Accountability, Quantification, and Law

Wendy Nelson Espeland and Berit Irene Vannebo

Department of Sociology, Northwestern University, Evanston, Illinois 60208;
email: wne741@northwestern.edu, b-vannebo1@northwestern.edu

Annu. Rev. Law Soc. Sci. 2007. 3:21–43

First published online as a Review in Advance on June 21, 2007

The *Annual Review of Law and Social Science* is online at http://lawsocsci.annualreviews.org

This article's doi:
10.1146/annurev.lawsocsci.2.081805.105908

1550-3585/07/1201-0021$20.00

Key Words

quantitative authority, professional discretion, sentencing guidelines, regulatory impact assessment, law school rankings

Abstract

Accountability can mean many things, but increasingly it is linked to quantification. This is true in many fields, including law. This review considers how the recent emphasis on quantitative accountability has influenced law and legal practices. Rather than offering a broad survey of quantitative techniques deployed in law, the article examines three legal contexts in which quantification has shaped how actors are held accountable: sentencing guidelines, cost-benefit analysis in regulation, and law school rankings. The conditions that promote rigorous quantification, its effects on professional discretion, relations of authority, and resistance are examined. The article suggests fruitful questions and strategies for analyzing more broadly the effects of quantification in law.

INTRODUCTION

Accountability is an old idea. As von Dornum (1997, p. 1483) argues, the ancient Greeks were "obsessed with keeping their officials legally accountable for their actions in office," and as early as the eighth century BCE, long before the emergence of classical Athenian democracy, there was an important archaic ideology of accountability emphasizing rectitude, visibility, and participation that regulated official conduct. The meaning of accountability has changed historically, but it has often been associated with governance, in which those with political power must somehow justify their actions to those subject to their power.

The idea of accountability, broadly construed, is fundamental to much law. Law is a site for interpreting accountability, a vehicle for establishing it, and sometimes its target. Accountability in law is conceptualized differently, but a great deal of law is intended to hold people and institutions responsible for their actions and accessible to their constituents. Law creates the infrastructure for political accountability and representation in government. Public international law plays a parallel role at the global level, creating rules of conduct for states and international organizations. Tort law creates and enforces the concept of liability. In the United States, the standard of the "reasonable person" helps define situations in which people are responsible for harm done resulting from their actions. Law of fiduciaries describes the duties of trustees who act on behalf of individuals or institutions. Contracts hold people and organizations responsible for doing what they say they will. Accountability is also central to criminal law, which determines the conditions under which people are held responsible for the crimes they commit. If someone is acting in self-defense, if their actions are unintended, or if they are mentally incompetent, their accountability is mediated. Law provides a rich and diverse language of accountability.

Understood as creating responsible people and accessible, responsive institutions, accountability is obviously a desirable goal. But the terms of accountability are dynamic and contested, and accountability can take on a more ominous tone. "Technologies of audit and accountability" create "new forms of governance and power" (Shore & Wright 2000, p. 57) and can consume vast resources; they produce unintended and sometimes harmful consequences; and they can transform the institutions and the self-conceptions of the people they target, sometimes in undesirable ways.

Because accountability is such a sweeping idea, we offer a highly selective treatment. We begin by describing three related trends. First, interest in accountability, especially in the past two decades or so, has increased dramatically. Second, its meaning is generalized beyond governance narrowly construed and is now associated with monitoring the performances of actors in many institutions. Third, accountability has become more closely associated with quantification and measurement. After briefly describing the first two trends, this article focus on the third, considering how the recent emphasis on the construction of accountability as forms of measurement has influenced law and legal practices. Rather than offering a broad survey of the various quantitative techniques deployed in law, we offer examples of how quantification has shaped how legal actors are held accountable in three different contexts, sentencing guidelines, the use of cost-benefit analysis (CBA) in regulation, and media rankings of law schools, focusing more closely on the first. Our examples are intended to be suggestive; by highlighting important dimensions and effects of quantification in these legal practices, we hope that other scholars will investigate similar or divergent patterns in other contexts. We conclude by identifying some conditions that mediate the flow of quantitative authority in law and suggest some fruitful questions that scholars might pursue.

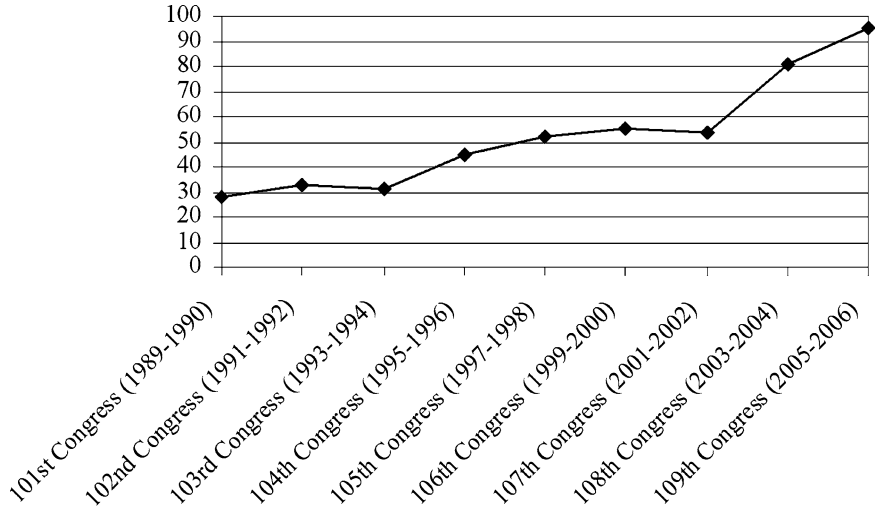

Figure 1

Graph based on Library of Congress records, created by counting legislation with "accountability" and "act" in the title from 1989–2006. Acts introduced in both the House and Senate are counted only once.

TRENDS IN ACCOUNTABILITY

Accountability has become a term to reckon with. For Williams (1976), a word or expression becomes a keyword when it dominates an issue or area, synthesizes or symbolizes a world view, or serves as a site of struggle over the meaning of goals or practices.[1] Accountability is a keyword in politics and policy. Demands for greater accountability are routine features of media, scholarship, and legislation. Efforts to promote a wide range of desired characteristics including excellence, productivity, efficiency, transparency, and justice are now often framed in terms of accountability. Accountability, it seems, has become an all-purpose solution to many problems.

The emergence of this trend varies by field. In U.S. education, Ohmann (2000) sees the 1970s as a pivotal moment for rising interest in accountability. Many scholars note a mounting concern with accountability since the mid-1980s. Power (1997, p. 3) points to an explosion in the term "audit" in the UK in the 1980s and 1990s, with more people subject to formalized styles of checking that involve creating verifiable accounts of behavior. For Strathern (1996, p. 1), "Audit and assessment seem the ubiquitous tools of accountability," and many see this audit explosion as reflecting and fostering the spread of a global "audit culture" (Strathern 2000, Shore & Wright 1999).

In the United States, this increasing concern with accountability is reflected in legislation. **Figure 1** shows a steady increase in the number of acts with accountability in the title introduced in Congress since 1989. The range of these bills is impressive, from the Federal Reserve Accountability Act to the Bounty Hunter Accountability and Quality Assistance Act. Because bills may be introduced for purely symbolic reasons, the introduction of legislation is an imprecise indicator; yet it seems significant that sponsors reframed legislation in terms of accountability.

Accountability is often associated with responsive political leaders and governmental institutions, but interest in accountability encompasses many different institutional domains. Few organizations and professionals are now exempt from pressure to demonstrate accountability, and law often mediates the terms of accountability. Laboratory scientists, for example, recently entered into complex negotiations with regulators about how to comply with environmental laws, the

[1]Emanuel & Emanuel (1996, p. 229) make this point about accountability in health care. Shore & Wright (1999, p. 558) describe audit as a keyword.

results of which may introduce "audit culture into the heart of modern science" (Silbey 2003, p. 3). Human rights advocates are under pressure to create indicators to hold states accountable (Rosga & Satterthwaite 2003). In 2005, 11 NGOs signed the International Non-Governmental Organisations Accountability Charter to demonstrate their commitment to transparency and accountability. Corporate and accounting scandals at firms such as Enron, WorldCom, and Tyco prompted Congress to pass the Sarbanes-Oxley Act in 2002 to improve financial disclosure and corporate governance. Beginning in New York in 1994, states began issuing public scorecards for physicians and hospitals performing coronary artery bypass surgery or angioplasty. Evidence about the scorecards' effects is mixed (Epstein 2006), but one study finds that they influence clinical decision making; 79% of the New York cardiologists reported that their decision to perform angioplasty or to intervene in critically ill patients was influenced by scorecards (Narins et al. 2005). As these examples suggest, the stakes of accountability are often high.

