2016

The Organization of Prosecutorial Discretion

William H. Simon
Columbia Law School, wsimon@law.columbia.edu

From Prosecutors and Democracy: A Cross-National Study, edited by Maximo Langer and David Sklansky (forthcoming 2017)

THE ORGANIZATION OF PROSECUTORIAL DISCRETION

William H. Simon

Discussion of prosecution reform is haunted by anachronistic conceptions of judgment and organization. These conceptions see professional judgment as inherently individual and ineffable and professional organization as inherently informal and opaque. The appeal of these conceptions is due in part to the assumption that the only alternative to the judgment and organization they prescribe is bureaucracy. In fact, post-bureaucratic forms of organization have become dominant in recent decades in several professions. They key elements of post-bureaucratic organization are presumptive rules, root cause analysis, peer review, and performance measurement. Each of these elements can be found in recent reforms in prosecution, but the field, like the legal profession generally, lags other occupations. Although post-bureaucratic reforms are sometimes resisted as inconsistent with democracy, they are better understood as democracy-reinforcing.

Contemporary understanding of prosecutorial discretion is influenced by anachronistic conceptions of judgment and organization. These conceptions have lost ground dramatically in professions like medicine, teaching, and social work. Yet, they remain prominent to a unique degree in law. They are embedded both in the general professional culture and in legal doctrine. Innovative prosecutorial practices have emerged in recent decades, but their progress has been inhibited by attachment to these older conceptions.

The older conceptions understand professional judgment as substantially tacit and ineffable decision by a single professional grounded in a relatively static and comprehensive discipline. The associated model of organization emphasizes decentralization, pre-entry training and certification, and a reactive, complaint-driven approach to error detection. This view contrasts professionalism to bureaucracy – decision driven by stable, rigid, and hierarchically-promulgated rules. The professions operate in realms where bureaucracy is often ineffective, and the case for professional judgment, traditionally understood, rests in part on the assumption that it is the only alternative to bureaucracy.

Yet, models of judgment and organization that are neither bureaucratic nor traditionally professional have established themselves in many sectors of both the private and public realms. These models, which might be called post-bureaucratic – or in one important variation, experimentalist -- see decision as governed by explicit but provisional norms and arising from multidisciplinary group deliberation. They imply forms of organization that combine local autonomy with centralized monitoring, foster continuous learning and revision, and take proactive approaches to error detection and correction.

I appeal in this paper to models of post-bureaucratic or experimentalist organization both to emphasize the extent to which prosecution has lagged other sectors in its understanding of judgment and organization and to connect the important innovations that have occurred in prosecution to developments in other fields.

The analysis of competing conceptions of organization has implications for the relation of prosecutorial discretion and democracy. Post-bureaucratic organization has two features that promise to enhance democratic accountability – greater transparency and greater potential for stakeholder participation.

I. Traditional Premises

The *discretion* part of "prosecutorial discretion" connotes a combination of flexibility and discipline that elides arbitrariness on the one hand and regimentation on the other. Our key paradigm for such activity is the traditional idea of professional judgment.[1]

In the paradigm, judgment is a decision by an individual applying a discrete body of university-based knowledge to a particular situation. The decision is presumptively all-things-considered, taking account both the full range of knowledge within the professional field (but not beyond the field's boundaries) and of all relevant aspects of the particular situation. It is substantially tacit and ineffable; it cannot be explained fully to lay people and its correctness cannot be determined confidently even by peers in a large fraction of instances. And the decision is difficult to observe, in part, because it is so sensitive to myriad particular facts and in part because many of these facts are confidential. The disciplines such judgments implement are understood as stable, and their general effectiveness can only be assessed in informal ways.

[1] See Burton Bledstein, The Culture of Professionalism: The Middle Class and the Emergence of Higher Education in America (1976); Talcott Parsons, "A Sociologist Looks at the Legal Profession," in Essays in Sociological Theory (Rev. ed 1954).

This type of judgment implies a distinctive form of organization. Work units tend to be organized by discipline, with workers supervised by members of the same profession and physically separated from people in other fields. Offices tend to be relatively decentralized. Workers are only loosely supervised. Instead, responsibility is assured in substantial part by licensing controls that certify the adequacy of training and ethical disposition on entry. Learning on the job occurs most characteristically through informal association with supervisors and mentors. These mechanisms are supplemented by processes of error-detection and correction that are initiated by complaints. Although the complaint processes are initiated by clients, the key judgments are made by, or strongly influenced by, professional peers. Errors are understood as idiosyncratic and are adjudicated and remedied one-by-one.

This vision of professional judgment has been nowhere more entrenched than in law, and in particular in the discussion of prosecutorial discretion. Consider three recent examples:

An article by Zachary Price on the political and constitutional dimensions of enforcement discretion has received a lot of attention, in part because of its pertinence to various controversial initiatives of the Obama administration, including guidelines for enforcement of immigration, controlled substance, and health care laws. In general, Price views as undesirable, constitutionally suspect, or worse most efforts by prosecutors to discipline or make transparent their enforcement decisions through explicit rules, guidelines, or general norms. His most encompassing objection derives from a conception of law and the separation of powers. It rests on a distinction between "categorical" or "across the board" norms and "individualized" or "case specific" judgments. "Executive nonenforcment discretion extends only to case specific considerations," he insists.[2]

This is because wholesale non-enforcement amounts to "making" or "re-making" law, which is a legislative power, while only retail nonenforcement is consistent with the executive function of "applying" the law. If, for example, it is impossible or undesirable to enforce the immigration laws fully against undocumented residents, the executive branch should not specify the criteria it will use to select residents for deportation but should instead permit such decisions to be made by frontline agents, asylum officers, and administrative law judges with minimal guidance other than the statutes and an informal sense of equity.

