible or counterproductive. Instead of being guided by central experts, therefore, the local actors exercise unfettered discretion to adopt strategies for solving the problems, which often promote the actors’ personal values or interests over those of the company’s owners or the voting public. Different actors will make different choices, destroying uniformity and predictability. And street-level actors will experience the rules as silly or irrelevant, prompting disrespect for the rules and the “higher-ups” who imposed them.48

48. See, e.g., Lipsky, supra note 47, at 16–25 (describing the demoralizing effect of inflexible bureaucratic rules on field-level employees of social service agencies who cannot do their job effectively without exercising a fair degree of discretion in response to the different circumstances in which their clients find themselves); Theodore R. Sizer, Horace’s Compromise: The Dilemma of the American High School 9–21 (4th ed. 2004) (noting the day-to-day difficulties of high school teachers and their lack of identification with the inflexible prescriptions of central administrators, including administrators trying their best to improve the instruction of students); Wilson, supra note 47, at 32 (“The individualistic, rule-oriented perspective of the courtroom is at variance with the situational, order maintenance perspective of the patrolman. The patrolman senses this conflict without
Neither immediate alternative to these difficulties is attractive. Higher-ups may adopt doctrines—for example, harmless error—that let themselves and field-level operatives off the hook from counterproductive rule-following. Doing so gives official status to the unfettered, often self-serving discretion that the rules were supposed to avoid. It also makes uninteresting nonevents out of near-misses—near-disastrous deviations from desired outcomes, which if properly addressed can motivate and inform effective precautions. This may create incentives to underenforce the rules and underachieve their objectives. Alternatively, higher-ups may fetishize the rules to the point—vividly described by Dan Simon and Laurie Levenson—that rule-following occurs for its own sake with little concern for whether it leads to the substantive outcomes to which the rules were supposed to be a means. Here, the result is to overenforce the rules while still underachieving their objectives.

Public-sector organizations may be especially drawn to rule following because it supplants what for them is the especially difficult process of identifying desired substantive outcomes and determining whether they have been achieved. Defining and measuring the success of for-profit operations often is easier. Their objectives usually have something to do with how much money they make, and their owners and managers have at least intermittently strong incentives to get an answer to that question as well as at least rudimentary accounting standards that provide a method for doing so. These conditions are less likely to occur in public-sector activity. Apart from observing whether the rules were followed, it is difficult to determine, for example, whether a prosecution, trial, and guilty verdict identified the real culprit and imposed a penalty with the desired—deontological? utilitarian?—effect. No wonder, then, that criminal justice actors are disposed to focus on rule-following, not on whether they achieve the desired result.


50. See infra notes 64, 127, 174 and accompanying text.

51. See Levenson, supra note 39, at 670–72 (criticizing the strict procedural bars that apply in the habeas corpus context for favoring finality over accuracy); Simon, supra note 39 (explaining that the goal of the criminal justice system is due process, not accurate outcomes). See also STUNTZ, supra note 11, at 79–80, 227–30, 260–65 (arguing that the Warren Court’s focus on procedure as opposed to accuracy “siphon[ed] the time of attorneys and judges away from the question of the defendant’s guilt” and gave legislators incentives to avoid procedural protections by criminalizing trivial behavior (e.g., possession of small amounts of drugs), increasing punishments, and resisting funding for indigent defense).
This Essay urges criminal courts to adopt an increasingly common strategy for avoiding the problems of rule-based bureaucracy, by (1) working to establish the conditions just mentioned—a solid, results-oriented conception of success, a strong motivation to inquire honestly into whether success has been achieved, and valid measures of whether it has; then (2) using those tools to evaluate the criminal justice systems’ collective experience and results; and (3) using those findings to adjust actions and improve results. We argue that the tools needed to implement this strategy are readily at hand and that the main obstacle to the strategy is an individualistic or libertarian conception of criminal process that, despite its intersection with the Criminal Procedure Revolution, is inconsistent with the broadest extension of the right of defendants, victims, and the public to accurate assessments of guilt and punishment and can be relinquished without risking the protections associated with that Revolution.

Part II describes the recent turn from rule-bound bureaucracy to rigorous analysis and response to collective experience in many spheres of activity and the resulting collapse of the distinction between operations and evaluation—administration and review—in favor of treating operations as continuous self-evaluation, adjustment, and improvement. Part III offers reasons why blending operations and evaluation through the systematic monitoring of results should come easily to courts. Part IV gives three illustrations of the criminal courts’ contrary tendency to resist systematic self-evaluation. Part V describes data-driven strategies the courts could adopt. Part VI addresses three objections to such strategies, ultimately tracing them to a benighted individualistic ideal that it urges criminal justice actors to abandon.

II. RIGOROUS SELF-EVALUATION AND ADJUSTMENT

The bureaucratic method of regulating activity to minimize harms is to specify an ameliorative step that every actor must take under preordained

---

52. See, e.g., Ansell, supra note 47, at 5–7, 175–83, 187–96 (describing numerous examples of public agencies using pragmatic problem-solving approaches in lieu of bureaucratic rule enforcement to solve public problems); Coglianese & Lazer, supra note 47, at 699–706 (explaining what management-based regulation is and when it is effective); Sabel & Simon, supra note 47, at 78–82 (providing examples of more flexible, less rule-based “experimentalist” methods of tackling public problems that previously had been impervious to bureaucratic solutions); James S. Liebman, Ending the Great School Wars, EDUC. WEEK (Dec. 11, 2012), http://www.edweek.org/ew/articles/2012/12/12/14liebman_ep.h32.html?tkn=UPLFHjswshh9%2BVL8j5dHAB4mAuFz1KAN%2By2q&cmp=clp-edweek (explaining how new institutional-learning strategies are being used to improve student results in what previously were poorly performing, highly bureaucratized public school systems).
circumstances. For example, the government, or a company with a large fleet of vehicles, may require all automobiles with internal combustion engines to have catalytic converters or require vehicles to run on natural gas or electricity. A municipality or a private condominium’s governing board may mandate that all windows in residences above the ground floor where children are present have specified window guards. Or, finally, the Supreme Court or state or local courts may prescribe steps that prosecutors, defendants, defense counsel, and courts must follow to resolve a felony charge by guilty plea. Such top-down, often intricate, requirements and arrangements work well when the harms addressed, conditions under which the harms arise, and regulatory costs are predictable and consistent across most sites or cases. But in many situations, one or more of those prerequisites is absent with the result that bureaucratic or technology-based regulations lead to inefficient over- or underregulation. Additional flexibility may be obtained by specifying a desired outcome and letting regulated actors decide for themselves how to reach or stay short of the specified level. Examples include limits on pollution emissions, the controlled allocation of tradable pollution permits, and rewards or consequences for teachers whose students’ rates of improvement on prescribed tests fall above or below a set level. Here, too, however, 

53. See Coglianese & Lazer, supra note 47, at 694, 701 (describing and criticizing this form of bureaucratic, or “technology-based,” regulation).

54. See, e.g., 42 U.S.C. §§ 7521(m), 7522(a) (2006) (requiring cars to have catalytic converters, requiring state inspection to assure the converters work properly, and limiting steps drivers and repair shops can take to modify the converters); Coca-Cola Joins the National Clean Fleets Partnership, COCA-COLA CO. (July 20, 2011), http://www.coca-colacompany.com/press-center/press-releases/coca-cola-joins-the-national-clean-fleets-partnership (announcing company’s commitment to reducing packaging weight, adopting “ecofriendly driving practices,” and “deploying diesel-electric hybrid trucks”).


57. See Coglianese & Lazer, supra note 47, at 705 (explaining why technology-based regulation is “a viable . . . strategy” only when most regulated “actors have similar operations” and the technology they use “tends to be stable over time”).

58. Id. at 701, 705.

59. See id. at 700–01 (limiting workers’ exposure to hazardous chemicals to a specified maximum level but not specifying the particular technology that regulated firms are required to use to
inconsistency, unpredictability, and variable costs across time and location can render the minimum or maximum level inappropriate in many circumstances, again causing inefficient under- or overregulation.\textsuperscript{60}

A third approach to regulation is to identify a desired direction of change and to ask actors to measure, explain, and develop site-specific plans for improving their outcomes with respect to that goal and to adjust actions and expectations based on what their own and other sites’ experience reveals is possible.\textsuperscript{61} This regulatory strategy minimizes inefficient under- and overregulation in two ways. First, it lets regulated parties customize their remedies to the precise conditions they face. Second, it lets officials compare the effectiveness of steps taken at similar sites to reveal effective technologies or targets that can serve as default solutions pending the results, or as rigorous benchmarks for evaluating the success, of subsequent rounds of regulatory experimentation.\textsuperscript{62}

diminish exposure affords firms flexibility in finding “less costly ways to achieve [those] performance levels”); Paul Krugman, \textit{Building a Green Economy}, N.Y. TIMES MAG. (Apr. 7, 2010), http://www.nytimes.com/2010/04/11/magazine/11Eonomy-t.html?pagewanted=all&_r=0 (discussing how market-based, cap-and-trade initiatives can achieve positive results in lessening climate change); Eric A. Hanushek, \textit{Teacher Deselection}, in \textit{CREATING A NEW TEACHING PROFESSION} 165, 165–78 (Dan Goldhaber & Jane Hannaway eds., 2010) (proposing that school systems improve student outcomes by setting minimum levels of improvement in test scores that students are expected to achieve on average each year and by dismissing teachers whose students repeatedly fail to make the desired gains).

\textsuperscript{60} See, e.g., EVA L. BAKER ET AL., ECON. POLICY INST. BRIEFING PAPER #278: PROBLEMS WITH THE USE OF STUDENT TEST SCORES TO EVALUATE TEACHERS 5–7 (2010), available at www.epi.org/files/page/-/pdf/bp278.pdf (arguing that “approaches to teacher evaluation that rely heavily on test scores can lead to narrowing and over-simplifying the curriculum, and to misidentifying both successful and unsuccessful teachers”); Coglianese & Lazer, supra note 47, at 701 (describing inefficiencies arising from performance-based regulations that require heterogeneous entities to achieve identical outcomes); Sabel & Simon, supra note 47, at 65–69 (discussing the inflexibility and ineffectiveness of some “cap and trade” regimes for regulating pollution); Stanley Reed & Mark Scott, \textit{In Europe, Paid Permits for Pollution Are Fizzling}, N.Y. TIMES (Apr. 21, 2013), http://www.nytimes.com/2013/04/22/business/energy-environment/europes-carbon-market-is-sputtering-as-prices-dive.html?pagewanted=all (describing the uncertainty and volatility in the political and real economy affecting European tradable carbon pollution permits, generating concerns that the carbon-permits regime “is not doing its job [of] pushing polluters to reduce carbon emissions”).

\textsuperscript{61} See Coglianese & Lazer, supra note 47, at 695–96 (explaining why “plan-based” or “management-based” approaches to regulation may be less costly and more effective than government-imposed regulatory standards because “they place responsibility for descisionmaking with those who possess the most information about risks and potential control methods”).