Accountability is a variable social and ethical relationship, and its terms are often ambiguous, hotly debated, and linked to diverse practices. Accountability can be informal or formalized, grounded in discretion, expertise, or standardized routines. It has been associated with voting, responding to constituents, and informing consumers; it is enacted with guidelines, inspections, certification, and public forums or even by carefully listening to clients. But, increasingly, pressures for accountability are linked to formalized quantitative measures that are designed to evaluate performances, facilitate decision making, or constrain discretion. Measures like cost-benefit ratios, performance indicators, rankings, benchmarks, scorecards, and standardized tests are used to evaluate behavior and explain decisions of many different organizational actors. Numbers circulate easily and make possible "government at a distance" (Miller & Rose 1990, p. 9).

CBA, for example, was first widely adopted by the Army Corp of Engineers and the Bureau of Reclamation (cf. Porter 1995, pp. 148–89). When disputes between these rival federal agencies generated embarrassing differences in the calculations of benefits for the same water project, with each agency's benefits reflecting their specialty, Congress responded with the Flood Control Act of 1936. This law required agencies to show that the benefits of proposed plans for federal water development exceeded their costs, which led to a dramatic expansion of benefits attributed to water development. It took decades for methods for conducting CBA to become more standardized, but now CBA is so naturalized that it has become the dominant logic of government.

The spread of quantitative measures is especially pronounced in health care. Physician scorecards are one example of this trend. The introduction of managed health care plans in the United States in the 1970s increased demands for information about the quality and affordability of health care plans. For example, in 1995, the Foundation for Accountability was created to evaluate performance measures for health care plans and to lobby for more accountable health care.

If, as Hoffer (2000, p. 529) suggests, accountability is understood as providing reasons for policies, schools have always been accountable in some fashion. But accountability in education has become more formalized in the past 20 years with the growing use of measurable outcomes, increasingly defined by scores on standardized tests. In the 1950s, standardized tests were used mainly to characterize and place students; however, this changed dramatically after the 1983 report *A Nation at Risk* (Natl. Comm. Excellence Educ. 1983) portrayed failing schools as threatening economic viability and pushed for more standardized assessments (Hoffer 2000, p. 533). Standardized testing was increasingly linked to teaching practices, which supporters refer to as measurement-driven instruction and critics call teaching to the test. Tests

are now used to evaluate schools and school systems, as well as students, shifting attention from individuals to institutions. The No Child Left Behind Act of 2001 amplified these trends.

What explains this proliferation of quantitative measures? Why has accountability become so aligned with measurement? Reasons for quantitative accountability vary across fields, but explanations often emphasize several conditions. For Porter (1995), the spread of quantification in decision making reflects pressures for "mechanical objectivity," forms of standardized knowledge that constrain the local, subjective, and personal. Rigorous quantitative methods entail highly structured rules for manipulating numbers, numbers that require enormous labor, coordination, and discipline to produce. Mechanical objectivity both shapes and reflects particular administrative and political cultures. Pressure for mechanical objectivity arises when decisions are subject to public scrutiny, especially from powerful outsiders, when there is conflict, when elites or experts are distrusted, or when communities must coordinate across social or geographical distances. Mechanical objectivity confers a robust, defensible authority because it makes the biases or personal characteristics of those making decisions less relevant. The loss of autonomy explains why mechanical objectivity is usually imposed on decision makers who cannot defend their authority, groups Porter characterizes as weak elites. Under these conditions, trust in persons is replaced by trust in numbers.

Others also see trust as mediating demands for accountability. Power (1997, pp. 134–35) views the emergence of the "audit society" as responding to a desire for the reassurance that ritualized monitoring provides. Auditing, which relies heavily on quantitative assessments, replaces trust in particular people with trust in auditing organizations, a shift comparable to Porter's. Ranson (2003, p. 460) suggests that different methods for ensuring accountability represent different ways of "securing trust in the public sphere."

Other scholars emphasize neo-liberal politics as an impetus for accountability and quantification (Power 2003, p. 191). Neoliberalism is both a source of distrust and a set of strategies for securing it. Shore & Wright (1999) trace the audit explosion in Britain to the conservative Thatcher regime and its efforts to shrink the welfare state. Proponents assumed that market controls were superior to bureaucratic ones. After privatizing a broad array of public services, the shrunken public sector was subjected to techniques of accountability that characterized the private sector. Accountancy and audits became vehicles for inserting the practices and norms of the private sector into the public realm.

Building on Foucault, Shore & Wright (2000, p. 61) describe this regulation as "neoliberal governmentality" in which demands for accountability are manifest as "political technologies" and market norms organize the economy, the state, and even "the conduct of individuals."[2] The paradoxical effect of this was an extraordinary expansion of governmental regulation. International financial institutions such as the World Bank and the International Monetary Fund, in pressuring member countries for transparency and accountability, have also contributed to the proliferation of quantitative measures. Having described reasons for the proliferation of quantitative accountability, we now turn to its implementation in law.

FEDERAL SENTENCING GUIDELINES

Sentencing guidelines are one area where quantification has deeply affected legal practices. Guidelines are a vivid example of efforts to create quantitative accountability. Creating more uniform sentences for those committing similar crimes was a central goal for

[2] Rose et al. (2006) see affinities between neo-liberalism and some techniques of governmentality but caution that the broad category of neo-liberalism obscures the specificity and vibrancy of governmentality.

sentencing guidelines. A key strategy for doing so was to limit the discretion of judges in sentencing people convicted of crimes. Although the language of accountability was not used to promote sentencing guidelines, the motives for guidelines are similar to the goals for accountability as it is commonly discussed today: Guidelines were intended to improve sentencing performance, provide oversight, make sentencing a more visible and reviewable process, and create uniformity (supporters eventually did adopt the language of accountability and transparency to defend them). Quantification is fundamental in federal sentencing guidelines as a technique for simplifying and classifying the characteristics of offender's and their crimes and as a means of restricting the range of sentences.

Policy makers have tried to constrain discretion in different jurisdictions with laws requiring mandatory minimum sentences, standardizing parole practices, and abolishing parole, but federal sentencing guidelines have been the most transformative and controversial pieces of sentencing policy. Federal sentencing guidelines are a product of the Sentencing Reform Act (SRA) that was part of the Comprehensive Control of Crime Act of 1984 (many states have imposed sentencing guidelines, but we focus on federal sentencing guidelines). The SRA abolished parole, made the sentences issued by district court judges reviewable by appellate courts, and created a new, independent administrative agency in the judicial branch, the U.S. Sentencing Commission (SC). The SC was charged with creating and implementing sentencing guidelines. The guidelines established by the SC in 1987 were mandatory until the Supreme Court ruled them unconstitutional in *United States v. Booker* (2005), which made the guidelines advisory. Guidelines were ruled unconstitutional for violating the Sixth Amendment right to trial by jury because they allowed judges to include facts in sentencing decisions that had not been presented to a jury. Both critics and supporters agree that sentencing guidelines have had a profound influence on the practice of federal criminal law. For Stith & Cabranes (1998, p. xi), they initiated a "revolutionary new system for punishment of federal crimes"; Bowman (1996, p. 680) contends that "[o]ne can hardly overstate the significance of the Guidelines, not merely for judges and judging, but for every aspect of federal criminal practice."

For most of the twentieth century, U.S. sentencing policies were informed by a penal philosophy emphasizing rehabilitation rather than retribution. Rehabilitation required that sentences be individualized to reflect the characteristics of the offender and the crime that was committed. Sentences were indeterminate in several senses. Sentences could be unpredictable because judges had broad discretion in issuing sentences. As long as they remained within statutory limits, there were almost no constraints on judges in sentencing. Furthermore, discretionary sentencing, was made more uncertain by parole officers' modifications of prison sentences.

Beginning around the mid-1960s, indeterminate sentencing came under fire from critics who objected to what they saw as disparate and irrational sentences that judges imposed on offenders committing similar crimes [Koh (1992), Bowman (1996), Stith & Cabranes (1998), and Nagel (1990) give useful summaries of the sentencing reform movement and the legislative history of the SRA]. Supporters of sentencing reform included a broad assortment of people, including academics, politicians, law enforcement, and even some judges, most notably Marvin Frankel (1973). On the left were those who believed that the sentence disparity reflected discrimination based on race or poverty, who advocated prisoner's rights, and who believed judges were too punitive and too reliant on prison as punishment. On the right were those who considered judges and parole boards too lenient, who thought sentences were too uncertain, and who saw the rising crime rate as evidence of a failing criminal justice system. A strong consensus emerged that rehabilitation

was not working and that discretion, especially judicial discretion, should be curtailed.

Passing the SRA was the major achievement of the sentencing reform movement. Senator Ted Kennedy, its original sponsor, had introduced multiple versions of the bill since 1974. After important modifications, the SRA enjoyed strong bipartisan support. President Reagan signed the law in 1984 and appointed the first sentencing commissioners. Part of the law's appeal was in demonstrating toughness on crime when crime control was deeply politicized. One of the main goals for the law was to "avoid unwarranted sentence disparities among defendants with similar records who have been found guilty of similar conduct" [18 U.S.C. §3553 (a)(6)]. The legislative record makes clear that constraining judicial discretion was seen as vital to this goal. The SC took this position to heart in issuing its guidelines.