[2] Zachary Price, *Enforcement Discretion and Executive Duty*, 67 Vanderbilt L. Rev. 671, 705 (2014).

The second example is *Connick v. Thompson*, in which the Supreme Court considered a claim that due process required the New Orleans district attorney to train his subordinates about the constitutional duty to turn over exculpatory evidence to the defense. At least one, and perhaps several, prosecutors in the office had violated this duty in connection with a trial of the plaintiff years earlier. At least four other violations by lawyers in the office had been condemned by the courts in the prior ten years. As far as the record showed, the agency did no relevant training. The plaintiffs invoked earlier cases holding that the failure of a police department to provide training in the use of deadly force could violate the Constitution.

Justice Thomas, writing for the Court, rejected the idea that the police cases were relevant to prosecutors. He emphasized that lawyers, to a far greater extent than police officers, must undergo lengthy education and then demonstrate their general knowledge on a demanding examination prior to entering the occupation. "These threshold requirements are designed to ensure that all new attorneys have learned how to find, understand, and apply legal rules." In addition, lawyers are screened at entry for "character and fitness" and subjected to a regime of peer discipline throughout their careers. The opinion concludes that the senior officials "were entitled to rely on the prosecutors' professional training and ethical obligations in the absence of a specific reason" to believe they were not qualified.[3]

Finally, Rachel Barkow has advanced a proposal for re-organizing prosecutorial activity that focuses on the problem of bias. Barkow is worried about the kind of bias that arises from the design of professional roles. US prosecutors are normally responsible both for investigating and referring for prosecution on the one hand and for charging, determining what punishment to seek, and negotiating with the defendant on the other. Bias arises from the tendency of prosecutor to identify cognitively and emotionally with the understanding of the case that emerges in the investigation stage. This makes her resistant to revising this interpretation as new information emerges later. Barkow's solution is to sub-divide functions, assigning separate lawyers to the tasks of investigation and "adjudication" (i.e., charging and plea bargaining).[4] The proposal departs from the traditional professional view in dividing the professional decision in two and bringing in a second decision-maker. But each of the now separate decisions is made in the traditional manner – by independent individuals under unspecified criteria. Moreover,

---

[3] *Connick v. Thompson*, 131 S.Ct. 1350, 1361, 1363 (2010).

[4] Rachel Barkow, *Institutional Design and the Policing of Prosecutors*, 61 Stan. L. Rev. 869 (2009).

responding to bias in this manner carries a serious cost: the second decision-maker, by virtue of his separation from the investigation, may lack information that should be considered in the "adjudication" decisions.

II. Recent Trends

This traditional view of decision-making is in strong tension with recent thinking in many fields about the nature of decision-making and its implications for institutional design. In field after field, practices have been re-designed on the basis of an opposed understanding. [5]

In the first place, this opposed understanding rejects any strong distinction between categorical and individualized decision-making. Psychologists demonstrate that thinking is always categorical.[6] People process decisions through implicit criteria derived both from idiosyncratic social experience and the surrounding culture. Social scientists observing individual decisions over many cases can infer the implicit criteria even though the subjects may be unaware of them. A mandate like Price's for individualized decisions does not result in unmediated contextuality, but rather decisions governed by tacit and perhaps unconscious criteria over more explicit and reflective ones.

Such a mandate has serious costs. To some extent the implicit criteria that generate ostensibly individual decisions will vary across decision-makers, thus violating the value of horizontal equity. Such inconsistency is often invisible, but immigration asylum decisions provide an unusually salient and troubling example of it. The rates of asylum decisions in favor of applicants vary enormously and persistently among adjudicators.[7] Since cases are randomly assigned and each adjudicator decides many cases, it is hard to account for these variations other than as manifestations of idiosyncratic adjudicator views. Moreover, even where the tacit criteria influencing decisions reflect widely shared social dispositions, they may be illegitimate. For example, the pervasive unconscious influence of racial bias has been elaborately demonstrated in many other areas.[8]

---

5 See generally, Charles F. Sabel, *A Real Time Revolution in Routines*, in The Corporation as a Collaborative Community (Paul Adler and Charles Hecksher, ed.s 2006); Charles F. Sabel and William H. Simon, *Minimalism and Experimentalism in the Administrative State*, 100 Georgetown L. J. 58 (2011).

6 See, e.g., Steven Winter, A Clearing in the Forest: Law, Life, and Mind (2003).

7 Jaya Ramji-Nogales, Andrew I. Schoenholtz, and Philip G. Schrag, *Refugee Roulette: Disparities in Asylum Adjudication*, 60 Stan. L. Rev. 295 (2007).