\textsuperscript{62} See id. at 702–03 (explaining how the flexibility that management-based regulation gives to firms to adopt solutions customized to their particular situations can result in “reducing the social harms that motivate regulation in the first place”); James S. Liebman & Charles Sabel, \textit{The Federal No Child Left Behind Act and the Post-Desegregation Civil Rights Agenda}, 81 N.C. L. REV. 1703, 1708–15 (2003) (discussing “New Accountability” reforms to public education systems under which administrative and pedagogical “rules are in effect provisional frameworks for action that are corrected” in light of evidence that they are not effective or that other approaches are more effective); Charles F. Sabel, \textit{Rolling Rule Labor Standards: Why Their Time Has Come, and Why We Should Be Glad of It}, in
This iterative and data-driven strategy has been used in many domains recently to regulate complex operations internally and externally. Toyota established a dominant position in the automobile industry by quickly diagnosing and correcting even the most modest production-line disruptions. From the start, the Nuclear Navy—and, in the wake of the Three Mile Island disaster, the United States nuclear power industry—achieved exemplary safety records by closely monitoring and broadly disseminating information about ameliorative steps for major and minor system and operations failures, near-misses, and inaccurate predictions of how designs and materials would function in practice. The commission investigating the Deepwater Horizon oil spill recommended the same approach for deep-sea drilling. By rapidly observing the effects of, then adjusting, ensembles of steps designed both to exploit and to preserve environmentally sensitive lands, collaborating developers, ranchers, first peoples, environmentalists, and public officials have found ways to use lands productively while protecting endangered species and reducing conflict. Counting and comparing relapse events has enabled drug courts to make better decisions about which probationary offenders and service providers to continue in their programs. The United States Department of

63. See, e.g., STEVEN J. SPEAR, THE HIGH VELOCITY EDGE: HOW MARKET LEADERS LEVERAGE OPERATIONAL EXCELLENCE TO BEAT THE COMPETITION 161–91 (2009) (explaining how Toyota organized all of its activities to maximize its ability to detect and fix production imperfections immediately as a way to make its manufacturing operations more efficient than those of other automakers).

64. See, e.g., JOSEPH REES, HOSTAGES OF EACH OTHER: THE TRANSFORMATION OF NUCLEAR POWER SAFETY AFTER THREE MILE ISLAND 1–7, 91–150 (1994) (discussing the effectiveness of regulatory mechanisms developed by the nuclear power industry after the Three Mile Island disaster); SPEAR, supra note 63, at 109–40 (discussing the “nuclear Navy”).


67. See, e.g., ANSELL, supra note 47, at 6, 92 (“[C]ourts have developed criteria to track the
Justice and private philanthropies have conditioned funding for local juvenile-justice agencies on data demonstrating the effectiveness of plans for reducing racial disparities at each stage of the process, identified effective practices, and reduced juvenile incarceration rates.68 Similarly iterative and data-attentive steps have limited consumers' exposure to toxins from the production and distribution of meat and leafy green vegetables;69 improved the learning outcomes of public school children from Auckland and Helsinki to New York City;70 reduced safety incidents as a strategy for improving the efficiency of aluminum production;71


69. Sabel & Simon, supra note 68, at 1278–85.


71. See, e.g., SPEAR, supra note 63, at 88–108 (describing major aluminum manufacturer Alcoa's monitoring and immediate implementation of local solutions for every harmful accident and
enhanced the autonomy and comfort of nursing home residents in Australia\(^7\) and the stability of in-home and foster placements for abused and neglected children in Alabama and Utah;\(^7\) provided disciplining and trust-building devices for producer-supplier and inter-divisional partners in the design of computers, jet engines, and large farm equipment when contractual price and quantity terms are unavailable for those purposes;\(^7\) helped urban police forces deploy law-enforcement resources where they are most needed;\(^7\) and identified tuna-fishing methods that are least likely to harm dolphins.\(^7\)

All these endeavors put carefully observed collective experience ever more systematically and effectively at the service of future improvement. Typically, organizations mine “Big Data” for patterns of error in more and more precisely defined sets of circumstances in search of predictive and diagnostic information about how to adjust operations to reduce error, then collect more data on the effect of each adjustment to allow still swifter and near-miss and sharing of solutions with other factories as a way to improve safety and productivity).


75. See, e.g., ANSELL, supra note 47, at 47–54, 108–15 (discussing CompStat and community policing); JOHN BUNTIN, HARVARD UNIV. KENNEDY SCH. GOV., ASSERTIVE POLICING, PLUMMETING CRIME: THE NYPD TAKES ON CRIME IN NEW YORK CITY, CASE PROGRAM C16-99-1530.0 (1999) (discussing implementation of CompStat in New York City); COLLEEN MCCUE, DATA MINING AND PREDICTIVE ANALYSIS: INTELLIGENCE GATHERING AND CRIME ANALYSIS 203–05 (2007) (discussing automated classification techniques used to facilitate investigations of drug-related crimes); ARCHON FUNG, EMPOWERED PARTICIPATION: REINVENTING URBAN DEMOCRACY 173–97 (2004) (discussing community policing in Chicago); Tina Rosenberg, Armed with Data, Fighting More than Crime, N.Y. TIMES OPINIONATOR (May 2, 2012, 7:00 AM), http://opinionator.blogs.nytimes.com/2012/05/02/armed-with-data-fighting-more-than-crime/?_r=0 (discussing implementation of CompStat and CitiStat in Baltimore); Somini Sengupta, In Hot Pursuit of Numbers to Ward Off Crime, N.Y. TIMES (June 19, 2013), http://bits.blogs.nytimes.com/2013/06/19/in-hot-pursuit-of-numbers-to-ward-off-crime/?_r=0 (“At the Seattle Police Department, the morning roll call begins with Google Maps and computer algorithms,” and police use the results of “software that crunches crime data and tries to predict where crimes are most likely to occur over the next few hours.”).

76. See, e.g., Sabel & Simon, supra note 68, at 1292–97 (explaining why the monitoring of steps tuna-fishing operations chose to reduce dolphin mortality under a general international mandate worked better than command-and-control regulation).
more targeted correction and fine-tuning. As the New York Times reported recently, "The story is the same in one field after another, in science, politics, crime prevention, public health, sports and industries as varied as energy and advertising. All are being transformed by data-driven discovery and decision-making" based on "analysis rather than experience and intuition—more science and less gut feel."

Evidence that the courts' rule-based regulation of criminal justice is failing puts criminal justice at the same crossroads these other operations traversed when deciding to supplant bureaucracy with data-driven self-reflection and adjustment. Below we encourage the courts to take the same path. Because doing so jettisons the distinction between operations and quality control in favor of rigorously measuring ongoing operations to assure quality during, not after, the fact, we first address a threshold objection: that the separation of powers requires courts to preserve the distinction between operations (for which the executive and administrators are presumptively responsible) and after-the-fact quality control (the work of the courts).

III. THE COURTS' SIMULTANEOUSLY OPERATIONAL AND REGULATORY ROLE IN CRIMINAL CASES

It might seem that the responsibility for the self-regulatory steps described above falls more naturally on and that the separation of powers

77. Steve Lohr, Sizing Up Big Data, Broadening Beyond the Internet, N.Y. TIMES (June 20, 2013, 11:09 PM), http://bits.blogs.nytimes.com/2013/06/19/sizing-up-big-data-broadening-beyond-the-internet/ ("Big Data . . . means three things. First, it is a bundle of technologies. Second, it is a potential revolution in measurement. And third, it is a point of view or philosophy about how decisions . . . should be . . . made.").

78. Id.


80. See supra notes 39–46 and accompanying text.

81. See, e.g, SPEAR, supra note 63, at 155–91 (discussing Toyota's self-conscious monitoring and continuous improvements of operations).
assigns them exclusively to police departments, forensic laboratories, prosecuting offices, and public defense organizations, and not the courts. Although we hope those actors take such steps—as many already do in contexts not regulated by the courts—giving them sole responsibility is misguided. As noted, the courts minutely govern large swaths of the administration of criminal justice. Law enforcement agencies are unlikely to innovate freely in those contexts, therefore, absent the courts’ exemption from those rules, endorsement of rigorous evidence-based action, and instructive example.

One might expect courts to be sympathetic to using the results of their actions to improve the administration of justice. In part, this is because judicial review in the criminal justice context differs from that in other domains. If, in Robert Cover’s terms, what courts do in constitutional regimes is review and constrain the “violence” that government actors inflict on people and communities by limiting their freedom to make meaning out of their lives, then the violence courts review in the criminal context—prosecuting, trying, imprisoning, and executing people and sometimes letting known and violent criminals escape punishment—is different, because the violence is substantially the courts’ own. Criminal courts often, therefore, simultaneously administer and regulate violence, so routinely using regulatory learning to improve administration should seem obvious and natural.

The role of appellate courts in death penalty cases is a stark example. Nearly all states forbid capital verdicts to be carried out until they are affirmed by the state high court on direct review, whether or not the convicted defendant opts to appeal. The appellate court’s role, thus, is

82. See supra notes 67–68, 75 (discussing data-driven techniques police, prosecuting offices, and courts use to fight crime, handle drug offenders, and minimize racial disparities in juvenile justice); Brandon Garrett, Remediing Racial Profiling, 33 COLUM. HUM. RTS. L. REV. 41, 107–47 (2001) (discussing similar techniques for combating racial profiling while improving law enforcement).

83. See supra notes 19–38 and accompanying text.


86. See Whitmore v. Arkansas, 495 U.S. 149, 173–74, 174 n.1 (1990) (Marshall, J., dissenting) ("[A]lmost all of the 37 [s]tates with the death penalty apparently have prescribed mandatory,
both to exercise quality control over the trial court’s product and to add something essential of its own. Rulings on stays of execution are similar. An essential step on the way to an execution is the telephonic “go-ahead” from the state high court and the United States Supreme Court. Technically, of course, the courts are refusing to stay a process otherwise outside their control. But that process is the execution of a court order; from the warden’s and public’s perspective, the phone call is necessary to deliver a superior court’s okay to execute the order and take the prisoner’s life.17

The same principle holds for the most mundane acts of criminal judges. Much of what judges do in the process of preparing for and conducting criminal trials and appeals is to decide whether to admit or exclude evidence. These acts are not just administrative steps to resolve the dispute before the court, but also regulatory steps to influence the conduct of police, prosecutors, defense attorneys, and lower courts in future cases. This fact is explicit when courts administer the Fourth and Fifth Amendment exclusionary rule with a focus not on the outcome of the case at hand but on the future behavior of police officers and on assuring that judges do not become “accomplices in the willful disobedience of [the] Constitution.” Conversely, rulings admitting evidence obtained through police searches and interrogations that are unsuccessfully challenged as unconstitutional not only put evidence before the trier of fact but also invite the continuation of the challenged practices. Mapp, however, did not invent the regulatory role of evidentiary rulings. It simply adapted that tool to the enforcement of constitutional rights. Every criminal trial and appellate judge recognizes, therefore, that most actions they take are simultaneously administrative or dispute resolving with respect to the case at hand and regulatory with respect to the ongoing collection, preservation, and presentation of evidence in criminal cases generally. Actors in the criminal justice system—judges included—are expected to learn something from each ruling about how future cases should be handled.