How Sentencing Works Under the Guidelines

Sentencing under the guidelines culminated in the Sentencing Table, which defines for district court judges the range of sentences for offenders (U.S. Sentencing Commission 2006, p. 381; all examples are from this edition). The vertical axis of the grid measures the seriousness of an offender's current crime based on 43 offense levels. The horizontal axis measures an offender's criminal record based on a scale of 1–6. The intersection of these two points determines the guidelines range for the sentence, expressed as months of prison. Judges decide within that range how long a sentence to impose. The SRA restricts the guideline range so endpoints cannot vary by more than 6 months or 25%. Along both axes, bigger numbers mean longer imprisonments. At first blush, the one-page Sentencing Table seems straightforward. That impression is incompatible with using the table. Identifying the boundaries among the 258 boxes that comprise the grid is a daunting challenge. This is borne out by the enormous *Sentencing Guidelines Manual* that provides the instructions for how to calculate the offense level and the criminal history (the 2006 *Guidelines Manual*, including the appendices, supplements, and index, exceeds 1800 pages; the 2005 paperback edition weighs more than 5 pounds).

Offense level is determined by three factors: the base offense level, the specific offense characteristics, and adjustments (Bowman 2005, p. 1325). Offense conduct is divided into 18 categories of offense, which are subdivided into 53 additional categories. For each category of crime, the guidelines identify a number that constitutes the base offense level, a ranking that represents the seriousness of the statutory crime of conviction. Depending on the specific offense characteristics, the baseline number is adjusted, usually upward, according to the number of points assigned to that offense characteristic. For example, the base offense level for aggravated assault is 14 (§2A2.2). There are six offense characteristics for aggravated assault. If the assault involved more than minimum planning, two levels are added. If a victim sustained bodily injury, three levels are added; if the injury is permanent or life threatening, seven levels are added.

The horizontal axis quantifies a defendant's criminal history. The six levels of criminal history are determined by the criminal history points a defendant is assessed for prior convictions. A defendant gets 3 points for each prior prison sentence that exceeds 13 months, or 2 points for sentences of more than 60 days but less 13 months (§3E 1.1).

Adjustments are modifications of the offense level that reflect characteristics that may apply to many kinds of offenses (§3.2). Adjustments are based on characteristics of the victim (e.g., a government official), the role a defendant played in committing the offense (e.g., a leader), or whether the defendant was obstructing justice or was convicted of multiple counts for the same crime. The most common adjustment is to decrease the level of offense by one or two levels for defendants who accept responsibility for their crimes (e.g.,

plead guilty), but most other adjustments increase offense levels.

The guidelines also describe the limited circumstances that permit a judge to depart from the proscribed sentence. Departures are generally discouraged by the guidelines (Stith & Cabranes 1998, pp. 72–77); for example, aggrieved parties can appeal departures but not sentences within guidelines ranges, which makes district judges cautious. Factors that previously informed sentencing are forbidden under the guidelines. A defendant's age, charitable work, health, military service, family responsibilities, or employment history cannot be the basis for departures. If an offender provides substantial assistance to law enforcement, prosecutors may recommend a downward departure. The guidelines also permit departures if the criminal history category does not accurately represent someone's criminal history; however, downward departures are more restricted than upward departures. Since *Booker*, departures are less scrutinized.

This brief overview makes clear two features of guidelines. First, extraordinary effort and resources were required to promote and create uniform sentencing. The political desire for uniform sentences was the product of a broad social movement. Locating and excising sources of unwarranted discretion involved the sustained involvement of Congress and the Department of Justice and the creation of a new, well-staffed, and well-funded bureaucracy. Second, quantification is central to guidelines. As Stith & Cabranes (1998, pp. 68–69) put it, "The most common offense characteristic found in the Sentencing Guidelines is *quantity* . . . . [T]he severity of a sentence is heavily dependent on quantifiable factors such as the amount of drugs. . ., the amount of money stolen. . ., or the number of unlawful aliens." Quantification, with its long, complex association with bureaucratic authority, was the master strategy the SC used to contain discretion and rationalize criminal law. Quantification integrated disparate and often piecemeal federal criminal law by making commensurate the crimes and criminal histories of all convicted defendants and the harm done to victims. This quantification demanded radical simplification of the context of criminality and the people involved, simplification that was hardly simple to create because it was sustained by an elaborate edifice of definitions, classifications, commentary, and rulemaking. But sentencing uniformity is a moving target, requiring constant vigilance. As new crimes, new laws, and new characteristics of people become germane, categories must be expanded and articulated. The length of the guidelines has more than tripled since they were first issued. The dynamism of criminality along with a Congress devoted to conspicuous if not always effective displays of law and order continually threaten to overwhelm aspirations for comprehensive classification. The pursuit of a "gapless" formal rationality in sentencing, to use Weber's terminology, turned out to be a frustratingly elusive and costly aim.

The Guidelines in Practice

The challenge of interpreting the guidelines parallels the complexity of creating them. Judges across the political spectrum complained bitterly about how difficult it was to use the guidelines. In 1994, Stephen Trott (reprinted in Trott 1995) wrote a rather desperate letter to Richard Conaboy, chairman of the SC. As associate attorney general under Reagan, Trott was a strong proponent of guidelines and helped design them. Six years after serving as an appellate judge, he describes the guidelines as "impossible and unnecessarily complicated," a "cure that is worse than the disease." Trott concludes, "[W]e have to give more latitude to sentencing judges, and then within reason *trust their judgment*" (emphasis in original). Margolick (1992, p. 1) summarized judges' reactions: "They complain that the new approach has taken the judging out of judging and replaced it with an oppressively mechanistic regime, one with the abstruseness of the Internal Revenue Code

and overtones of Franz Kafka, George Orwell and Rube Goldberg."

Judges were not the only group to condemn the guidelines. After nearly 20 years of experience with the guidelines, observers and practitioners express strong consensus that as practiced they are a resounding failure. Most see the reform movement as addressing serious problems, and many agree that the SRA "brought law and due process" to sentencing (Bowman 2005, p. 1327). Even the harshest critics believe that appellate review of sentencing decisions has improved accountability in sentencing (Stith & Cabranes 1998, pp. 170–72). But now even former supporters conclude that the guidelines require fundamental reform (Bowman 2005). Notable exceptions to this consensus include top officials in the Department of Justice, conservative members of Congress, and SC members (Tonry 1996).

Diagnoses of guidelines failure vary, but most commentators agree that the guidelines are too mechanistic and often produce unfair, incoherent sentences. Most commentators think that for structured sentencing to work the system must seem legitimate to those who execute it. Because judges pronounce sentences for particular persons rather than abstract classes of criminals, their greater knowledge of offenders' lives and crimes means they are often more concerned with the distributive effects of sentencing than those more remote from defendants (i.e., appellate judges, the SC, the Justice Department, or Congress). If sentencing judges (and prosecutors) feel that the sentences dictated by the guidelines are unfair, they may try to circumvent them to produce more appropriate outcomes (cf. Freed 1992, Koh 1992, Tonry 2005, Bowman & Heise 2001).

Others suggest that the political influence of Congress and the Justice Department over the SC undermined the autonomy that the SC needed to monitor the guidelines and make corrections (Bowman 2005, Tonry 2005). Critics point to severe drug sentences as one example in which politics prevented necessary amendments. Some blame the sentencing guidelines, characterizing them as an unaccountable bureaucracy (Stith & Cabranes 1998); problems at the SC include commissioners' efforts to curry political favor to advance their careers (Tonry 2005), ignoring some SRA directives while going beyond others (Miller 2004), and failing to justify their decisions (Koh 1992). Others see even more fundamental problems with the guidelines. Stith & Cabranes (1998, p. 172) reject the core premise of the guidelines that disparity reflects systemic judicial biases and that efforts to produce a "mechanical" uniformity across judges promotes fairness or justice.

Even if the diagnoses of guidelines failure are still debated, many of their effects are not. Everyone agrees that the rehabilitative model has been discarded and that the guidelines have sharply curtailed judicial discretion. Judges' formal discretion is relegated to the range within categories defined by the grid. As Bowman (2005, p. 1333) points out, the more complicated the grid, the more severely judicial discretion is constrained because more boxes create finer distinctions in sentencing ranges. It is also indisputable that guidelines have propelled the explosion of prison populations in the United States. Under the guidelines regime, more defendants are incarcerated, they are issued longer sentences, and they serve longer portions of their sentences. Blumstein & Beck's (1999, p. 17) analysis of the growth of prison populations concludes that the "dominant contributor to current growth for all offenses is time served."