8 See *Glossip v. Gross*, 576 U.S. (Breyer, J., dissenting) (2015), Slip Op. at 10-17 (discussing numerous statistical studies of the death penalty that conclude that its application does not correlate with plausible criteria of relative egregiousness and/or that it does correlate with factors that should not be considered, notably race). Note

In modern industrial organization, designers reject the tacit particularistic decision-making associated with traditional "craft"-style production (an industrial analogue to professionalism). They insist that tacit norms be made explicit. The craftsmen will rely on a learned, inarticulate sense of appositeness in deciding, say, how to apply stain to a table and what level of finish should be deemed adequate. Modern production insists these norms be made explicit and precise.[9] There are three reasons for this insistence. The process by which norms are articulated requires reflection that improves the quality of decisions. Explicit norms can be taught more quickly to newcomers. The learning model common to professionals and craft workers in which young workers learn from their seniors through a kind of informal osmosis has been discarded as inefficient. And most importantly, decisions under explicit norms are more transparent to observers; so they are more easily assessed and changed. The traditional model assumes a relatively stable body of specialized knowledge. But many fields face intensified pressure to adapt to changing circumstances.

Moreover, in the opposed understanding of judgment, the paradigmatic decision-maker is no longer an individual but a group that draws, not on a single discipline, but on several. Group decisions tend to be more consistent than individual ones, and they can synthesize a broader range of knowledge. Moreover, individual participants in groups feel pressure to consciously consider and articulate matters they would take for granted in solitary or more homogeneous settings.

Group decision-making is in part a response to the problem of professional bias that Barkow addresses. Bias is addressed by forcing individuals to articulate their premises and lay them open to challenge. A diverse group will likely contain people who do not suffer from any particular bias (or who may have offsetting ones). This approach avoids the disadvantage of Barkow's suggested remedy of sub-dividing the decision among different individuals. The second decision-maker avoids the bias of the first only at the cost of less information about the case.

At the same time, decisions tend to be multidisciplinary. This tendency responds to two developments. One is the evolution of perceptions of social problems. Some pressing problems

---

also Justice Thomas's reply in his concurrence that statistical results are unpersuasive in part because the analysts must abstract from the rich particularity observed by judges and juries at the trials. Slip Op. at 4-6. The reply misses the point of the analyses, which purport to show that the decisions are being driven by tacit criteria rather than ineffable particularity.

[9] *E.g.*, Productivity Press Development Team, Standard Work for the Shop Floor (2002).

that were not salient when the modern professional disciplines were established implicate multiple disciplines. Mental health and substance abuse, for example, are viewed as simultaneously public health and law enforcement problems. At the same time, evolving understandings of organization suggest greater capacity to coordinate interventions across disciplines and across institutional separations. When complex judgment at the organizational frontline was the exclusive province of the individual professional, and the individual professional was nested in a predominantly bureaucratic organization, coordination across organizational boundaries was difficult. More flexible contemporary organizational forms open up greater possibilities.

Professionals often resist the move away from the traditional understanding of judgment because they assume such a move would entail bureaucratic organization. They resist bureaucracy because it threatens individual fairness by regimenting judgment. In addition, while bureaucracy is superficially more compatible with public accountability than professionalism, in practice it can be equally opaque. Modern organizations that purport to operate in hierarchical, rule-governed manner described by Max Weber and Frederick Taylor in fact make room for a lot of frontline discretion. This discretion tends to be exercised informally on the basis of tacit peer cultures, and it tends to be substantially unobservable by supervisors and the public. Frontline agents ("street-level bureaucrats") can depart from the rules both for benign reasons (when the rules dictate patently unjust or inefficient decisions) and malign ones. Supervisors tolerate low-visibility rule departures either because the limits of their capacity to monitor leave them no choice or because they favor the benign departures. But benign or malign, low-visibility discretion is unaccountable except perhaps through the kind of socialization and recruitment controls that the critics assert make professionalism an inadequate mode of organization for the tasks in question.[10]

In fact, bureaucracy is not the only alternative to loose, informal decentralization favored by traditional professionalism. Major trends in important sectors of both private and public organization have produced a post-bureaucratic model of organization. Post-bureaucratic organization responds to the demands for adaptive and individualizing capacities in a world where uniform answers are undesirable and tacit cultural understanding is indeterminate. It

[10] Alvin Gouldner, Patterns of Industrial Bureaucracy (1954); Michael Lipsky, Street-Level Bureaucracy (1980).

repudiates both inflexible rules and low-visibility discretion. Decision-making in these regimes tends to be group and multidisciplinary. Accountability does not depend on either on monitoring compliance with fixed rules or socialization into an ineffable culture. The most distinctive mechanisms are (1) presumptive rules; (2) root cause analysis of unexpected events; (3) peer review; and (4) performance measurement.[11]

All these features can be observed in current prosecution practice. However, they seem less widespread and deep-rooted here than in other fields, and as Price, Thomas, and Barkow illustrate, they are often ignored or misunderstood. Indeed where we find prosecutors involved in sophisticated post-bureaucratic regimes, they often seem to have been pulled in by leaders in fields other than law. Problem-Oriented Policing, which has reconceived crime control strategies, and the Juvenile Detention Alternatives Initiative, which has transformed pretrial detention of juveniles, are examples.[12] Prosecutors play important roles in both, but most of the pioneering work has been done, in the first, by police officers and criminal justice academics and, in the second, by probation officers and sociologists. The greater prestige and longer history of law as a professional discipline relative to these other fields may have been liabilities that have inhibited reconception of practice.