Nor would one expect American courts to resist learning from the aggregate trend of their own and other courts’ rulings or from the run of nonwaivable appellate review of at least the sentence in capital cases.”).  

87. See Liebman, supra note 85, at 97 (“[T]he medium for giving the final order to kill—the court’s telephone call to the warden announcing the denial or lifting of a stay—dramatically ‘renders the execution constitutional violence’ by showing the violence to be the judge’s, not the executioner’s.”).

actions by police, prosecutors, and defense counsel in cases similar to the one at hand. Pattern-focused analysis of an aggregation of instances is, after all, the essence of the common law method and of courts’ reliance on standard commercial, professional, and community practices to resolve a myriad of cases.

IV. JUDICIAL RESISTANCE TO DATA-DRIVEN SELF-REGULATION OF CRIMINAL JUSTICE

As this part’s discussion of three criminal law curiosities illustrates, however, American criminal courts viscerally resist the aggregative, self-regulative analysis of their own and other criminal-justice players’ results despite strong evidence of the value of such analysis. Courts have eschewed pattern-focused review of death sentences, ignored error rates in capital cases, and discounted the summed effect of match and non-match evidence of the identity of criminals.

A. (NON)REVIEW OF DEATH-SENTENCING PATTERNS

If asked to name the Supreme Court’s lodestar death penalty case, any expert will probably cite Furman v. Georgia, which in 1972 overturned every capital-sentencing statute in the nation as excessively discretionary. This is itself an oddity. Furman’s outcome was premised on five separate and conflicting opinions, none joined by more than a single Justice. In the three opinions most often used to navigate the Court’s jurisprudence, Justice White concluded that there were too few death sentences to provide a deterrence or retribution, Justice Stewart reasoned that death sentences were too many and too capriciously distributed to signify the worst of the worst, and Justice Douglas ignored the proportion of crimes ending in death verdicts and focused on the verdicts’ misdistribution between, for

89. See, e.g., PETER L. STRAUSS, LEGAL METHODS: UNDERSTANDING AND USING CASES AND STATUTES 5 (2005) (defining the common law as an authoritative body of law that “draws its force from” reasoned explanations of conclusions reached “in relation to the existing body of decisions”).

90. See, e.g., Roth v. United States, 354 U.S. 476, 489 (1957) (using “contemporary community standards” to determine whether sexually explicit materials fall within the “obscenity” exception to First Amendment protection); Morisch v. United States, 653 F.3d 522, 531 (7th Cir. 2011) (requiring malpractice plaintiff to show that the medical practitioner failed to exercise the standard of care common in the relevant medical community).


92. Furman, 408 U.S. at 257 (Brennan, J., concurring); id. at 314 (Marshall, J. concurring); id. at 240 (Douglas, J., concurring); id. at 306 (Stewart, J., concurring); id. at 310 (White, J., concurring).
example, African American and white defendants.\textsuperscript{93}

The only respect in which the three opinions agree is their decisionmaking method. Each Justice based his conclusions on an evaluation of aggregate death-sentencing patterns observed over years. As Justice White explained, for example: "I can do no more than state a conclusion based on 10 years of almost daily exposure to the facts and circumstances of hundreds and hundreds of federal and state criminal cases involving crimes for which death is . . . authorized . . . that the death penalty is exacted with great infrequency . . . and that there is no meaningful basis for distinguishing the few cases in which it is imposed from the many cases in which it is not."\textsuperscript{94}

Four years later in \textit{Gregg v. Georgia}, as states struggled to develop death-sentencing procedures that satisfied \textit{Furman}, the Court upheld Georgia's new statute, pointing the way for other States.\textsuperscript{95} The Court approvingly noted that the Georgia Legislature had taken the Court's pattern-focused methodology to heart and required the state supreme court to conduct comparative "proportionality review" of every death sentence.\textsuperscript{96} Such review "inquire[s] . . . whether the penalty is . . . unacceptable in a particular case because [it is] disproportionate to the punishment imposed on others convicted of the same crime."\textsuperscript{97} Trial judges were required to file reports on every case in which death was a potential penalty with comprehensive data on the crime, offender, victim, and verdicts so the state supreme court could determine whether death sentences were consistently imposed in cases similar to the one under review.\textsuperscript{98} The Supreme Court lauded Georgia's aggregative analysis "as a check against the random or arbitrary imposition of the death penalty . . . which substantially eliminates the possibility that a person will be sentenced to die by the action of an aberrant jury."\textsuperscript{99} Many states emulated Georgia's comparative

\textsuperscript{93} \textit{Furman}, 408 U.S. at 255-57 (Douglas, J., concurring); \textit{id.} at 309-10 (Stewart, J., concurring); \textit{id.} at 311-12 (White, J., concurring). This analysis is developed further in Liebman, supra note 85, at 7-14; James S. Liebman & Lawrence C. Marshall, \textit{Less Is Better, Justice Stevens and the Narrowed Death Penalty}, 74 FORD. L. REV. 1607, 1612-48 (2006).

\textsuperscript{94} \textit{Furman}, 408 U.S. at 313 (White, J., concurring).

\textsuperscript{95} \textit{GA. CODE. ANN. § 17-10-35(c)(3)} (West 2006) (requiring Georgia Supreme Court to determine "[w]hether the sentence of death is excessive or disproportionate to the penalty imposed in similar cases, considering both the crime and the defendant"); \textit{Gregg v. Georgia}, 428 U.S. 153 (1976).

\textsuperscript{96} \textit{Gregg}, 428 U.S. at 204-06.


\textsuperscript{98} \textit{Gregg}, 428 U.S. at 166-68 (discussing \textit{GA. CODE. ANN. § 27-2537}, superseded in 2006 by the identical \textit{GA. CODE. ANN. § 17-10-35(c)(3)}).

\textsuperscript{99} \textit{id.} at 206.
proportionality provisions, and when the Georgia statute again came under constitutional attack in 1983, the Court again emphasized the importance of comparative review to the statute's constitutionality.

Starting in 1984, however, the Court gradually turned against the aggregative methodology that is the only unifying feature of Furman's lodestar opinions. In Pulley v. Harris, the Court ruled that the Constitution did not require capital states to conduct comparative proportionality review. Three years later, Warren McCleskey, a black man sentenced to die for killing a white police officer in a robbery, presented a study of hundreds of attributes of the crime, offender, victim, and courts in all potentially capital murder cases processed under Georgia's death-sentencing procedures. Controlling for those factors, the study showed that the state was 4.3 times more likely to sentence a defendant to death for killing a white victim than a black victim. Accepting the study as statistically valid—it of course was infinitely more informative than the unrefined data on which the Justices had relied in Furman—the Court ruled the pattern evidence irrelevant. Apart from evidence linking a particular prosecutor, judge, or juror to one or more identifiable and consequential acts of racial discrimination, the Court ruled, a capital defendant could not constitutionally challenge the death penalty based on a state's racially identifiable death-sentencing patterns.

Finally, in a series of cases culminating in Arave v. Creech, the Court held that it would no longer examine the pattern of a state's applications of any particular capital aggravating factor to determine whether the factor satisfied the Eighth Amendment requirement that it narrow the category of death-eligible killings to something short of all first-degree murders. If the state courts gave the factor a narrowing definition,

---


101. See, e.g., GA. CODE. ANN. § 17-10-35(c)(3) (requiring comparative review of capital verdicts); Zant v. Stephens, 462 U.S. 862, 890 (1983) (affirming capital statute in part because of an "important procedural safeguard, the mandatory appellate review of each death sentence by the Georgia Supreme Court to avoid arbitrariness and to assure proportionality").

102. Pulley, 465 U.S. at 43-44.


104. Id. at 291 n.7, 313.

105. Id. at 292.


107. Id. at 476-77 (declining review of "whether state courts properly have applied an aggravating circumstance" as opposed to "whether the circumstance, as narrowed, is facially valid." (citing Lewis v. Jeffers, 497 U.S. 764, 778-80 (1990)).
the Court would not consider whether the application of that definition in the run of cases in fact confined death sentences to less than all first-degree murders. The Court then upheld Idaho’s “cold, calculated killing” factor as sufficiently narrower than “premeditated and deliberate” first-degree murder.

Curiously, therefore, although Furman remains the lodestar for determining the constitutionality of death-sentencing procedures under the Eighth Amendment, no court today could replicate that decision’s analysis or conclusion in a case presenting the same statewide death-sentencing patterns as those before the Court in Furman.

B. (IN)ATTENTION TO RATES OF ERROR IN CAPITAL CASES

A second curiosity also relates to the death penalty. In 2000 and 2002, this Essay’s senior author and colleagues issued a two-part study of the outcomes of all 4578 state direct appeals, 248 state postconviction reversals, and 599 capital federal habeas cases in which a final judgment was reached upon review of a capital verdict between 1973 and 1995. We found that over two-thirds of all death verdicts fully reviewed during the period were reversed due to serious legal error, with elected state judges responsible for 90 percent of the more than 1800 reversals. In identifying conditions associated with especially high error rates in states

108. Liebman, supra note 85, at 63–64.
110. Liebman, supra note 85, at 121–22. For another example of the Court’s more recent rejection of systematic analysis it previously deemed essential, compare Witherspoon v. Illinois, 391 US 510, 517–18 (1968) (deferring a ruling on the constitutionality of excluding prospective capital jurors from deciding guilt or innocence because they are unwilling to impose a death sentence until additional social scientific evidence revealed the impact across cases of such exclusions on the fairness and accuracy of guilt-innocence determinations), with Lockhart v. McCree, 476 U.S. 162, 173 (1986) (acknowledging an array of new studies indicating that “death-qualified” juries are substantially more prone to convict than other juries, but ruling the studies irrelevant and limiting the constitutional question to whether any juror in the particular case was individually biased against the defendant).
113. See id. at 222 (finding that state courts staffed by elected judges were responsible for 90 percent of the reversals of death verdicts); id. at 218–23 (concluding that nearly all the reversals were for errors undermining the accuracy of the verdict).
and counties, we discovered that jurisdictions imposing the highest proportion of capital sentences relative to crimes and convictions were also the most likely to infuse any given death sentence with error.\textsuperscript{114} Holding other factors equal, the probability that a death verdict would be overturned was seven times higher in states and counties with the highest capital-sentencing rates than those with the lowest.\textsuperscript{115}

At the same time, states most likely to impose death for a given crime were seven times less likely to capture, convict, and incarcerate criminals than the least death-prone capital states and spent only a third as much on their criminal courts.\textsuperscript{116} In effect, high-death-sentencing jurisdictions use the death penalty as a cheap and dirty alternative to systematic law enforcement and to the reliable prosecution of criminal cases—explaining why use of capital punishment is strongly associated with serious error but not with lower crime rates.\textsuperscript{117}

Curiously, this boring bean-counting aggregation of appellate outcomes was a lead story in hundreds of newspapers and news programs nationwide.\textsuperscript{118} Because courts do not aggregate even the most routine information about the reliability of their own verdicts in the most consequential of cases, it is big news when someone else does. Courts’ failure to compare the frequency with which different jurisdictions’ capital or other verdicts are found to be seriously flawed leaves them unaware of vast discrepancies in the probability of error from one jurisdiction to the next and unable, for example, to raise the level of scrutiny applied in reviewing verdicts from jurisdictions with a demonstrably poor track record or to impose stricter sanctions when patterns of error persist.\textsuperscript{119}