Scholars and practitioners also agree that the guidelines have not eliminated discretion in sentencing but have merely shifted its location and form. Prosecutors can now largely determine sentencing because they control which and how many charges to file; prosecutors also largely determine which facts about a defendant's crime and criminal history are relevant to sentencing, resulting in fact bargaining; and prosecutors control whether to file a substantial assistance motion that permits downward departures. As Miller (2004, p. 1252) puts it, "The overwhelming and

dominant fact of the federal sentencing system. . .is the absolute power the system has given prosecutors over federal prosecution and sentencing. There is a lot of evidence to support this claim but it can be demonstrated with one simple and awesome fact: Everyone pleads guilty."

The role of defense attorneys and probation officers has also changed since the guidelines. Instead of offering guidance to judges, probation officers now prepare the presentence reports that the guidelines require. These reports help establish the facts of the case and include the sentence as calculated by the officer on the basis of those facts. Because parole officers typically have no formal legal training (most are social workers), some see this independent fact-finding as inappropriate (Stith & Cabranes 1998, pp. 85–91). Defenders now warn clients not to speak with probation officers for fear that their disclosures may influence their sentence. Because defendants are rewarded for accepting responsibility, which is broadly defined as pleading guilty, they must now choose between a vigorous defense and the promise of a shorter sentence, which diminishes the role of defense attorneys and can potentially subvert a defendant's right to counsel (Etienne 2004).

Scholars disagree about other consequences of guidelines, including their impact on sentencing disparity. An SC study (U.S. Sentencing Commission 2004) finds that guidelines have reduced some disparity among judges and between regions for most types of crime and that, although some gender disparity remains, there is little evidence of racial or ethnic disparity. Other studies find that the guidelines appear to have reduced sentencing disparity within districts but that disparity among judges in different districts and regions remains, especially in drug sentences (Anderson et al. 1999, Hofer et al. 1999, Bowman & Heise 2001). Albonetti (1997) and Mustard (2001) find that characteristics of offenders still matter in federal sentencing (but see Blackwell et al. 2003).

In making guidelines advisory, the *Booker* (2005) decision has created enormous uncertainty in sentencing. It is too soon to know how this will affect sentencing, but early evidence suggests that both sentencing judges and appellate judges still use guidelines to determine and review sentences. Regardless of *Booker*'s eventual impact, it is useful to ask why these legal actors were the target of mechanical objectivity. Does their position support or contradict Porter's (1995) claims about the conditions that propel quantification in decision making?

The reactions of federal judges support Porter's basic point that mechanical objectivity is typically imposed on rather than embraced by actors. Although some judges, Frankel (1973) most prominently, supported sentencing guidelines, most were skeptical, and their objections increased with experience. Yet we do not typically think of federal judges as the weak elites that Porter characterizes as the targets of mechanical objectivity. How do we reconcile their status with their vulnerability?

One response involves examining the distinctive role of federal judges. Judges are not supposed to act like politicians, and most refrain from direct engagement in politics. Also, during 12 years of Republican presidents (1981–1993), many judges supporting judicial restraint were appointed, a philosophy that is hard to reconcile with conflict over congressional directives. Consequently, judges are typically constrained in their pronouncements about legal issues. Few judges testified about the proposed guidelines during hearings for the SRA. Their more circumspect politics was poorly suited for the macho one-upsmanship of "tough on crime" politics. Furthermore, it was too easy to portray judges' objections in an unflattering light: as fearing change, resenting their loss of discretion, and protecting their self-interest. Of course, judges are not immune to politics. Lifetime tenure is intended to buffer federal judges from the sway of partisan politics, but ambitious judges are mindful of the political

ramifications of their decisions. The same holds true for members of the SC and the Department of Justice. The political predispositions and the political savvy of these actors help explain the trajectory of amendments toward harsher sentences and tighter control, which clearly reflected the wishes of Congress and the Department of Justice (Stith & Cabranes 1998, pp. 49–66; Tonry 2005).

The political appeal of sentencing reform was another important factor in judges' vulnerability to guidelines. Crime is highly publicized and politicized in the United States, a standard trope in popular culture and the object of much academic research. High-profile crimes attract broad media attention, and public conflict over the rights of victims and defendants is common. Crime politics plays to media strengths and public fears, which can be manipulated. Being perceived as tough on crime is politically expedient while being weak on crime is disastrous, especially for Democrats. Also, sentencing reform is an issue that spans many cleavages. Conservative Republicans and liberal academics supported the SRA. Each constituency could, in the abstract, project onto the SRA the outcomes it most desired. When comparing sentencing decisions with less visible legal practices, it is less surprising that they were a site for intervention.

Quantification in the Guidelines Approach

Why was quantification so central to the guidelines? The influence of two founding commissioners helps explain some of the sway of quantification. Paul H. Robinson is probably the commissioner most responsible for the eventual structure of the guidelines. Although Robinson was the lone dissenter in the final guidelines adopted, their structure nonetheless responded to his vision. A law professor with a utilitarian and retributional philosophy of sentencing, Robinson argued that every increment of harm caused by a crime should correspond to an increment of punishment. This stipulation required an overarching sentencing calculus of harm and punishment in which each mediating factor is prescribed some quantitative weight. Michael Block, a professor of economics, argued that efficiency should be the goal of sentencing, balancing the social costs and benefits of sentencing. Neither could convince their colleagues to adopt their sentencing philosophy as the guiding principle of the guidelines, but their commitment to quantification as the medium for expressing variation in both harm and mediating circumstances was adopted (Nagel 1990; Stith & Cabranes 1998, pp. 51–59).

Other factors also enhanced the appeal of rigorous quantification. The broad authority we grant to numbers is surely relevant. We tend to see numbers as more objective than other forms of information, perhaps because of their association with the rigors of mathematics and science. We believe that rules for deriving numbers are more constraining, less easy to manipulate, than are rules for other forms of expression. Organizational scholars know that, as information circulates, the assumptions, biases, and uncertainties that inevitably inform its production are obscured; the further that information travels from those who make it, the more certain and robust it appears. Judges who calculate sentences using the guidelines are deeply aware of their arbitrariness, but to congressmen or citizens the resulting sentences seem appropriate and considered. In Weberian terms, the legitimacy of guidelines is a by-product of standardization and impersonality that is the hallmark of bureaucratic practice. One great virtue of mechanical objectivity, as Porter notes, is that its logic is easily reproduced, which makes hard decisions easier to defend.

Another reason for the emphasis on quantification in sentencing guidelines is that once it is generally endorsed, it privileges things that are easy to measure. Consequently, the amount of money, the weight of drug, or the number of past criminal events weigh so

heavily under guidelines. Despite the absurdities of a drug sentence in which the medium on which a drug is sold—sugar cube or tablet—can mean the difference in years of prison, measures beget measures.

The guidelines profoundly affected relations of authority in criminal justice. Judges lost autonomy in relation to prosecutors and parole officers, but there are other important actors to consider. First, the SC itself had wide latitude in implementing the guidelines, especially the founding members. Congress set constraints on their discretion, but they gained control over the terms and structure of sentencing. The SC was responsive to the directives of Congress, which had formal oversight, and to the Justice Department, which had an official advisory role and was a nonvoting member, but the SC also clearly gained power over judges. The guidelines provided Congress with a useful vehicle for expressing its concern with crime because it must approve amendments to the guidelines annually (Bowman 2005). Furthermore, the statistics the SC produced provided a new kind of scrutiny for Congress, one that made it easy for Congress to assess the performance of judges by monitoring their use of the guidelines, the rate of departures, and so forth. And the executive branch via the Justice Department could easily make its interests known through its advisory role.

Although curtailing judicial discretion was a goal of sentencing reform, the increased power of prosecutors and probation officers and the diminished roles of defense attorneys were unintended consequences. One important aspect of these shifts is that they disperse discretion, making it less visible than in the past (Stith & Cabranes 1998, p. 127). Plea bargaining, fact bargaining, pressure to plead guilty, and the production of presentencing reports are practices that are harder to scrutinize than the sentencing decisions judges render publicly. If, prior to guidelines, judges had "unfettered discretion," at least one advantage of this is that it was clearer who was responsible for the sentence.

The example of guidelines also suggests that we should be critical of the transparency of numbers. As the hefty guidelines manuals show, the numbers that are the end product of guidelines do not reveal the hundreds of decisions and elaborate work involved in deriving that number. The context in which numbers are produced often results in numbers that are far murkier than they appear. Quantification depends on sturdy definitions of what something is and how it should be treated, and definitions generally proliferate, as the guidelines did, when new circumstances or challenges arise. Efforts to create an encompassing system of quantification require an exacting discipline, elaborate coordination, and vast resources and, given the dynamism of human agency, can never be fully exhaustive.