The most general contours of the move toward post-bureaucratic organization figure in what Catherine Coles describes as a trend away from "the felony case processing model" toward "the community prosecution model" The first model defines its goal as the maximization of convictions, weighted by seriousness of the crimes. Convictions are not ends in themselves, but the model assumes that they are the only relevant means of attaining the ultimate goals, so that there is no need for practitioners to refer directly to these goals in their decision-making. Decisions in this model are made by lawyers, often with frontline actors exercising substantial autonomy and "operat[ing]in relative isolation from other agencies" and stakeholders.[13]

---

11 For example, see Charles Kenney, The Best Practice: How the New Quality Movement is Transforming Medicine (2010); Anthony Bryk et al., Learning to Improve: How America's Schools Can Get Better at Getting Better (2015). On the lagging position of the legal profession, see William H. Simon, *Where Is the Quality Movement in Law Practice?*, 2012 Wis. L. Rev. 387.

12 See Herman Goldstein, Problem-Oriented Policing (1990); Juvenile Detention Alternatives Initiative, Juvenile Detention Risk Assessment: A Practice Guide to Juvenile Detention Reform (2006); Juvenile Detention Alternatives Initiative, Two Decades of JDAI (2009).

13 Catherine Coles, *Evolving Strategies in 20th Century American Prosecution* in The Changing role of the American Prosecutor 177-209 (John L. Worrall and M. Elaine Nugent-Borakove ed.s 2008).

In the Community Prosecution model, decision-makers are guided directly by the ultimate goals of public safety and quality of life. The model assumes that felony prosecution is not a uniformly effective intervention and that, even when it is effective, it is best combined with other strategies. The goal is to craft solutions tailored to specific problems. Lawyers work in offices with non-lawyer specialists and engage continuously with other agencies and stakeholders.

The term "community prosecution" connotes local initiatives, but the post-bureaucratic architecture Coles describes can be applied to initiatives on any scale. Ideally, local efforts are linked through central institutions that measure effectiveness and pool information on the relative success of different strategies. At the same time, national and international interventions can devolve operating initiatives to frontline actors while monitoring and analyzing their efficacy.

III. Elements of Post-Bureaucratic Organization

At any scale, the key features of post-bureaucratic organization are the presumptive rule, root cause analysis, peer review, and performance measurement. Note that each challenges the traditional dichotomy between bureaucratic and professional organization, and the associated premises that we must choose between rule-based and standards-based judgment and between centralized and decentralized organization.

A. Presumptive Rule

A presumptive rule is neither a rule (a norm that dictates decision on the basis of a limited number of specified factors) nor a standard (a general value to be furthered by an all-things-considered judgment). A presumptive rule is more specific than a standard, but unlike a bureaucratic rule, those to whom it is addressed are expected to depart from it in circumstances where it would be counter-productive to follow it. The departure, however, must be signaled, and it triggers an immediate review of the departure. When the departure is sustained, the rule gets re-written to reflect the new understanding achieved through review.

Practice under a regime of presumptive rules is more transparent because it conforms more tightly to the rules than in a conventional bureaucracy. Practice is also more self-conscious since actors must justify decisions that would be taken for granted in a rule-governed regime. In a bureaucracy, following the rule is always an acceptable explanation, and rule departures are generally unobserved or ignored. But in a post-bureaucratic regime, following the rule is not appropriate where doing so would be counter-productive, and departures must be transparent. A

key goal is to induce and facilitate learning. This occurs in two ways. The duty to depart when the rules are ineffective and to signal departure feeds back information from the frontline that facilitates revision. Second, as I will shortly emphasize, experimentalist regimes subject practices to testing, and only explicit practices can be tested with any rigor.

Constitutional doctrine on prosecution has shown little concern with internal administration. The courts insist that administrators respond to indications of frontline violations of rules they themselves have promulgated, but where those rules (and relevant statutes) leave prosecutors discretion, they seem indifferent to whether offices take initiative to structure that discretion.[14]

Practice is thus free to vary, and it does widely. In some quarters, judgments are left to informal processes and minimally supervised individual decision-making. But we also find sophisticated policy manuals that make use of the presumptive rule. The Department of Justice U.S. Attorneys Manual is a notable example. It sets out some policies in detail and then says that local offices may depart from them "[i]n the interests of fair and effective law enforcement" but only with the approval of the appropriate Assistant Attorney General and the Deputy Attorney General.[15]

The Attorneys Manual deals mostly with trans-substantive rules that apply to practices that recur across various initiatives. They do not cover some important practices, and they do not deal in detail with decisions about the allocation of resources across initiatives or with the strategic configuration of particular initiatives. In a fully articulated presumptive rule regime, the rules form a plan that reflects a coherent but provisional understanding of the relevant mission. The plan is revised periodically both in both piecemeal and overall re-assessments. A comprehensive plan embraces sets of more specific plans.