The criminal justice system does not even use readily available data to measure the errors we most care about: execution of the innocent.\textsuperscript{120} State

\begin{enumerate}
\item\textsuperscript{114} LIEBMAN ET AL., BROKEN SYSTEM, PART II, supra note 111, at 352–53.
\item\textsuperscript{115} Id.; Gelman et al., supra note 111, at 241; Liebman & Clarke, supra note 5, at 267–68.
\item\textsuperscript{116} Gelman et al., supra note 111, at 242–43.
\item\textsuperscript{117} Liebman & Clarke, supra note 5, at 273–77, 322–23.
\item\textsuperscript{119} Gelman et al., supra note 111, at 224.
\item\textsuperscript{120} For non-DNA evidence that American jurisdictions have executed innocent prisoners over
forensic labs and police lockers nationwide contain physical evidence with the never-tested DNA of the perpetrators of hundreds of capital crimes for which men and women have been executed since 1976. But freedom-of-information laws do not apply, and courts and law enforcement agencies have declined to conduct DNA tests or allow others to do so in all but three cases—one of which confirmed guilt, another of which was inconclusive, and the most recent of which pointed to innocence. Unlike other big operations that risk "innocent" life, that is, courts and law enforcement agencies remain utterly uninformed about what their own data show about their rates of risking and taking the lives of innocent people.

the last quarter century, see, for example, James S. Liebman et al., Los Tocayos Carlos, 43 Colum. Hum. Rts. L. Rev. 711, 1103–04 (2012) (reconstructing the process through which Texas executed Carlos DeLuna for a crime to which another man repeatedly confessed and very probably committed); Raymond Bonner & Sara Rimer, A Closer Look at Five Cases That Resulted in Executions of Texas Inmates, N.Y. Times (May 14, 2000), http://partners.nytimes.com/library/politics/camp/051400whbush-cases.html?scp-1&sq=%22david%20spence%22&st=cse (reporting that two of six jailhouse informants who testified against David Spence at his capital trial recanted their stories since he was executed in 2000 and that experts studying bite marks used as evidence against Spence concluded that the marks do not match Spence’s teeth); Terry Ganey, After Execution, Case Is Reopened, St. Louis Post-Dispatch (July 12, 2005), http://business.highbeam.com/435553/article-lGI-1339642625/after-execution-case-reopened (reporting that the critical witness who identified executed inmate Larry Griffin as the perpetrator of a street shooting had not been at the crime scene and that a man who was present but did not testify at trial knew Griffin and was sure he was not one of the shooters); David Grann, Trial by Fire: Did Texas Execute an Innocent Man?, New Yorker (Sept. 7, 2009), http://www.newyorker.com/reporting/2009/09/07/090907fa_fact_grann (casting doubt on evidence used to conclude that a house fire that killed Cameron Todd Willingham’s daughters was intentionally set, rather than accidental and exposing the scientifically unsound methods used to conclude that the fire was set intentionally); Lise Olsen, Did Texas Execute an Innocent Man?, Hous. Chron. (Nov. 20, 2005), http://www.chron.com/disp/story.mpl/front/3472872.html (raising doubts about the execution of Ruben Cantu, who was implicated by a single eyewitness who has since recanted his claim that Cantu was the killer and presenting a sworn statement by Cantu’s codefendant, who was convicted but not executed for the crime, that Cantu was not involved).

121. Murphy, supra note 39, at 645–52.


123. See Tests Inconclusive in Murder Case, Associated Press (Dec. 12, 2000), http://www.apnewsarchive.com/2000/Tests-Inconclusive-in-Murder-Case/ide9a2e62144737050cb8e7e6f528ca781 (reporting results of DNA tests arranged by news organizations after the execution of Ellis Wayne Felker, which found that tests comparing Felker’s DNA to that in fingernail scrapings from the victim were “inconclusive”).

124. See Dave Mann, DNA Tests Undermine Evidence in Texas Execution, Tex. Observer (Nov. 11, 2010), http://www.texasobserver.org/cover-story/texas-observer-exclusive-dna-tests-undermine-evidence-in-texas-execution (reporting results of postexecution DNA testing revealing that a hair found at a robbery-murder crime scene—which was described at Claude Jones’s trial as “matching” his hair based on since-discredited microscopic hair analysis and which provided the main evidence that Jones was the killer—was not Jones’s hair).

C. (DIS)COUNTING MATCH AND NON-MATCH EVIDENCE OF IDENTITY

Many modern operations collect and adjust strategies based on systematically analyzed information about their own collective experience. These analyses focus not only on large system failures but also on minor errors and "near misses," which can serve as canaries in the mine, signaling impending disasters. Because courts and other criminal justice actors rarely conduct such analyses, the best information we have comes from media and academic studies of prisoners exonerated by DNA or otherwise and occasional inquiries by law enforcement agencies into how innocent men and women became lead suspects or were convicted. Commentators' main focus in discussing these exonerations and reinvestigations is the mishandling of what might be called "big" evidence: confessions, eyewitness identifications, jailhouse informant testimony, fingerprints, and other evidence that exerts a major influence on jurors.

When DNA testing or confessions expose the actual offender, exonerations and reinvestigations also, however, highlight the mishandling of a kind of "small" evidence, which actually pointed to innocence from the

---


126. See supra notes 62–76.
127. For examples of operations that monitor minor incidents and/or near-misses for information about disaster-in-the-making, see sources cited supra notes 50, 63–64, 71 and infra note 174.
128. See, e.g., supra notes 41, 117–20 and accompanying text.
129. See, e.g., James S. Liebman et al., The Evidence of Things Not Seen: Non-Matches as Evidence of Innocence, 98 IOWA L. REV. 577, 647–48 (2013) (discussing FBI inquiry into how it falsely concluded that Oregon lawyer Brandon Mayfield was responsible for a terrorist attack in Madrid, Spain, based on faulty fingerprint analysis); Henry Lee, How Innocent Man's DNA Was Found at Killing Scene, SFGATE (June 26, 2013, 11:07 PM), http://www.sfgate.com/default/article/How-innocent-man-s-DNA-was-found-at-killing-scene-4624971.php (describing arrest of Lukis Anderson based on match between his DNA and fingernail scrapings from the murder victim, followed by his release five months later because of incontrovertible proof that he was drunk and unconscious in a hospital when the killing occurred and discovery that the same paramedics who transported Anderson to the hospital also moved the body of the murder victim hours later).
130. Liebman et al., supra note 129, at 583–84.
start but was ignored. These are traces found at the crime scene by police that were ignored because they did not match the defendant but later proved to be relevant because they do match the real culprit. For example, a reinvestigation of Texas’s 1989 execution of Carlos DeLuna for a crime a lookalike named Carlos Hernandez almost certainly committed, documented not only a likely mistaken, “big” cross-ethnic identification of DeLuna made under intensely suggestive circumstances, but also, buried in police reports and photos, twenty-two “small” markers potentially pointing to the culprit that were ignored because they did not match Carlos DeLuna. What the jury never knew, because it never learned of Carlos Hernandez’s existence, is that ten of those features matched Hernandez. So did all seven additional bits of small evidence in the case that were emphasized at trial because they matched DeLuna. None of the twenty-nine features matched DeLuna but not Hernandez.

We call this kind of evidence, which is present in most identity cases, “non-exclusionary non-matches.” It includes any trace that may or may not have been left behind or taken from the scene by the culprit and does not match the defendant. An example is a fingerprint on a convenience store counter that matches neither the robbery suspect nor the store’s employees. The “non-matching” fingerprint is “non-exclusionary” because it does not rule out the suspect: the culprit may have left the print (in which case the suspect in a single-perpetrator crime is innocent), or the print may

131. Id. at 580–81.
132. For examples, see id. at 580–87.
133. Id. at 585–87. Hernandez repeatedly confessed to family and friends that he, not his lookalike Carlos DeLuna, committed the crime. See Liebman et al., supra note 120, at 876–83.
134. Three of the twenty-nine clues matched neither DeLuna nor Hernandez, and nine were never tested to determine whom they matched. Liebman et al., supra note 129, at 585–86.

Our count of twenty-nine matchable features actually should be considerably higher. This is because, after being stabbed with a large buck knife, the victim struggled with the assailant as she bled to death, drenching the four-foot by six-foot crime scene in pools, drops and spatters of blood. None of the blood was collected or tested because the police assumed it all was the victim’s. Had police collected and tested all the blood traces — greatly increasing the count of matchable clues at the crime scene — one or more of them might have conclusively identified the killer. For a case in which the belated analysis of blood evidence had exactly this effect and may lead to an exoneration, see Lizette Alvarez, Lawyers Say DNA Clears Florida Inmate in Two Killings, N.Y. TIMES (May 13, 2013), http://www.nytimes.com/2013/05/14/us/lawyers-say-dna-clears-florida-inmate-in-killings.html?_r=0 (reporting that a DNA analysis of blood spatter collected at a crime scene seven years earlier revealed that, although none of the blood came from a man who spent seven years on death row for the crime and most of the blood was from the victims, some of the blood belonged to a mentally disturbed relative of the victims with a history of animosity toward the victims).

135. Liebman et al., supra note 129, at 584–85.
136. Id. at 585.
have been innocently left by a customer.\textsuperscript{137} Other examples include features of the offender remembered or misremembered by eyewitnesses that do not match the suspect (age, height, eye color, clothing, cigarette brand, facial tick), detritus found at but foreign to the crime scene that the culprit may or may not have left behind (a button, soil, carpet fiber, or animal hair), or similar items that are native to the crime scene and may or may not have been intentionally or accidentally carried off by the culprit.\textsuperscript{138}

A single non-exclusionary non-match often is unimpressive because of the high probability that it has nothing to say about who committed the crime.\textsuperscript{139} But when Bayes' Theorem is used to measure their aggregate effect, multiple non-exclusionary non-matches can easily add up to a reasonable doubt.\textsuperscript{140} Assuming independence,\textsuperscript{141} Bayes' Theorem calculates the impact of a new piece of information on the preexisting odds by multiplying those odds times a "likelihood ratio" associated with the new evidence.\textsuperscript{142} The likelihood ratio is a mathematical statement of the new information's probative weight.\textsuperscript{143} For instance, given a low probability of contracting an infectious disease (say, one-half of 1 percent, meaning the prior odds of infection are one to 199), Bayes' Theorem explains why even if you test positive for the disease using a 99-percent-accurate test, the odds are only one-to-two that you are actually infected (or two-to-one that you are not infected):

\[
\frac{\text{prior odds of infection} \times \frac{\text{probability you would test positive if infected}}{\text{probability you would test positive if not infected}}}{\text{subsequent odds of infection}} = \frac{1}{199} \times \frac{1}{0.01} = \frac{1}{199} = 33\% \text{ (odds restated as a probability)}
\]

In a recent article, we develop a stylized example of the potential effect of multiple non-exclusionary non-matches.\textsuperscript{144} We imagine a jury 98 percent certain of the defendant's guilt after hearing that an eyewitness identified the defendant and learning that the tops of the robbery-murder victim's nylon stockings were missing and that police found several knotted stocking tops in the suspect's dresser when they arrested him.