Many critics decried the mechanistic quality of guidelines sentencing. Instead of permitting judges to consider all the relevant factors in a particular case, judges could attend to the restricted range of guideline factors, and only in a narrowly prescribed fashion. This changed both the locus and the framework of responsibility. As Heimer & Staffen (1998) show, responsibility is an organizational as well as an individual accomplishment. Judges' investment in a sentencing decision inevitably changes if they feel their hands are tied (Koh 1992). Responsibility shifts from judges to guidelines, from persons to numbers, becoming less contingent and more bureaucratic. Detailed knowledge of a case means that judges pronounce sentences on real persons. The connection judges feel between their decisions and the punishment meted on offenders is diminished to the extent that they feel they are not really making the decision. Discerning the effect of this is difficult, but judges may feel that they ultimately bear less responsibility for the outcome (Stith & Cabranes 1998, pp. 84, 169)

Guidelines also illustrate a distinctive feature of law: Judicial decisions are subject to appeal. Law possesses a fundamental alternative to quantitative accountability because

knowing that decisions will be scrutinized and potentially overturned by other judges encourages prudence. Critics and supporters agree that making sentencing subject to appellate review provided an important constraint on judicial discretion. Judicial review could accomplish the goals of reform without the machinery of guidelines.

COST-BENEFIT ANALYSIS AND REGULATION

Regulation is another part of law that has been profoundly affected by efforts to construct quantitative accountability. Since the Kennedy administration, the executive branch has tried various strategies to curb the discretion of federal agencies, limit regulation, and force agencies to be more accountable for their decisions. Most strategies involved requiring agencies to conduct some version of CBA of proposed regulations. These regulatory impact analyses (RIA) were supposed to discourage agencies from excessive regulation and encourage more transparent, rational, and accountable rulemaking. Most scholars agree that RIA has not accomplished those goals, but after decades of debate they still disagree about its potential to do so.

The Flood Control Act launched CBA in federal agencies, but the practice of CBA did not spread much beyond the water development agencies until the 1960s, when Robert McNamara initiated the Planning Programming Budgeting System (PPBS) in the Defense Department. PPBS was a tool for comparing the costs and benefits of alternative plans. President Johnson, impressed by the success of PPBS in the Defense Department, soon required all executive agencies to use it. This directive failed, mainly due to strong opposition from agencies, but the idea that CBA could rationalize decision making remained alluring to policy analysts (Fuchs & Anderson 1987, p. 25).

The 1960s and 1970s were a boom time for new regulation, much of which was aimed at protecting consumers, workers, and the environment.[3] Over time, businesses expressed their objections with increasing force. The Reagan administration's enthusiasm for deregulation prompted the most sweeping change. The administration believed that agencies had been captured by special interest groups and partisan bureaucrats favoring costly regulation and that their influence needed to be checked (Bagley & Revesz 2006). To do so, Reagan issued two executive orders, EO12291 and EO12498. The first requires all executive agencies to perform CBA to justify all major regulations and to prove that the benefits of proposed regulations exceeded their costs. EO12291 also authorized the Office of Management and Budget (OMB) to review these analyses, delay suspect regulations, and establish guidelines for conducting these analyses. No executive agency could publish notice of proposed rulemaking before OMB reviews were complete and agencies had responded to OMB's concerns. The second order, EO12498, required agencies to prepare a regulatory agenda that must be approved prior to proposing new rules. These reviews were performed by the Office of Information and Regulatory Affairs (OIRA), a new branch of OMB, and represented an unprecedented extension of OMB's power as well as "cradle-to-grave presidential control over the rulemaking process" (Light 1988, p. 1012).

Reagan's executive orders had the desired effect, resulting in an increased number of proposed regulations that were returned to agencies to be revised or withdrawn. For example, more than 40% of OSHA's proposed regulations were not approved by OIRA (Weidenbaum 1997). Critics complained about the deregulation biases of OIRA and about how its reviews slowed an already cumbersome rulemaking process. Many agencies resented the new control OMB wielded over their regulatory powers, but the political mood of

[3]The Environmental Protection Agency (EPA) was created in 1970, the Office of Safety and Health Administration (OSHA) in 1971, and the Consumer Products Safety Commission in 1972.

Washington was unsympathetic (McGarity 1991, pp. 271–72). Reagan's order prevailed for more than a decade until Clinton rescinded it.

Scholars were surprised when Clinton's 1993 EO12866 left in place the core of Reagan's reforms. The Clinton directive emphasizes accountability and public participation in rulemaking, and, as the centralized reviewer of agency plans, OIRA retains its role as gatekeeper for new regulation. EO12866 requires agencies to identify clearly the problem any proposed regulation addresses; to analyze a range of alternative proposals, including not regulating; and, for rules costing more than $100 million, to perform CBA for each alternative, selecting only regulations with benefits outweighing costs. It also encourages risk analysis and performance-based standards. Agencies are required to consult with local governments, avoid interagency redundancies, and consider the cumulative burdens of proposed regulations. Clinton's directive remains in effect today.

The combined impact of the regulatory reforms of executive orders has been a profound reorientation of regulation, amounting to what some scholars term a "cost-benefit" state, with OMB emerging as a powerful arbiter of regulation (Sunstein 1996). Despite Clinton's interventions, a strong antiregulatory bias still shapes OIRA reviews. As Bagley & Revesz (2006) explain, OIRA reviews CBA only to assure that costs do not exceed benefits; it rarely reviews deregulation; it does not review agencies' failure to act; it delays regulations; because the Administrative Procedures Act does not apply to OMB, courts do not review its decisions; and ORIA need not disclose justifications for its actions.

Supporters of CBA see it as a tool for improving decision making by making agencies more rational, accountable, and democratic (Hahn & Sunstein 2002, Donohue 1999, Bagley & Revesz 2006). RIA can correct cognitive biases, stimulating innovation by encouraging the development of alternative plans and simplifying complex decisions that otherwise would be difficult to grasp (Sunstein 2000). Others see it as an important means of disciplining agencies (Posner 2002). Yet CBA shares some of the same biases as sentencing guidelines. In forcing value to be expressed as price, CBA favors things for which prices are readily available and naturalized (Ackerman & Heinzerling 2002). Although a range of techniques for nonmarket valuation have been devised, they remain controversial. Because it is easier to measure short-term costs than to project the value of long-term benefits, CBA often underestimates the value of environmental regulation or discounts the future. It also precludes expressing some values at all; for example, when something is valued precisely because it is not subject to market relations or when values are expressed as incommensurability, CBA cannot express these values except by subverting them (Espeland 1998). And CBA presumes a sharp distinction between means and ends, evaluating only means; to the extent that it becomes the focus of attention, it shifts attention from the ends of regulation and does nothing to help reconcile conflict over goals or account for the moral dimensions of decisions (Heinzerling 2000).

Like sentencing guidelines, RIA was designed to curtail the discretion of agencies and did so with a combination of quantification and review. The threat of review was a potent deterrent to new regulation, and, as the centralized arbiter of regulation, OIRA clearly gained power over agencies. In forcing agencies to submit their proposals and policy agendas to OIRA for approval, agencies are less autonomous, more accountable, and more cautious regulators. Agencies have hired more economists in response to RIA, and economists' influences may diminish that of other disciplines.

RIA, in changing how power was exercised, changed resistance. Because OMB played such a key and opaque role in mediating regulation, OMB Watch, a nonprofit advocacy organization, was created in 1983 to monitor and publicize OMB practices. Its

expressed purpose is to "increase government transparency and accountability" to a "powerful and secretive agency." Because CBA is now the dominant language of regulation, objections are often framed as methodological critiques rather than disputes over values, a turn favoring those with professional expertise. But by invoking the authority of science, OMB must comply with the norms of science, which leads to its own vulnerabilities. For example, when OMB submitted proposed guidelines for a standardized method for conducting risk assessment for all agencies for peer review, the National Academy of Science soundly rejected them, saying that uniform assessments were scientifically inappropriate and that OMB had overstepped its authority (Natl. Res. Counc. 2007).

The language of accountability and transparency has been used on all sides of the debates about CBA and OMB's role in reviewing agency rulemaking. As with much quantification, numbers that appear transparent on the surface can obscure the assumptions, values, uncertainty, politics, and constraints that shaped their production. With CBA, much depends on how nonmarket values are constructed, the periodization of evaluation, or how future interest rates are calculated. For example, economists have used different methods to estimate the value of preventing lead poisoning in their children. Because lead poisoning damages children's brains, the EPA estimates the benefits of presenting this damage in terms of diminished future earnings, calculating the value as $9000 per IQ point. But another study (Lutter 2000) measures the benefits as ranging from $1100–$1900 per IQ point (these examples are from Ackerman & Heinzerling 2002). These estimates are derived from data based on parents' willingness to pay for chelation treatment for lead poisoning. Lutter argues that this estimate is superior because it is based on observable data and that the government significantly overestimates benefits of regulating lead poisoning. However, his study does not address that chelation is an unproven treatment for ameliorating lead poisoning or that the problem disproportionately affects poor families. Furthermore, neither study includes the extra cost to the state of educating children with neurological damage.

If the goals of enhancing accountability were implicit in sentencing guidelines and explicit in RIA, it is revealing to consider an example in which accountability was an unintentional consequence.