Plans of this kind are most readily found in some self-consciously reformist initiatives, such as drug courts and problem-oriented policing. Problem-oriented policing was developed

---

[14] See Charles F. Sabel and William H. Simon, *The Duty of Responsible Administration and the Problem of Police Accountability*, 34 Yale J. on Reg. 165 (2016). For example, *Connick v. Thompson* notes that some of the prosecutors in the case were uncertain whether they had a *Brady* duty to turn over or test blood evidence that *might* turn out to be exculpatory if tested but had not been tested. The opinion assumes that, if *Brady* did not apply, there was no constitutional problem. No one suggested that the prosecutors had a duty to clarify this issue internally with their own rule, though that is what basic norms of good management required once the issue was identified.

[15] US Attorneys Manual 9-27.140. In a fully developed post-bureaucratic regime, the rules would be periodically reconsidered and re-written in the light of approved departures. There is no indication that this happens systematically in the Department of Justice.

mostly in the policing field, but it necessarily involves prosecutors. At the frontline, it involves local plans focused on geographical sites associated with recurring criminal activity or individuals or groups engaged persistently in criminal activity. Multidisciplinary teams engage with stakeholders to craft interventions and then periodically re-assess their efficacy. The initial intervention is codified in an explicit plan that gets reconsidered in the light of experience. The plan is thus a set of presumptive rules.

The shift from informal standards associated with professionalism to the presumptive rule is salient in the Juvenile Detention Alternatives Initiative. The initiative is a network of local criminal justice agencies supported by a foundation and responsive to a federal statutory mandate that the agencies produce and implement plans to reduce disparate racial impacts from their activities. A central reform that has emerged is the development of numerical Risk Assessment Instruments to govern pretrial detention decisions.

The story parallels the one told about baseball recruiting in Michael Lewis's *Moneyball*. Traditionally, probation officers made detention issues through minimally supervised individual all-things-considered judgments. Reformers believed that these judgments tended to be inconsistent, but there was no way to tell for sure because of limited review and the absence of articulated norms. Gradually this process has been replaced by one in which a scorecard dictates decision on the basis of numerically scored indicators, such as prior offenses, school attendance, or substance abuse. The scorecard is a presumptive rule. The decisions it dictates can be overridden, but only with the approval of a supervisor. When the scoring norms are periodically reviewed, reviewers look at overrides to see if they suggest inadequacies in the rules. The reforms have led to more consistent judgments and have made it possible to investigate the predictive power of the indicators. They seem to have contributed to declines in detention in most jurisdictions that have adopted them and to have reduced racial disparities in some.

B. Root Cause Analysis of Significant Operating Events

A significant operating event is an occurrence involving actual or potential harm that is unexpected or cannot be immediately explained. Examples include abnormal adverse health events in hospitals or "near misses" in aviation. Bureaucracy tends to treat such events as idiosyncratic. It tends to ignore the ones that do not involve tangible harm. It tends to respond to harm by sanctioning those responsible and/or compensating those who suffer the harm.

Post-bureaucratic organization requires more. Rather than viewing such events as idiosyncratic, it sees them as symptoms of potential systemic problems. Thus, it subjects them to root cause analysis. It traces the causal stage back through the system. The "5 Whys" slogan from the Toyota Production System suggests as a rule-of-thumb that the analysis goes back five stages. The goal is to use the event as a learning opportunity by exploiting its diagnostic significance.

For example, a *Brady* violation could signal a need for training of a particular prosecutor, a need for better information technology to track evidence and disclosure, or better communication between police and prosecutor, or clearer assignment of responsibilities for *Brady* compliance among those responsible for a case. Assigning blame to a particular prosecutor will not necessarily distinguish among these explanations, nor will sanctioning the prosecutor or compensating the defendant guarantee that the problem will be remedied. Root cause analysis insists on ambitious diagnosis and remediation.

Such practices, however, are little developed in prosecution. The courts rely mainly on end-of-the-pipe punitive and compensatory remedies. A defendant who can show misconduct may get evidence suppressed or a case dismissed or damages. However, such remedies are available only in the case of actual tangible harm, and they require proof that is often unavailable. Moreover, they have small deterrent effect, since the responsible officials virtually never bear their costs. Post-conviction exonerations have been numerous in recent years, often prompted by DNA analysis. The discovery of a wrongful conviction is an unexpected adverse event of the sort that would prompt root cause analysis in many fields. Hospitals, for example, conduct searing "mortality-morbidity" reviews in comparable circumstances. But no such practice is standard in prosecution.[16]

Disciplinary sanctions for prosecutor misconduct are rarely considered, much less applied. Justice Thomas in the *Connick* case did not even consider it relevant to ask whether the office in question had a functioning disciplinary process or what the likelihood was that the Louisiana bar would sanction an erring prosecutor. Judge Alex Kozinski of the Ninth Circuit recently expressed great frustration at this situation and took the extraordinary step of demanding

[16] *See, e.g.*, James Liebman, *The Overproduction of Death*, 100 Columbia Law Review 2030 (2000).

that the California attorney general explain why her office had not filed criminal charges against a prosecutor who lied in an early stage of the case before him.[17]

It is important to sanction willful violations, but doing so is not adequate, and the focus on egregious cases can have perverse effects. Post-bureaucracy urges intervention, not just to induce compliance with clear obligations, but to promote learning. Many events that do not imply willful wrongdoing may yield diagnostic intervention. A search for causes can be informative and lead to valuable reforms. The "after action" reviews undertaken sometimes in high-profile cases can facilitate valuable inquiry of this sort.[18] If retrospective inquiry is predominantly associated with punishment or humiliation, it may have two unfortunate effects. Actors will hide or misreport information for fear it will used to inculpate them. And peers will be reluctant to express reservations about each other's performance, since criticism implies incompetence or immorality.