\textsuperscript{137} Id. at 584–85.
\textsuperscript{138} Id.
\textsuperscript{139} Id. at 610.
\textsuperscript{140} Id. at 587–88.
\textsuperscript{141} See infra notes 155–57 and accompanying text.
\textsuperscript{142} Liebman et al., supra note 129, at 593. Applied to evidence of identity, the likelihood ratio is the probability that the new evidence would be present if the defendant were guilty divided by the probability that the same evidence would be present if the defendant were not guilty.
\textsuperscript{144} See Liebman et al., supra note 129, at 588–97 (using a hypothetical example to demonstrate the value of multiple non-exclusionary non-matches as evidence capable of establishing a reasonable doubt about guilt).
Using Bayes’ Theorem, we then estimate the impact of five non-exclusionary non-matches: (1) a police investigator’s initial theory that a left-handed assailant delivered the fatal blows, though the defendant is right-handed, reduces the 98 percent prior probability of guilt to 94.1 percent; (2) a neighbor’s description of a short balding man seen near the victim’s home just before the offense, though the defendant is taller and has a full head of hair, further reduces the guilt probability to 93.7 percent; (3) a pedestrian’s testimony that she was nearly knocked over by a blue-eyed man running in the opposite direction from the crime scene after the crime occurred, though the defendant has brown eyes, reduces the probability to 88.1 percent; (4) an unsmoked cigarette found at the crime scene, though neither the victim nor the defendant smokes, reduces the probability to 83.2 percent; and (5) the eyewitness’s description to police of a brown sweatshirt worn by the man she saw at the crime scene and officers’ failure to find a brown or similar sweatshirt on the defendant or at his apartment reduces the probability to 78.5 percent, almost certainly raising a reasonable doubt."145

Despite the availability of analyses such as these, it remains the practice today, following California’s oft-cited decision in People v. Collins, to forbid judges, prosecutors, and defense counsel to help jurors aggregate the quantified weight of individual matches and non-matches between crime-scene evidence and the defendant, as an aid in solving the whodunit.146 This is so even if jurors (1) are only invited and not required to use Bayesian analysis and are instructed in its use via explanations proven to enable lay persons to use the theorem reliably;147 (2) are allowed to estimate the prior-odds and the likelihood ratio’s numerator and denominator themselves in the usual way jurors assess evidence; (3) are aided in making estimates by the introduction of population frequencies (e.g., of left-handedness and cigarette smoking) from reliable time- and place-specific censuses of human features and elective behavior (e.g., wearing brown sweatshirts at a particular time and place) that ubiquitous security cameras and consumer information services now make possible;148

145. Id. at 590–95.
146. See id. at 597–600, 670–73 (discussing People v. Collins, 438 P.2d 33 (Cal. 1968)).
147. See infra note 213 and accompanying text.
148. See, e.g., United States v. Jones, 132 S. Ct. 945, 963 (2012) (Alito, J., concurring) (describing massive amounts of data collected on individuals’ whereabouts by “automatic toll collection systems” and “wireless carriers [that] track and record the location of users” of wireless devices); id. at 957 (Sotomayor, J., concurring) (describing cellular and Internet service providers’ and online retailers’ collection of consumers’ phone numbers, email addresses, and book, grocery, and medication choices); Liebman et al., supra note 129, at 620–22 (discussing breadth and acuity of information captured by security cameras); Charles Duhigg, How Companies Learn Your Secrets, N.Y. TIMES MAG. (Feb. 16,
and (4) can plug their estimates into a tool similar to a mortgage calculator that performs the Bayesian arithmetic for them.\textsuperscript{149}

Curiously, the “gold standard” of identity evidence—DNA—is a product of this same sort of aggregative analysis.\textsuperscript{150} DNA is not a “genetic fingerprint” unique to a single human being. Rather, it is a collection of genetic traits, each present in hundreds of millions or even billions of individuals, that are powerful evidence only when the aggregate probability that, say, nine or thirteen such traits would occur in a single, randomly selected individual is calculated.\textsuperscript{151} Fingerprints are not best understood as a unique trait either.\textsuperscript{152} They are a mass of unremarkable lines, swirls, and intersections at measurable distances from each other that, again, are powerful evidence only when the probability of all of them occurring together is considered.\textsuperscript{153} Indeed, as the verbatim recording of confessions becomes the norm, even this most classic form of “direct” evidence will best be understood not as a unique marker of the culprit but as circumstantial evidence whose value depends on the power of the matches and non-matches between actions and conditions described in the confession and those known to have characterized the crime but not been communicated to the confessing suspect by police.\textsuperscript{154}

True, with DNA (but not fingerprints or confessions), there is a finite number of discrete features to consider, with frequencies that we now know are nearly independent of each other.\textsuperscript{155} But it has taken twenty-five years of intensive trial and error and adversarial scrubbing to get technicians to use precise enough methods to identify the discrete features accurately and to confirm their independence.\textsuperscript{156} Because the latter uncertainty persisted for so long, statisticians developed ameliorative procedures for estimating

\textsuperscript{2012}), http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html (discussing the wide variety of information available to and used by companies to determine and influence consumer buying habits); Sengupta, \textit{supra} note 75 (discussing police use of drones and of tools that automatically analyze surveillance camera videos and comb through “data bases, from license-plate readers to sensors that can pick up heightened radiation levels to arrest records,” which can “find a car associated with a possible lawbreaker and analyze where that car has been seen over the last several weeks”).

\textsuperscript{149} See Liebman et al., \textit{supra} note 129, at 676–78 (describing commercially available Bayesian calculators).

\textsuperscript{150} \textit{Id.} at 601.

\textsuperscript{151} \textit{Id.} at 605–06.

\textsuperscript{152} \textit{Id.} at 641–42.

\textsuperscript{153} \textit{Id.} at 601, 641–42.

\textsuperscript{154} \textit{Id.} at 603–04.


\textsuperscript{156} See Liebman et al., \textit{supra} note 129, at 612–17 (describing the improvements that have taken place over the last twenty years in data-mining, statistical analysis, and laboratory practices with regard to DNA evidence).
probabilities to neutralize the effect of any interdependence that might have been present, which now can be used in aggregating the probabilities of features that certainly are interdependent.\textsuperscript{157} Nor did the courts wait to use DNA to make consequential decisions until all uncertainties were conclusively removed, including some that persist today.\textsuperscript{158} The history of DNA thus suggests that a good way to reach the gold standard is to acknowledge the value of aggregating the weight of multiple matches and non-matches and to use adversarial procedures to discipline the process.\textsuperscript{159} Doing so could improve current—often quite risky—practice not only as to confessions and fingerprints,\textsuperscript{160} but also as to large swaths of other evidence, such as a description of the culprit as a 5'8" white woman in her twenties and a red skirt, and evidence that the defendant had the same traits when seen nearby at the time.

By ignoring the aggregate effect of matches and non-matches, courts surrender potentially powerful evidence not only of a particular defendant's guilt or innocence but also of the comparative effectiveness of police and prosecutors.\textsuperscript{161} Imagine two identity cases in which five identical bits of matching evidence link the suspect to the crime scene. Imagine, as well, that police find five identical markers at both scenes that do not match the suspect, and an additional ten non-matching items at the second, but not the first, of the two scenes. Bayesian analysis can show that the suspect in the latter, five-of-twenty-matching-markers case is more probably innocent than the suspect in the five-of-ten-matching-markers case. Of course, few individual cases are likely to be this comparable, but the run of cases that different police departments and detectives face are likely to be comparable, so that a department or detective that makes arrests in cases with substantially higher proportions of non-matches on average than other agencies or officers is probably generating more faulty arrests and

\textsuperscript{157} Id. at 617.

\textsuperscript{158} Id. at 609 n.134 (discussing Virginia's 1994 execution of Timothy Spencer based largely on a DNA match made when the procedure was in its infancy and subject to multiple uncertainties and noting uncertainties that remain today). See also, e.g., Alvarez, supra note 134 (describing apparent DNA misidentification due to contamination of evidence); Osagie K. Obasogie, High-Tech, High-Risk Forensics, N.Y. TIMES (July 24, 2013), http://www.nytimes.com/2013/07/25/opinion/high-tech-high-risk-forensics.html? (discussing continuing problems with DNA, including contamination and a relatively high probability of misidentification from "cold hit" comparisons of crime-scene DNA to profiles collected in DNA databanks).

\textsuperscript{159} Liebman et al., supra note 129, at 617.

\textsuperscript{160} See, e.g., id. at 601-04 (describing the aggregative analysis responsible for the power of DNA evidence and explaining how the application of similar "disciplined measures to expose the aggregate power of matches and non-matches" could be used to increase the accuracy of confession and eyewitness testimony evidence).

\textsuperscript{161} Id. at 687.
convictions.

Police departments and detectives, however, rarely calculate the number of markers associated with crimes that could potentially match a suspect to the crime or, as good practice requires, lay down crime scene grids and videotape the contents of each small cell. Nor do they use interview checklists to mine witnesses for all categories of identifying information they might have. Even markers that police reports and photos capture are rarely counted, and non-matching traits are rarely treated as Brady material and turned over to the defense so that it can count the matches and non-matches. The courts are partly to blame for these gaps because they routinely bar efforts to quantify the effect of matches or non-matches in court. Nor are courts likely to admit information about police agencies' or detectives' comparative match and non-match rates to help jurors assess an investigation's credibility or the reasonableness of doubts it leaves. Courts thus forgo powerful incentives for police to document all trace evidence and consider the aggregate weight of the matches and non-matches in each case and the pattern of matches and non-matches in all cases.

D. (DE)VALUED LEARNING FROM CAREFULLY OBSERVED COLLECTIVE EXPERIENCE

Three curiosities cannot capture the dispositions of all criminal courts, but they suggest a disturbing resistance to using the rigorous examination of well-documented collective experience to reveal the truth and expose error. This resistance extends to the full range of techniques used effectively in other domains: basing decisions on how the system operates in all cases, not just the one at hand; aggregating instances; identifying patterns; and comparing results at each site to prior results there and contemporaneous results at other sites. In all three examples, moreover, courts resist learning from experience despite their own or the public's recognition of the value of quantifying and comparing past experience to

162. Id. at 623, 681.
163. Id. at 681.
164. Id. at 661, 663–64.
165. Id. at 671–72.
166. See id. at 667–70 (describing the practical and legal obstacles to finding and admitting evidence suggesting the guilt of third parties who were previously under investigation in connection with the crime at issue).
167. Police have been at the forefront of using Big Data in developing crime-fighting techniques not routinely regulated by courts. See supra notes 75, 148 and accompanying text.
168. See supra notes 62–76 and infra notes 174–75.
improve the system. Drawing on some fitful innovations by the Supreme Court itself, the next part proposes several strategies criminal courts could adopt to harness the corrective power of rigorously observed collective experience.

