1. Introduction

When John Adams crafted wording that the state shall establish "a government of laws and not of men" into the constitution of the Commonwealth of Massachusetts in 1780, he did not specify precisely where, in the real world, the boundaries of law should end and where officials should begin to exercise discretion in interpreting and enforcing the laws. The founding fathers gave us organizing principles and ideals, not detailed blueprints for governance and bright lines for policy. For the more than two centuries to follow, reasonable people have disagreed often and loudly over those boundaries. This has been a recurring source of tension in matters of civil and administrative justice, and in the criminal and juvenile justice systems.

Adams's dictum was really more a statement of protest against tyranny and a linchpin of his deeply held commitment to republicanism than it was a guide for governance. He revealed, day-in and day-out, in his life of service as legislator and then President, that under the rule of law officials routinely exercise discretion when they make decisions: they interpret the laws so that a prudent course of action can be taken in each case, given the particulars of the case and the effects, both direct and indirect, of alternative decisions on the people involved. By "prudent" we mean effective, evenhanded, and mindful of the social costs of the alternatives.

One cannot understand our system of justice without understanding the central role that discretion plays in its operation. Discretion is frequently discussed, often mischaracterized, and occasionally analyzed, but rarely defined. We propose the following definition, adapted from Shavell (2007): Discretion is the latitude granted officials to act under a formal set of rules and in a public capacity. It includes the decision not to act -- for the police not to arrest or for the prosecutor not to file charges or to drop charges filed earlier. These decisions are shaped largely...
by resource constraints and commonsense interests of justice. Discretion is sometimes described as "unbridled" or "unfettered"; while such characterizations apply accurately to earlier periods of justice in America, they are less applicable today. Discretion is now more closely restricted than in earlier times by formal and informal systems of accountability and by higher levels of political transparency, largely technologically induced.

Discretionary decisions are essential to move cases either forward or out of the system. They are often based on information not delineated in rules, information that in principle should be related to the mission of the agency for which the actions are taken or not taken, but that in practice may be unrelated. Decisions made in the name of discretion may even run counter to the mission of the agency. Such decisions can be especially harmful to the community when they routinely go unchecked.

In a well-functioning system, discretionary decisions will relate closely to the goals of the parent agency, and achievement of those goals will serve the public interest. Full evaluations of these decisions should take into account the levers available to the official, as well as the possibly conflicting interests of that official. They should also account for the costs and benefits to the clients of the system -- defendants, victims, witnesses, and the community at large -- as well as the costs and benefits to other officials in the system. In the best case scenario, we can study the motives of the actors directly, but in most cases we must draw inferences about the motives of the actors through the study of the outcomes that emanate from those decisions. Accordingly, empirical studies of discretion need to be mindful of the goals that the actors aim to achieve, the restrictions under which the decisions are made, and the costs and benefits of the decisions to all involved.

Criminological research on discretion often focuses on unwarranted disparity in sentencing outcomes for convicted offenders, ignoring the decisions that gave rise to those outcomes and the goals and boundaries that shape them. While useful as a starting point -- unwarranted disparity in sentencing is, after all, a real and serious problem -- such analyses do little to advance theories of behavior or to provide generalizable conclusions into how discretion is used or how it might be managed differently to produce outcomes that are more evenhanded, consistent with the goals of justice, and less costly.

In this paper, we explore the prospect that new and potentially useful insights into the performance of the system may be generated by analyzing the system as a set of interacting decisions with costs and benefits, both explicit and implicit. We go about this by clarifying the essentials and re-framing the discussion on the exercise of discretion, drawing from well-established principles and concepts that have proven useful in non-justice settings. In the next section we review the work of criminologists and others on the question of discretion. We then identify the key decision nodes at which discretion is exercised, and we look to methods that can help identify the discretion of individual agents. Having empirically identified individual discretion, we can then attempt to distinguish between abuse and "responsible use". Next, we discuss the exercise of discretion at the policy-setting level We then consider theoretical pathways to advance our understanding of the interaction of rules and discretion -- including
rational choice and the use of heuristics in decision making; the prospect that sentences may reduce the social costs of crime but impose other costs that can be much more than offsetting; and the idea of discretion as a source of both random and systematic errors of justice. We then bring evidence to bear to demonstrate the viability of testing aspects of these ideas empirically. We conclude by setting forth an agenda for further research.

2. Thinking About Discretion

Survey Research on Discretion. Criminologists and legal theorists have been thinking about discretion for decades, if not longer. Suspicious of widespread criminal justice failures and aware of how little was known about critical decisions made in the system in the 1920s, legal scholars undertook a series of surveys: in Cleveland in 1922, Missouri in 1926, and Illinois in 1929, culminating in the Wickersham Commission survey and analysis (Bettman, 1931; Ohlin, 1993). The primary goal of these surveys was to shed light on incompetence and corruption in case processing; one of the products created to serve this end was a set of "mortality tables", indicating rates of case dropout at each major stage of case processing. The surveys also revealed that police, prosecutors, and judges exercise considerable discretion in determining what happens to cases after the police first become aware of them as criminal incidents.