C. Peer Review

In the broadest sense, peer review refers to review of practice decisions by people working in the same field as those who made the decisions in question. As such, it overlaps the other elements of the post-bureaucratic approach. Here, however, I use the term more specifically to refer to relatively intense and qualitative review by peers of representative or exceptionally challenging decisions or practices. "Peer" is a capacious and somewhat ambiguous term in a world where decisions are typically multidisciplinary. The key desideraturm is that review involve people working on comparable problems. Police officers or social workers might be appropriate members of a peer review team for a prosecutor or prosecutorial office. Reviewers could come from inside or outside the office.

Peer review is above all a learning process. The lawyer under review learns both by self-assessment and explanation of his decisions and by critical response from the reviewers. At the same time, peer review promotes the exchange of information across lawyers in the same office, and where the reviewers are outsiders, across offices. This means that lawyers can learn alternative approaches and benefit from others' experiences with them. It also tends to make

---

17 *Johnny Baca v. Derral Adams*, YOUTUBE (Jan. 8, 2015), https://www.youtube.com/watch?v_2sCUrhgXjH4.; see also the critique of de facto immunity for prosecutorial misconduct in Alex Kozinski, *Preface: Criminal Law 2.0*, 44 Geo. L. J. Ann. Rev. Crime. Proc. xxxv-xli (2015)

18 Erin Murphy and David Sklansky, *Science, Suspects, and Systems: Lessons from the Anthrax Investigation*, 8 Issues in Legal Scholarship 1, 34-39 (2009).

practice decisions more consistent to the extent the peers develop a shared understanding that informs their decisions.

Peer review is most extensively developed in medicine. It takes various forms. In addition to "mortality-morbidity" reviews of adverse events, there is professional recertification review in which a particular practitioner's practice over a period of time is examined. And there is institutional certification review in which a hospital's operations and structures are assessed periodically. In addition, peer review can focus on particular practices; new treatments, where formal clinical trials are impracticable, are assessed through informal peer discussion.[19]

All these variations could be readily applied to law and to prosecution in particular. Perhaps the most ambitious involve the kind of intense qualitative discussion of particular cases of a sort exemplified in "morality-morbidity" reviews. The review need not be focused on cases with bad outcomes. It could draw random samples of cases, stratified to capture relevant categories where appropriate. Kathleen Noonan, Charles Sabel, and I have described such a procedure employed by social workers in some child welfare systems.[20] It is hard to find ambitious versions of such systems in law, and they are sometimes actively resisted. Gary Bellow once proposed and experimented with a version of such a system among civil legal aid programs. Observers were surprised both by the volume of errors or suggestions for improvement that reviewers found or made and by the amount of resistance by practitioners to the process, even when it was divorced from personnel or compensation decisions. [21]

Outside reviewers in law may create risks to preserving confidentiality. These concerns are less severe with prosecutors than in other areas. Since the client – the government -- has a long-term interest in the quality of its lawyers' work, consent should be easier to get. And in any event, the concern is also present in medicine but has been overcome there with the help of facilitative legislation.

Even when review is done by insiders,, the bar has tended toward indifference if not hostility. The American Bar Association flirted with the idea of prescribing that firms institute

---

[19] Robert J. Marder, Effective Peer Review: The Complete Guide to Physician Performance Improvement (3d ed. 2013).

[20] Kathleen G. Noonan, Charles F. Sabel, and William H. Simon, *Legal Accountability in the Service-Based Welfare State: Lessons from Child Welfare Reform*, 34 Law and Social Inquiry 523 (2009).

[21] Gary Bellow, *Turning Solutions Into Problems: The Legal Aid Experience*, 34 NLADA Briefcase (August 1977); available at http://www.garybellow.org/garywords/solutions.html.

internal peer review procedures but quietly gave up the idea.[22] The bar has moved beyond the traditional idea that key professional learning takes place prior to certification by mandating "continuing legal education". But these programs, even when well prepared, rarely focus on particular practice decisions in richly observed contexts.

D. Performance Measurement

Performance is measured by translating the institution's goals into metrics and then periodically applying them. This was once a radical idea in the professions, and it is still controversial, but it has received increasing attention. Writing about a major federal gun-control initiative, Coles observes: "[I]t is likely the case that most U.S. Attroneys knew little about their cities' homicide rates. Project Safe Neighborhoods has changed that, prompting attention to the nature of decline (or increase) in [their] jurisdiction's homicide rate."[23]

Performance metrics can measure process (such as charges filed) or outcome (such as convictions, or looking to ultimate outcomes, crime rates). Process metrics indicate whether plans are being implemented; outcome metrics indicate whether they are working. Without the process metrics, we don't know what practices are contributing to the outcomes; without the outcome metrics, we don't what the effects of the interventions are. A good set of metrics includes both types in a "balanced scorecard".[24]