LAW SCHOOL RANKINGS

Legal education offers a unique venue for examining the effects of quantification on law.[4] In addition to providing training, socialization, and networks, law schools are gatekeepers of the profession and key arbiters of professional status. Like most bureaucracies, numbers, whether in the form of test scores, grades, tuition, budgets, or accreditation statistics, play a pivotal role in decisions made in law schools. But in the past 15 years or so, a new number has become important to law schools: *U.S. News and World Report* (USN) rankings.

USN began publishing its annual rankings of law schools in 1990 as part of its issue devoted to graduate school rankings. Initially, rankings did not generate much attention. A few deans denounced rankings, but most ignored them, believing they were too silly to take seriously. When deans realized that their applicants and alumni were taking rankings seriously, they were forced to as well. Deans' first response was to fight them. Some deans unsuccessfully tried to organize a boycott of the information USN requested. Some deans lobbied USN to abandon its rankings. Others urged USN to rank only the top programs, to measure factors differently, or to include other qualities. Eventually, in 1998 law schools, through their professional organizations, commissioned a study debunking

[4]This example draws on Sauder & Espeland (2006) and work done collaboratively with Michael Sauder.

ranking's methodology, wrote a letter sent to all students taking the law school admissions test (LSAT) to be wary of rankings, and held press conferences disparaging rankings. It was too late. Rankings found their market.

Over time, the constituencies for rankings grew as new groups discovered new uses for them. Along with harried applicants who used them to decide where to apply and attend law school, many found rankings convenient and accessible markers of standing. Employers used school rankings to evaluate job candidates; faculty use them to evaluate job opportunities; central administrators use them to compare units across the university, track progress over time, or distribute resources; law schools, alumni, and law students use them for bragging rights or to assess professional status; law review editors use them to help sort submissions; and administrators, even some who criticized the rankings, use them to market their schools or lay claim on resources.

The appeal of rankings seems straightforward: They provide useful information about complicated organizations to busy people. But this value belies the dramatic effects rankings have had on law schools. Rankings have affected many aspects of legal education, including admissions, budgets, program development, and employment. Few decisions are made without considering potential ramifications on rankings. As one law dean commented, "Rankings come up all the time. We're embarrassed to admit that but it's true. They are there lurking behind pretty much every issue" (fieldnotes, 2007). At a recent professional meeting, when a ballroom full of law professors and administrators were asked whether at their institutions most decisions included a discussion of potential ranking impacts, most raised a hand.

Status is the coin of the realm in legal education, and law schools have always been stratified in ways that defined careers. There has long been agreement about the elite schools, but the relative status of most schools was ambiguous, the stuff of partisan arguments. In formalizing and publicizing the reputations of law schools, USN changed both the terms and the stakes of status for law schools. Rankings deconstruct status as discreet components rendered commensurate. The end product is that the status of each school is expressed as a precise number on a shared metric that creates a specific relationship to every other law school. This reconfiguration of the status of law schools has encouraged law schools to think and act differently.

One way rankings have affected law schools is by providing strong incentives to change admissions policies. LSAT scores are an important factor in the selectivity component of rankings, and schools have used LSAT scores in admissions decisions for years. However, admissions directors report that they now weigh test scores more heavily, for fear of jeopardizing their rank. They offer merit scholarships to applicants with high scores, place lower-scoring students in part-time or evening programs where their scores do not count toward the school's rankings, and solicit high-performing transfer students from local law schools in their second year for the same reasons.

Rankings also affect how law schools allocate resources. In addition to spending money on merit scholarships, a big shift from prior practices, deans now spend much more money on marketing that is directed toward trying to influence the reputational surveys that USN sends out. Although most deans suspect this is a huge waste of money, they continue to send vast numbers of glossy brochures in an effort to boost their rank. To do so, they divert revenue from other areas, including need-based scholarships, hiring, or developing new programs.

Work practices inside law schools have also changed in response to rankings. In career services, what it means to be employed has shifted. Instead of tracking how students fare in obtaining jobs in law, now any job counts. Because USN does not restrict its definitions, a school that uses a more meaningful definition will be punished in the rankings.

Personnel also divert time from counseling students and contacting employers to tracking down graduates who fail to return employment surveys because USN assumes that three-fourths of nonresponders are unemployed. Because it is hard to verify employment statistics, the temptation to fudge the placement statistics is high. The suspicious numbers submitted by some schools suggest that not all have resisted this temptation. Consequently, there is much distrust and gossip surrounding these statistics.

Rankings also shape how law school communities think about themselves because organizational status envelopes its members. Students, alumni, and faculty all feel empowered or diminished by their rank, and a downward slide is often accompanied by reactions described as "hysteria" and "demoralization." Collective action and overt hostility have diminished over time as rankings have become normal features of law schools. Rankings are now so well institutionalized that legal educators no longer harbor hopes they will disappear. But they can still dream. As one dean quipped, "I wish Al Queda would make USN their next target."

Law schools experience many of the same effects of quantification that judges and regulators do. Rankings simplify complex organizations so that the whole field of American legal education is reduced to several pages. By necessity, rankings neglect important information. They rely on quantitative proxies for complex attributes: Student ability is captured by test scores and grade averages, good teaching by faculty-student ratios, reputation by survey responses, libraries by the number of books. Rankings exclude attributes that are hard to measure, such as the depth of community, commitment to public service, or accessibility of faculty.

Rankings subvert traditional forms of professional authority. As an unwelcome intrusion foisted on schools by outsiders, and despite spirited resistance, rankings powerfully control professional status. A magazine now defines the meaning and expression of excellence in legal education. Because rankings are universalistic and based on elite standards, schools with distinctive missions are penalized in rankings that do not attend to that mission. So USN influences not only the terms of excellence and evaluation but also what a particular law school is for.

Law schools are now accountable to new audiences in new ways. Rankings have succeeded in making information about law schools more accessible. Rankings circulate broadly, so distant people can easily scrutinize the performance of law schools over time and compared with each other. Deans describe how even a small downward change in rank generates enormous anxiety and demands for explanations. Applicants, law students, and alumni now feel free to challenge deans to explain their numbers, something that rarely happened before. And trustees and presidents relish having a handy disinterested measure of performance.

Like sentencing tables or cost-benefit ratios, rankings look simple enough. They seem clear, reasonable, and objective, and they make visible some aspects of an organization that outsiders might otherwise have trouble seeing. But this apparent transparency is the culmination of long, complex processes of accumulating, culling, and transforming that reflects the assumptions and practices of the institutions and people who created them. This requires enormous work and coordination among large bureaucracies, including the colleges that generate the GPAs, the Law School Admissions Council that develops and administers the LSAT, and the American Bar Association that accredits law schools. The elaborate processes behind rankings are largely invisible to those who use rankings. Few students bother to examine USN's methodology.

Rankings also illustrate the moral stakes of quantification. One reason why they remain controversial is their influence over the distribution of scarce resources: access to legal education, budgets, and employment opportunities. Even more fundamentally, rankings may also subtly subvert the missions of some

law schools or challenge broad professional values, as they punish schools who cater to disadvantaged students or schools committed to diversity or public service.

In addition to these familiar effects of quantification, there are distinctive features of the accountability that rankings impose. Most importantly, rankings are relative, so the rise of one school may affect the position of many schools. This relativity is deeply fateful for schools and helps explain why rankings generated so much controversy, change, and distrust. If one school benefits by devising a new strategy to game the rankings, any school that fails to follow suit could be jeopardizing their ranking. The prisoner's dilemma quality to rankings generates enormous pressure to scrutinize the rankings, scrutinize the schools with similar rankings, and be innovative and expansive in generating organizational statistics. It also makes collective action harder to accomplish.

The media play a crucial role by amplifying the effects of rankings, often by reinforcing their tendency to become self-fulfilling prophecies. As media products, rankings started with broad circulation and then expanded further as they became news reported by other media. The more visible rankings become, the more their legitimacy and influence grow.

Sentencing guidelines, RIA, and rankings are three locations where quantification has dramatically changed law. But it is also important to consider, if only briefly, how law sometimes precludes quantitative practices. Supreme Court rulings on the use of racial preferences in university admissions are a good example. The 1978 *Regents v. Bakke* decision struck down University of California's "inflexible quota systems" in admissions. Powell's controlling opinion held that schools could consider race in admission decisions as one of many factors in an individualized review of applicants in order to create a diverse student body but that using quotas was unconstitutional. Race and ethnicity are "suspect classes" under the Equal Protection Clause of the Fourteenth Amendment, meriting the standard of "strict scrutiny" in judicial review. The government may use these categories only in "narrowly tailored" ways to achieve a "compelling government interest." Educational diversity may be a compelling interest, and race can be considered as one of many factors in admissions, Powell concludes; however, a quota is not "narrowly tailored," does not treat applicants as individuals, is discriminatory, and violates a person's individual rights.