Three decades later, the American Bar Foundation (1957) conducted a survey aimed at providing a more systemic view of the criminal justice system, focusing on the decisions made by police, prosecutors and judges in daily practice (Ohlin, 1993). The ABF survey revealed that practitioners prefer discretion and flexibility to rules; that discretion is usually exercised most extensively at the lowest levels of the organization (especially in policing and prosecution) and with limited transparency. Frank Remington reported in 1993 that the problem of lack of transparency was particularly acute among prosecutors, who feared that casting light on their exercise of discretion would lead to criticism and the imposition of formal rules that would restrict their ability to protect the community.

One of the central findings of the study was that the criminal justice system is complex and interconnected, that attempts to control decisions at one stage affect decisions made by agents at other stages of the process, in a manner akin to a hydraulic system. Another was that most discretionary decisions are made by agents on the front line -- especially beat cops and screening room prosecutors -- rather than by higher-ups in the agency. Yet another was that these decisions are based largely on practical considerations -- especially, competing demands for service and other pressures of the moment, and resource constraints -- rather than rigid adherence to the law. The work of the ABF study culminated in a series of five books edited by

4 Bettman's 1931 analysis of the Wickersham Commission survey data focused on the following topics: criminal career case histories; adolescent experiences and crime; warrants and arrests; juveniles in the criminal justice system (CJS); municipal court systems; plea bargaining; grand juries; waiver of jury; trial by jury; the insanity defense; trial conduct; sentencing; probation; public defenders; bail; CJS information; and CJS organization.
Frank Remington, one of the principals of the original study. The books, published from 1965 through 1969, focused on five areas: investigation (Tiffany et al., 1967), arrest (LaFave, 1965), prosecution (Miller, 1969), conviction (Newman, 1966), and sentencing (Dawson, 1969).

**Discretion as a Problem.** Robert Dawson's assessment of discretion in sentencing is especially germane to our inquiry. He opened his analysis by observing that a full assessment requires attention to discretion exercised not only by judges, but by police, prosecutors, and correctional officials (especially the probation officers who conduct and document the presentence investigations that provide a primary basis for both the prosecutor's recommendation on sentence and the judge's decision, and parole boards; Dawson, 1969, pp. 215, 220). Dawson went on to observe that unwarranted disparity in the exercise of discretion at each stage imposes a variety of harms: it undermines public confidence in the administration of justice; demoralizes prisoners and disrupts attempts to rehabilitate them; and casts grave doubt on the ability of the criminal justice system to reduce crime and achieve the goals of justice generally (p. 216).

The ABF survey findings were followed by those of the 1967 President's Crime Commission, which found that informal and often opaque discretionary actions taken by police, prosecutors and judges produce unwarranted disparity for similar cases. This finding was more fully developed by Kenneth Culp Davis in his landmark 1969 book, *Discretionary Justice: A Preliminary Inquiry*. Davis argued for more control over the exercise of discretion.

In 1993, criminologist Samuel Walker elaborated on Davis's call for greater control of discretion in the criminal justice system. He opened his analysis by describing the ABF survey and the crime commissions, noting that they failed to see factors associated with what he called the "sociology of criminal justice": the impact of workloads, concerns about bureaucratic self-interest and the need to maintain good relations with other institutions, and use of the law for purposes that were unrelated to the need to punish criminal violations. Walker saw discretion largely as unbridled, less as a tool for prudent justice and more of a "national problem" in need of taming, especially discretion exercised by the police (pp. 10-11, 146).

A related concern often raised in the literature on discretion and sentencing is that -- consistent with the hydraulic model -- attempts to reduce judicial discretion run the risk of shifting sentencing authority from the judge to the prosecutor, and that this is unwise and unfair, especially a threat to the rights of the innocent (see, for example, Remington (1969, p. 89; Alschuler, 1978, p. 551; Tonry, 1996, p. 135). The prospect of displacing discretionary power from the agent charged with fair adjudication to the agent representing the state against the defendant should, indeed, raise concerns among fair-minded people. While tougher sentencing laws have unquestionably reduced the discretionary authority of judges and increased sentence terms, reviews of the empirical evidence provide little support for the proposition that changes in sentencing laws have induced prosecutors to exercise more discretionary power than before (Engen, 2008). The evidence appears to provide more support to the role of prosecutor as fair

---

5 Remington was Director of Field Research of the ABF Pilot Project and Staff Director of the ABF's Survey of the Administration of Criminal Justice in the United States.
official of the court than as a pursuer of maximum punishments on defendants. With regard to increases in mandatory minimums, prosecutors are more inclined to circumvent overly harsh sentencing laws by reducing charges and dropping cases (Bynum, 1982; Farrell, 2003; Ulmer, et al., 2003). Engen concludes, "(T)he more severe the increase in sentence that would result from the mandatory minimum, the less likely prosecutors were to use it." (2008, p. 77) With regard to voluntary sentencing guidelines, Engen finds that judges maintain more control over the imprisonment decision than do prosecutors (p. 82). Engen concludes that research on discretion should focus less on the displacement hypothesis and more on the broader question of the relationship between sentencing laws and discretion (p. 86).