V. STRATEGIES FOR CAREFULLY ANALYZING COLLECTIVE EXPERIENCE TO IMPROVE CRIMINAL JUSTICE

A. LEARNING FROM PATTERNS OF BEHAVIOR

The simplest strategy for using collective experience to discipline criminal justice is to attend to patterns of behavior of which the action at issue in the case is an example. The indifference shown by the Supreme Court to racial death-sentencing patterns in Georgia and varying interpretations of a facially overbroad capital aggravating factor in Idaho and by criminal courts generally to patterns of error in capital cases and to counts of matches and non-matches in identity cases are not a given. In Furman and several contemporaneous cases, the Court used patterns of behavior across entire fields of action to inform itself about how individuals and institutions tend to behave, why they do so, and even what the law-in-action is as revealed by patterns of official behavior. These precedents reveal that courts can easily gather evidence of their own collective experience and that of repeat players in their courtrooms, create incentives for administrators and researchers to generate still more information and rely on the revealed patterns of behavior to guide

169. See supra notes 102–110 and accompanying text.
170. See supra notes 111–167 and accompanying text.
171. See, e.g., Ballew v. Georgia, 435 U.S. 223, 232–37 (1978) (invalidating criminal juries with fewer than six members based on studies indicating that juries with fewer than twelve members tend to deliberate less effectively and accurately than twelve-person juries); Castaneda v. Partida, 430 U.S. 482, 496–98 (1977) (shifting burden to state to disprove that jury-selection procedures are discriminatory once defendant establishes a substantial pattern of underrepresentation of Hispanics on a jurisdiction’s juries); Vill. of Arlington Heights v. Metro. Hous. Dev. Corp., 429 U.S. 252, 266–68 (1977) (ruling that discriminatory motivation may be inferred in part from patterns of behavior or departures from normal substantive or procedural practices that tend to disadvantage members of racial or ethnic minorities); Keyes v. Sch. Dist. No. 1, Denver, Colo., 413 U.S. 189, 208 (1973) (holding that a finding of intentional school segregation within a “meaningful portion” of a school district creates a rebuttable presumption that patterns of segregation elsewhere are also intentional); Griggs v. Duke Power Co., 401 U.S. 424, 431 (1971) (prohibiting employment practices unrelated to job performance that have a disproportionate negative effect over time on minorities); Witherspoon v. Illinois, 391 U.S. 510, 517–18 (1968) (deferring judgment on whether juries made up entirely of individuals without conscientious scruples against the death penalty are more prone to convict pending further study of patterns of decisions by death-qualified and non-death-qualified juries); supra notes 91–94 and accompanying text (discussing Furman v. Georgia, 408 U.S. 238 (1972)).
172. See, e.g., Ballew, 435 U.S. at 232–37 (discussed supra note 171); Witherspoon, 391 U.S. at
decisions or at least trigger further analysis.\textsuperscript{173}

**B. PUBLISHING DATA ABOUT CRIMINAL JUSTICE OPERATIONS**

Publicizing information about error rates, near-misses, and costs, then comparing results across courts, agencies, subdivisions, and individuals, can incentivize and empower actors to shrink the number of potentially embarrassing outcomes and increase efficiency.\textsuperscript{174} Some commentators have recently proposed this strategy to deter excessive, coercive, retaliatory, or racially selective Terry stops, arrests, interrogations, charges, and plea bargains.\textsuperscript{175} Criminal courts could adapt this strategy to their needs by receiving evidence on the numbers and rates of matches and non-matches between suspects and crime-scene clues, eyewitness descriptions, and details of confessions and by calculating rates of judicial, law-enforcement, and defense error exposed on appellate review.\textsuperscript{176} Data on the relevant outcomes typically are available in the judicial system’s own records.\textsuperscript{177} Data on potentially causal factors can be generated by asking police, prosecutors, and defense counsel to key-in biographical, demographic, situational, and procedural information about defendants, victims, offenses, investigations, and procedures as they make stops or

\textsuperscript{517-18} (discussed supra note 171).

\textsuperscript{173.} See supra notes 91–94, 171 (describing decisions relying on statistics and behavioral patterns).

\textsuperscript{174.} See, e.g., REES, supra note 64, at 91–120 (discussing a debate over whether nuclear power plant rates of errors and near-misses developed through industry self-regulation should be disclosed to the public or only to the leadership of the regulated companies; concluding that the industry’s decision to adopt the latter course contributed to industry-wide improvements); Bradley C. Karkkainen, Information-Forcing Regulation and Environmental Governance, in LAW AND NEW GOVERNANCE IN THE EU AND THE US 293, 301–05 (Grainne de Burca & Joanne Scott eds., 2007) (explaining how the penalty default created by the all-or-nothing nature of the Endangered Species Act creates incentives for developers and environmentalists to cooperate and share information with each other to develop Habitat Conservation Plans that, when approved by the Department of the Interior, replace the act’s enforcement scheme); Erlend Nier & Ursel Baumann, Market Discipline, Disclosure and Moral Hazard in Banking, 15 J. FIN. INTERMEDIATION 332 (2006); Sabel, Fung & Karkkainen, supra note 66, at 116–18, 121–23 (discussing the ameliorative effect of public disclosure of the amount of emissions of toxic chemicals under the federal Toxics Release Inventory and the Massachusetts Toxics Use Reduction Act).

\textsuperscript{175.} See, e.g., STUNTZ, supra note 11, at 120, 212, 301 (proposing public disclosure to deter coercive and abusive interrogation, charging, and plea bargaining practices); Garrett, supra note 82, at 107–47 (proposing ways to use data on the number and outcomes of police stops under different circumstances as a tool for limiting abuses and maximizing law-enforcement effect); Johnson, supra note 68, at 407–16 (discussing congressional statute requiring juvenile justice authorities to record racial patterns of arrest and disposition and develop plans for addressing disparities).

\textsuperscript{176.} Liebman et al., supra note 129, at 685–87.

\textsuperscript{177.} Gelman et al., supra note 111, at 211, 214, 216–17, 235.
arrests, charge suspects, or enter appearances for defendants.178

C. USING DATA TO REVEAL THE DEMOCRATIC GOING RATE ON DIFFICULT INTERPRETIVE QUESTIONS

This and the next section discuss decisionmaking innovations by the Supreme Court itself that illustrate the ability of criminal courts to use data not only to reveal the effectiveness of their own actions and solutions but also to trigger innovations that broaden the solution set. The data-and-innovation-forcing mechanism in each case is a “penalty default”—a threat by the Court to mandate a single inflexible solution to the problem at hand unless the regulated entities craft their own locally tailored solution and measure its effectiveness.179

A regulatory difficulty criminal courts often face is deciphering the meaning of constraints the Bill of Rights places on the administration of criminal justice. Deciding whether the manner in which a state imposes or executes capital sentences is “cruel and unusual” is a good example.180 Furman and later decisions developed a partially aggregative solution to this problem. In Furman, the Court overturned every capital-sentencing statute in the nation by reading the Eighth Amendment to forbid wholly discretionary procedures, but it was initially unable to discern the procedures that provision allows. The Court addressed this problem by adopting the penalty default of abolition and challenging States that wanted the death penalty to formulate their own procedures that (1) were consistent with the “evolving standards of decency” on which the Court had previously relied in “cruel and unusual” punishment cases and (2) generated sentences proportional to the killings for which they were imposed.181

---

178. See, e.g., supra notes 96–101 and accompanying text (describing the statutorily mandated collection of information on the offenders, victims, and crimes involved in death penalty cases to facilitate judicial review of the comparative proportionality of death verdicts imposed).


180. U.S. CONST. amend. VIII. See also Liebman, supra note 85, at 113–14 (describing the difficulty of interpreting the Eighth Amendment).

181. See Furman v. Georgia, 408 U.S. 238, 242 (1972) (per curiam) (Douglas, J., concurring) (premising cruel and unusual punishment decision in part on whether the practice in question was consistent with the “evolving standards of decency that mark the progress of a maturing society”).
Having thus prompted states to expand the potential solution set, the Court applied the “evolving standards” test to each option by counting the number of states endorsing each one and in some cases counting the number of death sentences juries imposed under particular conditions. In essence, the Court used the preponderance of legislative and jury choices to reveal the nation’s current moral coordinates with respect to each option. Over a quarter century, the Court ruled the death penalty potentially constitutional for deliberate homicide, but unconstitutional for the following: if it was mandated for all individuals who committed a broadly or narrowly defined murder; imposed for rape of an adult or child; imposed for being an accessory to murder absent major participation in an accompanying felony or an aggravated mental state as to the killing; or carried out against someone who was “mentally retarded” at the time of the killing, a juvenile at the time of the killing, or insane at the time of the execution.182 By forcing states wishing to avoid abolition to generate new information about death-sentencing options and the degree of democratic consensus supporting each, the Court increased both the predictability and democratic legitimacy of its judgments.183

182. Head-counting decisions include Kennedy v. Louisiana, 554 U.S. 407, 426 (2008) (holding that state may not punish aggravated rape of a child with death: “it is of significance that, in 45 [state] jurisdictions, petitioner could not be executed for child rape of any kind”); Roper v. Simmons, 543 U.S. 551, 564 (2005) (“The evidence of national consensus against the death penalty for juveniles.... [includes that] 30 States prohibit the juvenile death penalty, comprising 12 that have rejected the death penalty altogether and 18 that maintain it but, by express provision or judicial interpretation, exclude juveniles from its reach.”); Atkins v. Virginia, 536 U.S. 304, 315-16 (2002) (“The large number of States prohibiting the execution of mentally retarded persons... provides powerful evidence that today our society views mentally retarded offenders as categorically less culpable than the average criminal.”); Ford v. Wainwright, 477 U.S. 399, 408, 410 (1986) (“[N]o State in the Union permits the execution of the insane.”); Enmund v. Florida, 458 U.S. 782, 792-96 (1982) (“weigh[ing] on the side of rejecting capital punishment for the crime at issue” in the case that (1) “only a small minority of jurisdictions—eight—allow the death penalty to be imposed solely because the defendant somehow participated in the robbery in the course of which a murder was committed,” and (2) the eight states that theoretically allow the death penalty under these circumstances rarely impose it in such cases); Coker v. Georgia, 433 U.S. 584, 596 (1977) (“The current judgment with respect to the death penalty for rape is not wholly unanimous among state legislatures, but it obviously weighs very heavily on the side of rejecting capital punishment as a suitable penalty for raping an adult woman.”); Gregg v. Georgia, 428 U.S. 153, 169, 179-82, 187 (1976) (plurality opinion) (noting, in the process of holding that the death penalty is not per se unconstitutional, that most state legislatures chose to reinstate capital punishment after the Court overturned preexisting capital statutes several years earlier). More recently, the Court has used the preponderance of state practices to regulate the punishment of life without parole for juveniles. See Miller v. Alabama, 132 S. Ct. 2455, 2470-74 (2012); Graham v. Florida, 130 S. Ct. 2011, 2023-26 (2010).