Powell's distinction between preferences and quotas was upheld in 2003 in two pivotal cases involving the University of Michigan's admission policies. The university assigned additional points to minority students applying to their undergraduate program. In *Gratz v. Bollinger* (2003), the Supreme Court struck down the university's point system, arguing that their policies were not "narrowly tailored." The law school, in contrast, evaluated individual applicants in order to create an unspecified "critical mass" of minority students. In *Grutter v. Bollinger* (2003), the court upheld this program as flexible enough to assure that race or ethnicity were not the defining features of admissions. These cases suggest that the explicit, mechanical use of numbers precludes treating applicants as individuals who could be compared to all applicants and violates their equal protection under the law. Even if "flexible" admissions policies generate outcomes similar to those achieved by more mechanical practices, the discretion of admissions staff, it seems, preserves the constitutionality of affirmative action.

CONCLUSION

Accountability is a term to reckon with, one that is easy to invoke and hard to oppose. Accountability can mean many things, but increasingly we link accountability to quantification. This is true in many fields, including law. We trust that numbers will be more transparent and objective than other forms of knowledge because we believe that numbers

are impersonal; that the rules for producing them are clear, shared, and constraining; that their validity can be checked by others; and that their meanings are broadly interpretable.

But quantification is rarely the neutral intervention that we might wish for, nor does it always produce disinterested knowledge. Cohen's (1984, pp. 207, 225) advice in her pivotal book on American numeracy still merits heeding: Each instance of quantification has a particular history; the meanings of systems of classification that permit quantification are relative; numerical precision and accuracy are not the same thing; quantification possesses its own distinctive biases; and "statistics were political before they were quantitative."

Because quantification disturbs power relations, because it transforms and shifts discretion, its effects can reverberate within institutions. Quantitative accountability is often appealing precisely because we wish to curtail the authority of some parties. The legitimacy of quantitative authority depends on procedures, on methodology, rather than on the discretion of particular persons or the expertise of trained professionals. But as our examples suggest, the locus and shape of discretion may change, but it does not disappear, sometimes moving to locations that make it harder to observe. Tracking discretion can be difficult, as assumptions, uncertainty, and ambiguity are buried in layers of small decisions, the traces of which are hard to recover. In such situations, responsibility becomes more diffuse and abstract than when it is exercised by known individuals or groups whose decisions must be defended publicly. It may be harder to indict procedure than persons.

Changes in authority relations are one reason why quantification is transformative. Another reason is that quantification powerfully shapes our cognition. As Hutchins (1995) shows, cognition is socially and temporally situated, and quantification can disrupt cognition systems. As a strategy for reducing and integrating information, quantification directs attention in particular ways; like a spotlight that illuminates part of the stage, it shapes what we notice and ignore. Focusing attention can improve performances. The creation of the Apgar score for measuring the health of newborns, for example, has dramatically improved mortality rates by forcing doctors to attend to babies, when previously they had focused more on the mother. But diverted attention—from goals, context, or distributive effects—can also produce poor outcomes. What counts as a relevant instance or fact is dictated by the rules that govern how and what we aggregate.

The transparency of numbers stems from their capacity to highlight processes that had formerly been obscure. Some decisions are obscure because they are made by dispersed people within bureaucracies that are hard to penetrate. Complexity is also a source of obscurity. Numbers are valuable partly because they can summarize complex information and make it accessible and easy to circulate. But simplification in one form can mask complexity by making other forms of knowledge harder to retrieve.

Quantification is designed to be transgressive. Numbers make things easy to compare by expressing difference as intervals on a shared metric. Commensuration usurps other forms of classification (Espeland & Stevens 1998). Other ways of marking difference are made irrelevant. As Strathern (1996) argues, measurement is a moral issue. With sentencing guidelines, quantification mediates the distribution and meaning of punishment, such that justice is rendered to abstract categories of persons rather than particular individuals. In regulation, quantification defines efficiency, which is a moral value. And quantification now largely defines excellence and access in American legal education.

Because of the potential for quantification to initiate dramatic changes in institutions, both positive and negative, and because it is often costly to produce, it is important to study its effects empirically. Our examples suggest that legal scholars should analyze how the authority we ascribe to numbers interacts with other forms of authority, including the

formal rational authority of law. Following Weber's example, this entails understanding authority as a relational concept. In law, for example, formalizing and interpreting rules is a different process than designing methods for quantifying, and reasoning based on cases, precedent, and intentions is different from statistical or scientific reasoning or mechanical objectivity. Understanding the varying effects of how different forms of authority impinge upon one another is an important question for legal scholars.

If quantification changes relations of authority, what people notice, and how they make sense of situations, it is not surprising that it changes how people resist. Debate often focuses on methods, on how qualified someone is to make or interpret quantitative findings, or on how instances are not covered by the rules for quantifying. Efforts to game the numbers, or interpret rules in creative ways, or create alternative rankings are ways of reclaiming some more narrow forms of discretion. But these are different strategies than making more direct claims about decisions or performance.

Which parts of law are immune to quantification or able to resist it and which are not? What makes certain legal sites, actors, and practices vulnerable to the imposition of quantitative accountability? Is this vulnerability a function of status, a particular kind of visibility, a political weakness, or a disposition to certain forms of conflict? When can claims to individualized treatment be sustained? Both sentencing guidelines and RIA were an effort to check discretion and impose accountability, but accountability was an unintended consequence of rankings. Legal scholars pay close attention to the intentions of legislators when interpreting law, but as Miller (2004) suggests, we must broaden our focus to understand the unintended effects of quantification.

Although Porter emphasizes status, conflict, and the need to communicate across cultures and distance as conditions fostering the spread of quantification, the media can also forcefully shape vulnerability to an imposed accountability. Federal judges with lifetime tenure do not seem particularly vulnerable, but because the sentences they impose are public and because crime is a highly politicized issue, Congress was able to impose guidelines. We also do not think of law school deans as particularly weak. But because rankings are media products, they were widely disseminated from the beginning, which permitted them to acquire constituencies independently of legal educators. The media can be enlisted in ways that make even influential constituencies vulnerable. And numbers, because we believe their meanings are self-evident to the numerate, can circumvent gatekeepers. Understanding how techniques for quantitative accountability accumulate powerful constituencies is another crucial question to examine.

Lastly, what are the effects of institutionalizing quantitative accountability in law? The conditions that gave rise to quantification may change as it becomes more broadly used and legitimated. If accountability becomes closely associated with quantification, it becomes routine rather than deliberative. Once quantification is understood as "the way things are done," other forms of accountability are harder to sustain and validate, which encourages a broad conformity among organizations. Sometimes this conformity may be superficial or decoupled from the real business of organizations, as a symbolic bid for legitimacy, or it may produce significant changes within organizations. Institutionalization also mediates the effects of quantification. Law schools now view rankings as a permanent if unwelcome part of the landscape, and resistance has shifted from trying to stop them to constructing more credible rankings. And although the effects of *Booker* are unfolding, it is not surprising that after two decades of experience with guidelines, many judges still use them in sentencing and reviews. Understanding how the causes and effects of technologies of accountability in law change over time and as they spread is another key dimension to examine.

Pressure for accountability will likely continue and its link to quantification grow. Efforts to collaborate internationally, to link economies and governments, and to develop international law all broaden the demand for accountability. And quantification seems a universal language. But understanding where quantification intrudes, where it is resisted, how it reconfigures power, and how context is invoked in interpreting it require nuanced research attentive to variation. We have investigated how just three examples of quantitative accountability have influenced parts of American law. Clearly, a more global and comparative perspective is needed.

DISCLOSURE STATEMENT

The authors are not aware of any biases that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

Both authors contributed equally to this review. We thank Abeer Alwan, Tom Durkin, Marwa Elshakry, Gabrielle Ferrales, Carol Heimer, Faisa Mushtaq, Michael Sauder, Peter Siegelman, Susan Silbey, Christine Stansell, and Mitchell Stevens for helpful suggestions, and Sophia Wang, Kathleen Pakuszewski, and especially Cassandra Malik for expert research assistance. The first author gratefully acknowledges the Radcliffe Institute for Advanced Study at Harvard University for its support.