Metrics can be used to induce compliance with instructions, or they can be used diagnostically to revise and adapt instructions. The two functions are not entirely complementary. In order to use metrics to reinforce incentives, one needs to be confident about what practices one wants to induce people to undertake, and one must be able to define them with reasonable precision and comprehensiveness. Metrics attached to rewards and sanctions can have well-known perverse effects, especially when the metrics are incomplete. They may drive behavior to goals captured by the metrics and away from ones not captured. "Teaching to the test" in education is the classic example. Maximizing convictions is the corresponding phenomenon in prosecution. Conviction rates alone do not tell us whether convictions were achieved ethically, how much resources were used to obtain them, the collateral social costs of

22 Susan Fortney, *Am I My Partner's Keeper: Peer Review in Law Firms*, 66 U.Col. L. Rev. 329 (1995).

23 Coles, cited in note 13, at 154.

24 David Norton and Robert Kaplan, The Balanced Scorecard (1992).

the convictions, the relative priority of the crimes prosecuted, or the deterrent effect of the convictions.

In situations where there is uncertainty about the relevant practice and monitoring is designed to facilitate learning, the stakes for individuals have to be lowered. Metrics have to be provisional, and provisional metrics do not fit well with rewards and punishments because low scores are as likely to reflect the inadequacy of the measures as the quality of the performance.

From the learning perspective, metrics have three functions. First, the process of defining the metrics and interpreting their application structures and disciplines ongoing assessment of the relevant practices. For example, discussion has recently arisen with respect to Compstat-style assertive policing regimes about whether the number of arrests should be treated as a measure of success or as a cost. Many departments have viewed it as a measure of success, but critics assert that this practice ignores the harm such arrests do in creating criminal records that impair the life chances of a broad segment of the community. Requiring that the program specify metrics may cause the issue to surface earlier and the discussion to become more precise.

Second, measurements produce information about the system that can guide reform. Pretrial juvenile detention is an interesting case because there are only two permissible grounds for such detention -- likely failure to appear for court proceedings or re-offense – and both are easily observable. Thus, once decision criteria are made explicit as they are in the scorecards, their predictive power can be readily studied. The Juvenile Detention Alternative Initiatives regime mandates that the criteria be validated initially and periodically thereafter in the light of experience. The validation studies are sometimes quite sophisticated, and the scorecards have often been revised.

Third, the metrics, when applied across comparable institutions or individuals, indicate relatively effective and ineffective actors. The relatively successful are studied for lessons about what produces success. The laggards are subject to intensified supervision and technical assistance. In the diagnostic perspective, failure is presumed until proven otherwise to result from incapacity rather than willfulness.

The most common use of metrics in prosecution appears to be in assessing the relative effectiveness of individual prosecutors for promotion purposes.[25] However, there are more

[25] U.S. Government Accountability Office, GAO-04-422, US Attorneys: Performance-Based Initiatives Are Evolving (2004); see also M. Elaine Nugent-Borokove et al., Exploring the Feasibility and Efficacy of Performance

ambitious efforts. Some offices monitor charging practices in order to ensure consistency and compliance with policies about evidence quality and prosecution priorities.[26]

The most sophisticated efforts combine aggregate metrics with ongoing rule revision, root cause analysis, and peer review. Examples can be found in initiatives inspired by the Vera Institute Project on Racial Justice. In the manner of the Juvenile Detention Alternatives Initiative, the program prescribes ongoing monitoring of racial disparities in the effects of prosecution practices, root cause analysis of disparities, and scanning for reforms that might mitigate the disparities. In Milwaukee, for example, sophisticated implementation that included revisions of charging practices and the development of diversion programs has significantly reduced disparities.[27]

Another example is the "focused deterrence" strategy that starts by identifying violence-prone actors through intensive surveillance and then offers them a package of moral exhortation, threats of prosecution for past offenses, and offers of social services (for example, job training or substance abuse treatment). A distinctive component of the regime is the "call in" which invites (or in the case of probationers, requires) attendance at a meeting where prosecutors, community leaders, police, and social workers make presentation. In addition to creating tangible incentives for compliance, the intervention is designed to leverage peer relations by threatening or promising group punishments or benefits. Many focused deterrence regimes have been studied with rigor. An example from Cincinnati illustrates how measurement has been sufficiently fine-grained to yield information useful for reconfiguring the program to eliminate or revise specific ineffective elements.[28]

IV. Democracy

We generally think of democratic accountability in terms of elections or the more diffuse pressures of public opinion. There is some ambiguity about the range of prosecutorial activity that should be controlled democratically. In some respects, prosecutors resemble judges. They make decisions of great consequence that should be made disinterestedly and reflectively on the

---

Measures in Prosecution and Their Application to Community Prosecution (2009).

[26] Ronald F. Wright and Marc L. Miller, *The Worldwide Accountability Deficit for Prosecutors*, 67 Washington & Lee Law Review 1587, 1614-18 (2010).

[27] Angela Davis discusses the Vera Initiative in her contribution to this volume. See also Vera Institute of Justice, A Prosecutor's Guide for Advancing Racial Equity (Nov. 2014).