183. See Liebman, supra note 85, at 111-13 (concluding that the Court’s early decisions on comparative proportionality review in effect created a system in which the Court shared constitutional responsibility with state courts and legislatures for determining whether particular categories of uses of the death penalty constitute cruel and unusual punishment).
By encouraging states to adopt “comparative proportionality review,” Gregg v. Georgia initiated a similar innovation for answering the second constitutional question: whether sentences generated by a state’s capital statute were proportional to the severity of the particular killing. The difficulty here is to reconcile the twin aims of the Court’s death penalty jurisprudence and of centuries of criminal law: clear standards that generate consistent outcomes and the flexibility to dispense mercy based on the “diverse frailties of humankind.” By requiring state high courts to compare all sentences meted out in potential capital murder cases and to invalidate outlier death sentences imposed for insufficiently aggravated murders, the Court used state judges’ extraction of the aggregate moral judgment embedded in the run of hundreds of prosecutors’ charging and jurors’ sentencing decisions to expose each state’s capital “going rate.”

The responsibility then assumedly would fall to the U.S. Supreme Court to use the going rates of all states to extract a national going rate and overturn outlier—thus cruel and unusual—death sentences for particular crimes (as it occasionally did in the 1980s and 1990s), for entire categories of murder (as it also did), or for state capital-punishment schemes that generated patterns of outcomes outside the national mainstream.

As we note above, however, starting in 1984, the Court disavowed any requirement that state courts engage in comparative analysis of death-sentencing patterns in search of moral “going rates” and outliers and then refused to conduct such analysis itself. Still, the head-counting the Court continues to do and the comparative proportionality structures it

184. See supra text accompanying notes 95–101.
186. See Liebman, supra note 85, at 113–18 (arguing that “[t]he Court’s ingenious set of post-Furman proposals for sharing constitutional decisionmaking with local institutions was a striking example of” the use of aggregative and comparative analysis to improve judicial interpretation of the Constitution and achieve greater consistency in the application of the death penalty).
187. See, e.g., id. at 41–50 (discussing cases in which the Supreme Court appears to have conducted its own substantive review of capital outcomes).
188. See, e.g., Enmund v. Florida, 458 U.S. 782, 797 (1982) (holding the death penalty categorically disproportionate when imposed for the crime of being an accessory to a robbery in the course of which someone else killed the victim and absent proof that the accessory intended or attempted to kill the victim).
189. See Liebman, supra note 85, at 59–62, 73, 85, 113–18 (“[T]he cases close to the line, especially where the legislative category, sentencing pattern, or case outcome was different from those generated by most other local actors, the Court’s scheme contemplated that it would review the matter for itself.”).
190. See supra text accompanying notes 102–10.
temporarily envisioned reveal how the aggregate decisions of juries, prosecutors, courts, and legislators can guide and confer democratic legitimacy on the courts' search for legal meaning.

D. BENCHMARKING PROMISING PRACTICES

The Court also has recognized that systematically comparing the results of local experimentation in the shadow of penalty defaults can improve criminal justice. This approach works best when the difficulty is not defining legal standards but deciding how best to satisfy them, and when evaluation of local innovations is not by head-counting but by comparing how each performs. "Plan-" or "management-based" strategies of this sort are used in many domains to induce regulated entities to determine how the relevant problem manifests at each of their sites and to develop a site-specific plan for solving it.191 All plans then are evaluated centrally, including by benchmarking the quantitatively and qualitatively assessed results of each plan against those at similar sites and approving the better performers while disapproving the laggards.192 This technique allows more local flexibility than the bureaucratic specification of steps that every regulated entity must take or the performance-based specification of minimum or maximum outcomes that all regulated entities must or may reach.193

Furman's invalidation of discretionary death-sentencing and invitation to States to avoid the penalty default of abolition by developing their own alternatives again provides a crude example.194 States responded by adopting various versions of two approaches—mandatory death sentences for more or less narrowly defined categories of murder, and "guided discretion" mechanisms that use statutorily enumerated aggravating circumstances to narrow death-eligibility, then net out the aggravating and mitigating effect of available sentencing information. After comparing dozens of iterations of the two approaches, the Court eventually rejected the former approach in all its forms and insisted that states adopt a version of the latter approach that permits defendants to rely on any mitigating factor that at least one juror might accept as a basis for a sentence less than

191. See supra notes 64–76 and accompanying text.
192. For discussions of how various management-based strategies are implemented, see, for example, Coglianese & Lazer, supra note 47, at 717–18; Dorf & Sabel, supra note 179, at 433–34; Johnson, supra note 68, at 407–16; Olatunde C. A. Johnson, Beyond the Private Attorney General: Equality Directives in American Law, 87 N.Y.U. L. Rev. 1339, 1362–70, 1393–99 (2012); Liebman et al., supra note 129, at 679–81.
193. See supra notes 53–62 and accompanying text.
The Court's plan-based strategy remained crude, however, because beyond head-counting, it reached conclusions by speculating about the patterns of death sentences each option might generate, while eschewing data about how well each actually performed.196

The Court's mid-1960s decisions in *Miranda v. Arizona*197 and *United States v. Wade*198 provide additional examples, though, again, neither realized its potential. Both decisions reflect a conclusion that the Court's prior strategy of case-by-case evaluation of, respectively, confessions and eyewitness identifications under broad "involuntariness" and "reliability" standards was ineffective, along with a concern that any single procedure for obtaining confessions and identifications that the Court required would be unsuited to many situations. So, instead, the Court imposed default procedures—*Miranda* warnings and rules for conducting lineups, each enforced by exclusionary rules—and invited localities to avoid the default by adopting locally tailored alternatives that were "at least as effective."199

---


196. See, e.g., supra notes 182–83 and accompanying text.


199. Id. at 239 (inviting jurisdictions to forgo Court-imposed procedures if they adopted "[l]egislative or other regulations, such as those of local police departments, which eliminate the risks of abuse and unintentional suggestion at lineup proceedings"); *Miranda*, 384 U.S. at 467, 490 (encouraging law enforcement agencies "to continue their laudable search for increasingly effective ways of protecting the rights of the individual while promoting efficient enforcement of criminal laws" by allowing them to adopt substitutes for the Court's postarrest warnings and exclusionary-rule that "are fully as effective"). See also Dorf & Sabel, supra note 179, at 452–54 (discussing the Court's prophylactic holding in *Miranda*); Henry P. Monaghan, *The Supreme Court 1974 Term—Foreword: Constitutional Common Law*, 89 HARB. L. REV. 1, 4, 20–26 (1975) (arguing that prophylactic rules of the sort the Court created in *Miranda* to shore up Fifth Amendment protections can be justified as a form of constitutional common law that fosters beneficial self-regulation by law enforcement).
In deciding whether alternatives to *Miranda* were "as effective," for example, the Court assumedly planned to consider information about the numbers and proportion of arrests resulting in interrogations, interrogations leading to confessions, confessions leading to convictions, convictions challenged and reversed on appeal, and whether interrogations were recorded and demonstrated the voluntariness of statements made. We will never know, however, because the Court's default rules were so undemanding that police and prosecutors never developed alternatives.  

Wisconsin's recent treatment of eyewitness identifications demonstrates what an effective penalty default rule and benchmarking strategy might look like. In *State v. Dubose*, the Wisconsin Supreme Court interpreted the state constitution to exclude eyewitness identifications produced by "unnecessarily suggestive" procedures, a standard more likely to protect innocent defendants than the federal rule. The Wisconsin Legislature effectively turned *Dubose* into a penalty default by allowing local law enforcement agencies to avoid its effect by developing their own plans for "reduc[ing] the potential for erroneous identifications." It instructed localities to adopt procedures shown by research to improve "objectivity and reliability," to file their new policies with the Wisconsin Department of Justice, and to revise them every two years based on statewide experience. As an aid, the state Justice Department issues and periodically revises model procedures. Texas has adopted a similar regime.

By comparatively benchmarking the plans against each other based on factors such as how often each results in positive identifications, how often identifications are validated by conviction, and how often challenged convictions are upheld on appeal, the Justice Department can now help the

---

200. Liebman et al., *supra* note 129, at 682.
203. *See* Manson v. Brathwaite, 432 U.S. 98, 106, 114 (1977) (permitting the introduction of an eyewitness identification made as a result of suggestive procedures as long as the trial court determines that the identification is likely to be "reliable" as indicated by a set of forgiving criteria).
205. *Id.* § 175.50(5).
Wisconsin Supreme Court apply Dubose with the aid of comparative data identifying outlier counties and procedures. To increase the effectiveness of the system even further, Wisconsin might empower jurors, when evaluating the weight of identifications, to compare the procedures used in the county at hand to those used in other counties, along with the procedures’ error rates. Doing so would incentivize police to adopt procedures they are comfortable defending to well-informed jurors.  

Generalizing this approach, courts could encourage law enforcement to conduct more reliable investigations without imposing inflexible rules by allowing defense counsel to present evidence of investigative procedures used elsewhere in the state and the procedures’ comparative results. For example, Massachusetts allows trial judges to instruct jurors to consider whether shortcomings in the state’s investigation affected the “quality, reliability or credibility of the [state’s] evidence.” Allowing defense counsel to compare a locality’s practices to those used elsewhere would encourage localities to adopt the superior practice of other jurisdictions.

E. REGULATING MORE BY MANDATING LESS

By attaching consequences to comparative patterns of results, the strategies discussed here would supplant bureaucratic enforcement of inflexible rules via nuclear remedies with more effective, site-specific regulation driven by democratic and adversarial innovation. Courts could accomplish more while mandating less by (1) impelling judges and other criminal justice actors to measure their own results and error rates, (2) setting penalty defaults that apply only when jurisdictions fail to tackle the problems the data reveal, and (3) overturning outliers based on deviations from standards established by democratic institutions, not by off-the-cuff judgments of unelected judges.

---

208. Cf. James M. Lampinen et al., The Reactions of Mock Jurors to the Department of Justice Guidelines for the Collection and Preservation of Eyewitness Evidence, 27(2) BASIC & APP. SOC. PSYCH. 155, 155–62 (2005) (finding that mock jurors evaluating eyewitness identifications after being informed of U.S. Justice Department guidelines for conducting identifications rated questionable eyewitness-identification procedures and testimony less credible than did the study subjects who were not informed about the guidelines).


VI. (UNPERSUASIVE) REASONS FOR COURTS’ INDIFFERENCE TO COLLECTIVE EXPERIENCE

This part assesses three reasons why criminal courts may resist the corrective use of collective experience, despite those courts’ and other domains’ recognition of its value. We conclude that the first two reasons are outdated, and the third—an ideological preference—explains but does not justify the courts’ negative disposition toward collective experience.

A. TECHNICAL INCOMPETENCE

The Supreme Court sometimes dismisses the evaluation of aggregate criminal justice experience as “not readily susceptible to the kind of analysis the courts are competent to undertake.”211 This section discusses the courts’ competence to collect and analyze data. The next section discusses their competence to draw inferences from the results of such analyses about matters of policy that arguably are the domain of the political branches.