LITERATURE CITED

Ackerman F, Heinzerling L. 2002. Pricing the priceless: cost-benefit analysis of environmental protection. *Univ. Penn. Law Rev.* 150(5):1553–84

Albonetti CA. 1997. Sentencing under the Federal Sentencing Guidelines: effects of defendant characteristics, guilty pleas, and departures on sentence outcomes for drug offenses, 1991–1992. *Law Soc. Rev.* 31(4):789–822

Anderson JM, Kling JR, Stith K. 1999. Measuring interjudge sentencing disparity: before and after the Federal Sentencing Guidelines. *J. Law Econ.* 42(1):271–307

Bagley N, Revesz RL. 2006. Centralized oversight of the regulatory state. *Columbia Law Rev.* 106(6):1260–329

Blackwell K, Schanzenbach M, Yaeger M, Kesler C, Gleeson J, Stith K. 2003. Panel I: Disparity in sentencing—race and gender. *Fed. Sentencing Report.* 15(3):160–64

Blumstein A, Beck AJ. 1999. Population growth in the U.S. prisons, 1980–1996. *Crime Justice* 26:17–61

Bowman FO. 1996. The quality of mercy must be restrained, and other lessons in learning to love the Federal Sentencing Guidelines. *Wis. Law Rev.* 1996(4):679–749

Bowman FO. 2005. The failure of the Federal Sentencing Guidelines: a structural analysis. *Columbia Law Rev.* 105(4):1315–50

Bowman FO, Heise M. 2001. Quiet rebellion? Explaining nearly a decade of declining federal drug sentences. *Iowa Law Rev.* 86(4):1043–136

Cohen PC. 1984. *A Calculating People: The Spread of Numeracy in Early America*. Chicago: Univ. Chicago Press

Donohue JJ. 1999. Why we should discount the views of those who discount discounting. *Yale Law J.* 108(7):1901–10

Emanuel EJ, Emanuel LL. 1996. What is accountability in health care? *Ann. Intern. Med.* 124(2):229–39

Epstein AJ. 2006. Do cardiac surgery report cards reduce mortality? Assessing the evidence. *Med. Care Res. Rev.* 63(4):403–26

Espeland WN. 1998. *The Struggle for Water: Politics, Rationality and Identity in the American Southwest*. Chicago: Univ. Chicago Press

Espeland WN, Stevens ML. 1998. Commensuration as a social process. *Annu. Rev. Sociol.* 24:313–43

Etienne M. 2004. The declining utility of the right to counsel in federal criminal courts: an empirical study on the diminished role of defense attorney advocacy under the sentencing guidelines. *Calif. Law Rev.* 92:427–85

Frankel ME. 1973. *Criminal Sentences: Law Without Order*. New York: Hill & Wang

Freed DJ. 1992. Federal sentencing in the wake of guidelines: unacceptable limits on the discretion of sentencers. *Yale Law J.* 101(8):1681–754

Fuchs EP, Anderson JE. 1987. The institutionalization of cost-benefit analysis. *Public Product. Rev.* 10(4):25–33

Gratz v. Bollinger, 539 U.S. 244 (2003)

Grutter v. Bollinger, 539 U.S. 306 (2003)

Hahn RW, Sunstein CR. 2002. A new executive order for improving federal regulation? Deeper and wider cost-benefit analysis. *Univ. Penn. Law Rev.* 150(5):1489–552

Heimer CA, Staffen LR. 1998. *For the Sake of the Children. The Social Organization of Responsibility in the Hospital and the Home.* Chicago: Univ. Chicago Press

Heinzerling L. 2000. The rights of statistical people. *Harvard Environ. Law Rev.* 24(1):189–207

Hofer PJ, Blackwell KR, Ruback RB. 1999. The effect of the Federal Sentencing Guidelines on inter-judge sentencing disparity. *J. Crim. Law Criminol.* 90(1):239–322

Hoffer TB. 2000. Accountability in education. In *Handbook of the Sociology of Education*, ed. MT Hallinan, pp. 529–43. New York: Kluwer Acad./Plenum

Hutchins E. 1995. *Cognition in the Wild*. Boston: MIT Press

Koh SY. 1992. Reestablishing the federal judge's role in sentencing. *Yale Law J.* 101(5):1109–34

Light P. 1988. Reagan's regulatory reform. *Pub. Administ. Rev.* 48(6):1012–13

Lutter R. 2000. *Valuing children's health: a reassessment of the benefits of lower lead levels*. Work. Pap. No. 00–2, AEI Brookings Joint Cent. Regul. Stud.

Margolick D. 1992. Justice by the numbers: a special report. *NY Times*, April 12. Sect. 1, Part 1, p. 1

McGarity TO. 1991. *Reinventing Rationality. The Role of Regulatory Analysis in the Federal Bureaucracy.* New York: Cambridge Univ. Press

Miller ML. 2004. Domination and dissatisfaction: prosecutors as sentencers. *Stanford Law Rev.* 56(5):1211–69

Miller P, Rose N. 1990. Governing economic life. *Econ. Soc.* 19(1):1–31

Mustard DB. 2001. Racial, ethnic, and gender disparities in sentencing: evidence from the U.S. Federal Courts. *J. Law Econ.* 44(1):285–314

Nagel IH. 1990. Foreword: structuring sentencing discretion: the new Federal Sentencing Guidelines. *J. Crim. Law Criminol.* 80(4):883–943

Narins CR, Dozier AM, Ling FS, Zareba W. 2005. The influence of public reporting of outcome data on medical decision making by physicians. *Arch. Internal Med.* 165(1):83–87

Natl. Comm. Excellence Educ. 1983. *A Nation at Risk: The Imperative for Educational Reform.* Washington, DC: USGPO

Natl. Res. Counc. Comm. Rev. OMB Risk Assess. Bull. 2007. *Scientific Review of the Proposed Risk Assessment Bulletin from the Office of Management and Budget, Executive Summary.* Washington, DC: Natl. Acad. Press

Ohmann R. 2000. Historical reflections on accountability. *Academe* 86(1):24–29
Porter TM. 1995. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton Univ. Press
Posner EA. 2002. Using net benefit accounts to discipline agencies: a thought experiment. *Univ. Penn. Law Rev.* 150(5):1473–88
Power M. 1997. *The Audit Society: Rituals of Verification*. Oxford: Oxford Univ. Press
Power M. 2003. Evaluating the audit explosion. *Law Policy* 25(3):185–202
Ranson S. 2003. Public accountability in the age of neo-liberal governance. *J. Educ. Policy* 18(5):459–80
Regents of the University of California v. Bakke, 438 U.S. 265 (1978)
Rose N, O'Malley P, Valverde M. 2006. Governmentality. *Annu. Rev. Law Soc. Sci.* 2:83–104
Rosga A, Satterthwaite M. 2003. *Trust in indicators: trying to measure human rights.* Presented at Annu. Law Soc. Assoc. Meet., Pittsburgh, June 5–8
Sauder M, Espeland WN. 2006. Strength in numbers? The advantages of multiple rankings. *Indiana Law J.* 81:205–27
Shore C, Wright S. 1999. Audit culture and anthropology: neo-liberalism in British higher education. *J. R. Anthropol. Inst.* 5(4):557–75
Shore C, Wright S. 2000. Coercive accountability: the rise of audit culture in higher education. See Strathern 2000, pp. 57–89
Silbey S. 2003. *Governing green laboratories: differential responses to legal regulation.* Presented at Annu. Law Soc. Assoc. Meet., Pittsburgh, June 5–8
Stith K, Cabranes JA. 1998. *Fear of Judging: Sentencing Guidelines in the Federal Courts*. Chicago: Univ. Chicago Press
Strathern M. 1996. From improvement to enhancement: an anthropological comment on the audit culture. *Cambridge Anthropol.* 19(3):1–21
Strathern M, ed. 2000. *Audit Cultures. Anthropological Studies in Accountability, Ethics and the Academy.* London: Routledge
Sunstein CR. 1996. Congress, constitutional moments, and the cost-benefit state. *Stanford Law Rev.* 48(2):247–309
Sunstein CR. 2000. Cognition and cost-benefit analysis. *J. Legal Stud.* 29(2):1059–103
Tonry M. 1996. *Sentencing Matters*. Oxford: Oxford Univ. Press
Tonry M. 2005. The functions of sentencing and sentencing reform. *Stanford Law Rev.* 58(1):37–66
Trott S. 1995. Letter to Richard Conaboy, Chairman of the U.S. Sentencing Commission, November 9, 1994. *Fed. Sentencing Report.* 8(3):199
U.S. Sentencing Commission. 2006. *Guidelines Manual*, §3E1.1 (Nov.). Washington, DC: U.S. Sentencing Commission. **http://www.ussc.gov/2006guid/gl2006.pdf**
U.S. Sentencing Commission. 2004. *Fifteen Years of Guidelines Sentencing: An Assessment of How Well the Federal Criminal Justice System Is Achieving the Goals of Sentencing Reform*. Washington, DC: U.S. Sentencing Commission. **http://www.ussc.gov/15_year/15_year_study_full.pdf**
United States v. Booker, 543 U.S. 220 (2005)
Von Dornum DD. 1997. The straight and the crooked: legal accountability in ancient Greece. *Columbia Law Rev.* 97(5):1483–518
Weidenbaum M. 1997. Regulatory process reform: from Ford to Clinton. *Regulation* 20(1):20–26
Williams R. 1976. *Keywords: A Vocabulary of Culture and Society*. New York: Oxford Univ. Press

Annual Review of Law and Social Science

Volume 3, 2007

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Law and Social Science* articles may be found at http://lawsocsci.annualreviews.org