[28] See Robin S. Engel et al., *Reducing Gang Violence Using Focused Deterrence: Evaluating the Cincinnati Initiative to Reduce Violence (CIRV)*, Just.Q. 1, 28-32 (2011).

basis of general, public, and prospective norms. Since public pressure can be infected by considerations that prosecutors are obliged to ignore in these decisions, it risks compromising fairness. At the same time, prosecutors are executive officials commanding resources and exercising discretion in ways that have broad impact on their communities. The public has a clear stake in the general efficiency and fairness of prosecutorial practice and in the ways prosecutors exercise discretion within the interstices of enacted law.

Public accountability seems most productive and least dangerous to fairness values when it focuses on general patterns of practice rather than individual decisions. Unfortunately, this has not been the traditional focus of discussion. Prosecution often has a low profile in elections and public debate. Incumbents running for re-election are often unchallenged and usually re-elected. Moreover, where there is appraisal, it tends not to focus on general patterns. Discussion of practice tends to focus on a few high-profile cases. Otherwise, discussion is pre-occupied with the background qualifications and character of the candidates.[29]

This situation is in part a function of the traditional conception of prosecutorial work that emphasizes individual, ineffable judgment. We have seen that the traditional conception puts great emphasis on character and qualifications because it assumes that individual judgments are difficult to assess. In addition, the traditional conception assumes that judgment is necessarily idiosyncratic and ineffable; so it resists efforts to cabin discretion through explicit rules or to measure its effects. Practice under these assumptions is necessarily opaque.

The post-bureaucratic trends in the organization of prosecutorial discretion have two broad implications for democratic control of prosecutorial power. First, the basic tendency of post-bureaucratic reform is to make the broader system transparent in a way that increases control and adaptive capacity by insiders and outsiders alike. These reforms potentially enhance both fairness and accountability. Charles Sabel and I have argued that there is (or should be) a duty of responsible administration that requires administrators to adopt reforms to manage transparently so that courts and citizens can assess their compliance with substantive norms.[30] We find this duty in convergent themes of constitutional, statutory and common law, as they have been applied to a range of public institutions, including, prisons, police departments, and welfare programs. Courts have been reluctant to put such pressure on prosecution offices, in part

---

29 Ronald F. Wright, *How Prosecutor Elections Fail Us*, 6 Ohio State Journal of Criminal Law 581 (2009).

30 Cited in note 14 above.

because of the persistence of the traditional conception of prosecutorial judgment and the related assumption that accountability must take bureaucratic forms that would rigidify practice inappropriately. But initiatives from prosecutors themselves have demonstrated that there are ways of structuring discretion that enhance transparency without strait-jacketing practice. Courts could draw on these efforts to induce reforms by recalcitrant offices.

The "duty of responsible administration" idea runs directly counter to arguments like Price's that find the self-conscious structuring of prosecutorial discretion as an illegitimate assertion of law-making powers by executive officials. Price's argument implies that internal regulation enhances the power of senior administrators, rather than making it more accountable. This is wrong. A top administrator who wants to impose her will on the frontline has many tools for doing so without rule-making and transparent forms of review. She can, for example, make hiring, promotion, and compensation decisions on the basis of low-visibility signals of loyalty to her goals. Moreover, even where top officials leave broad autonomy to the frontline, there is no reason to assume that frontline decisions are benign. Without structure, frontline decisions may reflect the prejudices of the agents or may turn out to vary in arbitrary ways.

Second, the post-bureaucratic reforms often appeal to a conception of democracy somewhat different from the one that emphasizes elections. This alternative conception has attractive features, and it suggests the possibility of a thicker form of political legitimation.

The alternative conception is stakeholder democracy. Here decisions should be made, where feasible, locally by the people most affected and knowledgeable about them. General elections are inadequate both because they bundle far too many issues for people to make and register informed decisions about, and because they weigh all votes equally on all issues without regard to intensity of knowledge or interest.[31] (Some account of intensity is taken in the design of jurisdictions and the assignment of issues to them, but within even local jurisdictions there is a wide variation in knowledge and personal stake on many issues.)

Stakeholder democracy has to deal with the problem of who to admit to participation and how to reconcile differences when stakeholders disagree. But to the extent that representatives of diverse interests can come close to consensus on local interventions, they may confer a kind of democratic legitimacy that is unavailable in other processes. Even when stakeholders do not

[31] See the discussion of the "problem of intensity" in electoral democracy in Robert Dahl, A Preface to Democratic Theory (1956).

agree, their engagement may produce information that can influence official decision in ways that make it more acceptable.

The stakeholder conception resonates with various initiatives associated with "community prosecution". These initiatives are driven by the perceptions that, to the extent that the process is concerned with justice for victims, it should be more directly responsive to them; and to the extent it is concerned with deterrence, its efforts are most efficiently configured when coordinated with actions of other institutions and citizens and when they are configured in the light of information that can best be extracted through broad consultation.

In the stakeholder conception, the legislature's role is not to authorize specific decisions prospectively. Rather, the legislature sets basic parameters and provides resources for local deliberations and for central review of their efficacy. The legislature then retrospectively assesses the success of various intervention, perhaps mandating continuing experimentation where they are ineffective and perhaps codifying or promoting specific ones where success has been demonstrated.

Stakeholder participation is not necessarily beneficial. It can involve unproductive and expensive process costs and capture by unrepresentative sub-constituencies. But it has the potential to vindicate a different but complementary ideal of democracy from the one usually assumed in discussion of prosecutorial discretion.