For reasons already mentioned, the difficulty of obtaining and analyzing data no longer justifies incompetence concerns. Nearly all the required data already exist, often in the records of criminal courts and criminal justice agencies themselves, or can easily be willed into existence through the penalty default and other “plan-based” regulatory mechanisms discussed above.212 Tools for analyzing the data also are available.213 And compared to other uses of data, the ones proposed here raise less serious privacy concerns, because they mainly rely on material already in public records and are entirely summative, not person-specific, which is consistent with uses permitted by the strictest forms of legislation protecting vulnerable populations from disclosure of government data.214

213. Compare, e.g., Laurence H. Tribe, Trial by Mathematics: Precision and Ritual in the Legal Process, 84 HARV. L. REV. 1329, 1335–37, 1360–63 (1971) (doubting that jurors could competently use Bayesian analysis to aggregate the probabilities of matches between the defendant and crime-scene markers and arguing that the effort was futile in any event given a lack of frequency data on relevant traits such as how many men in a particular neighborhood have long hair or wear a particular type of shoe), with Liebman et al., supra note 129, at 674–79 (describing recent studies establishing the accessibility of Bayesian analysis to lay decisionmakers when it is properly explained to them, and the broad availability today of frequency information from ubiquitous security cameras and the like).
Nor should courts and lawyers continue to be excused—or amused—by the old saw, “if I were good at numbers, I would not have gone to law school.” If a corporate or white-collar-criminal lawyer demurred on this basis when asked to analyze the financials or market analyses in a case, she would lose her job. It is time for judges and lawyers in blue-collar criminal cases to catch up to this reality. Even concerns about how jurors handle aggregative analysis are overblown. Although jurors do not understand Bayesian equations very well, recent studies show that they reach appropriate Bayesian judgments when instructed using demonstrative diagrams instead of equations. Once jurors get the basic concept, analogously to their understanding of long division, it is appropriate for them to use Bayesian calculators to process their analyses after using the usual inductive reasoning to fill in the variables.215

B. HUMILITY

Neither is humility a good reason for criminal courts to resist aggregation and comparison. True, worries about doing things others do better and about doing too much are peculiarly judicial concerns warranted by the separation of powers between courts and the political branches. As we already have noted, however, criminal courts’ core function includes the daily administration of criminal justice in close coordination with police officials, forensic scientists, prosecutors, and public defenders.216 For courts to disclaim the responsibilities of administrators, therefore, is to place the system’s accuracy, as well as its legality, at risk.

Still, systemic analysis may pose particular threats to the courts’ legitimacy if it triggers expansive remedies that are especially the province of democratically disciplined officials. After attempting large scale social engineering through its criminal justice rulings of the 1960s and 1970s, the Supreme Court has little stomach left for broad procedural edicts.217 Recently, for example, it refused to give prisoners a right to postconviction testing of DNA evidence possessed by the state in order to avoid “taking the development of rules and procedures in this area out of the hands of legislatures and state courts shaping policy in a focused manner and turning it over to federal courts applying the broad parameters of the Due Process Clause.”218

215. Liebman et al., supra note 129, at 676–79.
216. See supra notes 13–38, 82–90 and accompanying text.
217. See supra notes 44–46 and accompanying text.
Again, however, our earlier discussion explains why humility is a bad reason to forbear aggregative analysis, given the courts’ responsibility to regulate their own procedures, the ill effects of judicial inaction on legislative and administrative innovation, and the ability of courts—in lieu of policy-monopolizing undemocratic mandates—to use penalty defaults, comparative benchmarking, adversarial testing, and other mechanisms to base decisions on the innovations and moral judgments of democratically constrained officials and juries.

Indeed, by allowing sophisticated comparison, gradation, and exposure of outliers, aggregative analysis allows courts to employ even humbler remedies than were previously available. Statistical aggregation and comparison have many of the advantages—and their greater precision avoids many of the disadvantages—of the most modest forms of judicial decisionmaking, such as case-level balancing and totality-of-circumstances analysis. An example is provided by the Court’s dubious *McCleskey* decision. Faced with evidence that Georgia death-sentencing patterns were explainable on no basis other than race, Justice Powell’s majority opinion apocalyptically predicted that granting relief would effectively bar jurisdictions from using not only the death penalty but also all forms of contextualizing and humanizing discretion in resolving criminal cases. As Justice Stevens noted in dissent, however, the data themselves pointed the way to far more modest relief, given that racial disparities did not appear in highly mitigated and highly aggravated cases. Relief thus could be limited to a targeted subset of midrange cases. Or, the Court could require Georgia to choose among procedures that other states use to limit death sentences to superaggravated cases, to use comparative proportionality review to screen-out low-aggravation outliers, or to

---


221. *See supra* Part V.


224. *Id.* at 367 (Stevens, J., dissenting).

225. *Id.*

226. *See Liebman & Marshall, supra* note 93, at 1665 (cataloguing steps jurisdictions can take to narrow the death penalty).

overturn low-aggravation cases on a totality-of-the-circumstances basis.\footnote{See supra note 187 and accompanying text.}

As these examples indicate, resort to deductive data analysis will not absolve courts of inductive judgment and line-drawing. But induction disciplined by data and structured analysis allows an intermediate kind of judgment that the Pragmatists call “abduction.” As they have shown, such analysis is far closer to the gold standard of deductive reasoning than is unstructured intuition.\footnote{See ANSELL, supra note 47, at 202–03 n.12 (explaining that “abduction is more deliberate than intuition”); CHARLES SANDERS PEIRCE, PRAGMATISM AS A PRINCIPLE AND METHOD OF RIGHT THINKING: THE 1903 HARVARD LECTURES ON PRAGMATISM 217–18 (Patricia Ann Turrisi ed., 1997) (defining abduction as modes of reasoning that study facts, devise a hypothesis to explain them, and test the hypothesis based on the observation of common experience).}

C. PROTECTION OF LIBERTY

There is a more visceral explanation for the courts’ resistance to rigorous evidence of collective experience: the libertarian concern that aggregative analysis is a form of inquisitorial decisionmaking that jeopardizes sacred freedoms. Most at risk, it is said, are adversarial protections and the assurance of individualized decisionmaking that animates the ban on character evidence, preferences for “direct” over “circumstantial” evidence and for individualized over statistical proof of discrimination, and restrictions on defensive evidence implicating third-party suspects. A classic of this genre is Laurence Tribe’s 1971 article Trial by Mathematics, which contends that efforts to quantify the probability that a defendant is guilty by aggregating probabilities associated with individual matches between crime-scene evidence and the defendant would violate our commitments to the presumption of innocence, trial by jury, proof beyond a reasonable doubt, and punishing people for their actions and not for being members of a certain class.\footnote{Tribe, supra note 213, at 1355–78. See also Liebman et al., supra note 129, at 600 (discussing Tribe’s critique of aggregative evidence).}

Tribe’s concerns notwithstanding, there is every reason to believe that banning aggregative analysis is more harmful than helpful to the rights of criminal defendants who are innocent or undeserving of particular penalties. This was true of Carlos DeLuna, who could not use Bayesian analysis to show that, collectively, the many individually trivial non-matches between himself and the crime scene created more than a reasonable doubt about his guilt.\footnote{See Liebman et al., supra note 129, at 585, 653–54 (listing non-matches that could have been aggregated to show that DeLuna was innocent or at least that there was a reasonable doubt as to his guilt).} It was also true of many dozens of
innocent prisoners whose exonations reveal that they were harmed by our system’s overweighting of confessions, eyewitness identifications, fingerprints, and other seemingly “individualizing” proof, in lieu of rigorous efforts to aggregate the effect of all kinds of match and non-match evidence. And it was true of Warren McCleskey, who could not rely on systemic analysis to show the influence of race on his death sentence.

The crucial role of the adversarial system in domesticating DNA analysis and current efforts to enhance the adversarial testing of confession, eyewitness, and fingerprint evidence similarly reveal that aggregative analysis and adversarial process are not at odds. Likewise, Bayesian analysis can be shown to justify both the common-law character rule, which deters heuristic departures from statistical truths that result from inattention to base rates, and also that rule’s “modus operandi,” “common scheme,” “habit,” and “business routine” exceptions, all of which use a form of aggregative analysis to aid the search for truth. Again, there is no inherent conflict between aggregative analysis and liberty-protecting procedures.

Sometimes, however, the criminal law’s individualizing and anti-aggregative dispositions are justifiable only as a libertarian, antigovernment prejudice and not because they protect the liberty of innocent, overcharged, or overpunished defendants. This is true, for example, of (1) the preference for direct over circumstantial evidence, which thankfully the courts are gradually abandoning; (2) the still-strong disposition of law-enforcement, jurors, and judges to overvalue evidence they falsely consider to be “unique” and “individual” such as eyewitness, confession, and fingerprint evidence; (3) the failure to adopt Bayesian analysis of match evidence; supra text accompanying notes 133–34.

232. See Liebman et al., supra note 129, at 584–87, 625–28 (demonstrating the use of Bayesian analysis to show that multiple individually inconsequential non-matches can aggregate to a reasonable doubt of guilt).

233. See supra notes 103–05 and accompanying text.

234. Liebman et al., supra note 129, at 610–18. See also supra notes 155–60 and accompanying text.


236. See Liebman et al., supra note 129, at 657–58 (discussing modern trend toward weighing direct and circumstantial evidence equally).

237. See id. at 627–28 (discussing social psychological evidence that humans instinctively rely more heavily on individualized evidence and give less weight to circumstantial evidence than both types of evidence deserve).
and non-match evidence;\textsuperscript{238} (4) the courts' stubborn resistance to evidence of a "third-party's" guilt and demand that criminal trials be an individualized examination of the guilt of the defendant alone, even when a comparative analysis of the probability of the defendant's guilt and that of other suspects to whom solid evidence points would lower the risk of false conviction;\textsuperscript{239} and (5) the rejection of pattern evidence to establish selective prosecution and sentencing.\textsuperscript{240}

VII. CONCLUSION

Our analysis reveals the tendency of libertarian and individualizing dispositions to accomplish in the realm of criminal law and procedure what they are better known for accomplishing in the realms of economic and social policy: cementing longstanding patterns of advantage and disadvantage that systematically work to the detriment of poor and minority populations.\textsuperscript{241} These dispositions are the real reason for the criminal justice system's benighted resistance to aggregative and comparative analysis of conditions crucial to the accuracy and efficiency of criminal justice. In turn, our analysis reveals the danger of nostalgia for the fleeting intersection in the Warren Court's Criminal Procedure Revolution of libertarian dispositions and the promotion of the interests of disadvantaged populations. For too long, observers enamored of the Criminal Procedure Revolution have been entranced by its momentary, no longer effective, co-optation of that libertarian streak in service of the protection of innocent criminal defendants and the cause of truth. Romanticizing aside, however, it is now time to get in touch with our technocratic selves. Without jettisoning juries, adversarial proceedings, or the character rule, we must put the tools of rigorously analyzed collective experience at the service of the criminal justice system's self-improvement and search for the truth.

\textsuperscript{238} See id. at 667–73 (discussing current legal doctrine's antipathy to aggregative analysis of the weight of evidence of matches and non-matches in identity cases); supra text accompanying notes 146–67.

\textsuperscript{239} See Liebman et al., supra note 129, at 667–70 ("[T]he realities of our criminal justice system and the rules limiting evidence of a third party's guilt make it difficult to mount an effective 'I didn't do it, but I'll tell you who did' defense.").

\textsuperscript{240} See STUNTZ, supra note 11, at 119–22 (criticizing three Supreme Court decisions that have set the legal foundation for the impermissibility of using pattern evidence to show racial motivation by prosecutors and other criminal justice decisionmakers).

\textsuperscript{241} See supra notes 2–8 and accompanying text.