However, the evidence used by Engen to support his conclusion is admittedly fairly weak. The problem is that it is difficult empirically to identify the relative discretion of all actors, particularly prosecutors, whose behavior is not as easily measurable as that of judges. The advantage of the mandatory minimum laws is that we know what the prosecutor is supposed to do, and we can then compare that to what they actually do. But the fact that prosecutors do not always do what the legislators want them to do does not mean they have less discretion -- the threat of the mandatory minimum might allow them to coerce guilty pleas in cases where the defendant would have otherwise taken his chance at trial. In fact, one could argue that the fact that the mandatory minimums are applied selectively means that the prosecutors have more discretion, not less, as a result of the mandatory minimum. Even if the average sentence does not change, prosecutors could be redistributing the sentences, based on who is most susceptible to the threat.

Anne Piehl and Shawn Bushway (2007) deal with this empirical challenge with respect to the strictness of the guidelines in Washington and Maryland. They define prosecutorial discretion as the distance traveled between the potential sentence at indictment and the actual sentence as a result of the plea. After using a decomposition method to control for different case characteristics in Washington and Maryland, they show that there is more discretion in the Washington system then in the Maryland system. Moreover, the variation at this stage from case to case is more likely to be correlated with race and gender, which at least raises the possibility that the strict guidelines in Washington displace discretion to prosecutors, who then use extra legal factors in their decision making.

Clearly, the hydraulic model and the associated displacement hypothesis have some validity: changes at one stage can indeed influence operations at adjacent stages up and down the line. Common statements suggest, however, that this displacement of discretionary power is automatic and inevitable. The hydraulic model is in fact a crude metaphor for the adaptations

---

6 For example Rosett and Cressey (1976) write, "The exercise of discretion is a response to a fundamental need for flexibility ... Like a closed tube of toothpaste, if squeezed in one spot it merely pops out in another." (p. 170, emphasis added) Similarly, Remington writes, "... the ABF research findings contributed to an increased understanding that an effort to eliminate or reduce discretion at one stage ... will create a risk that discretion will merely shift to another stage where
that actually occur throughout the system when changes are made at any particular stage. The model ignores the unique decision processes at work and uniqueness of setting. The adaptations will play out differently depending on the nature of the initial change in policy, the response options available to agents at other stages, workloads and available resources, the incentives of the agents at each stage, unique characteristics of the jurisdiction, and other factors. The challenge, empirically, is to identify the actions of the different stages using available data, and be clear about what it means to have discretion at any given stage. We will address the empirical challenge in a number of sections below.

Legal theorists have generally taken a less systemic view of discretion, focusing largely on the exercise of judicial discretion. Perhaps the most authoritative legal-theoretic account of discretion has been advanced by Ronald Dworkin (1977), who begins by distinguishing discretion from ordinary personal decision making, which is not accountable to a set of standards or a particular higher authority (p. 31). Dworkin goes on to distinguish further between "weak" and "strong" discretion (pp. 31-32). Weak discretion refers to the need for judgment in the presence of vague rules; for judges, umpires, and other interpreters of rules, it is a fairly mechanical and uncontroversial business that comes with the job. Strong discretion is discretion that is neither mechanical nor trivial; it involves judgments supported by principles and invoked under a duty that is larger than that ordinarily associated with the problem at hand, involving an "ultimate social rule or set of social rules" (p. 69).

Hole in the Literature on Discretion: Social Cost. Discretion has a bad name in many circles, much of it well-deserved. It is associated with a style of justice administration that is arbitrary, capricious, and too often discriminatory. It lacks transparency. It often produces variation in outcomes that lack any rational justification and are, therefore, unwarranted. Discretion is, for some, a mostly negative feature of the criminal justice system, and it needs "taming" (Davis, 1969; Walker, 1993). Discretion has in fact been effectively curtailed in policing (as in the case of policies and rulings (especially, Tennessee v. Garner, 1985) that have ended the practice of shooting at harmless fleeing felons and that use arrest as the preferred or mandatory policy in domestic violence cases) and in sentencing (with the reduction in unwarranted disparity under sentencing guidelines), and it could be effectively reduced elsewhere (Walker, 1993).

Other observers have been more cautious, seeing discretion as an essential lubricant of our system of justice, needed to deal rationally with important intricacies that cannot be covered by laws or guidelines, nuances that inevitably show up in most real-world criminal justice cases (Gottfredson and Gottfredson, 1988). This lubricant is often needed to counter tendencies to either systematically overuse criminal justice interventions, as in the case of draconian mandatory sentencing laws, or to underuse them, as in the inclination of law enforcement officials to fail to collect and process relevant and available evidence to solve crimes and reduce instances of wrongful arrest and conviction.

its exercise is less visible." (p. 110, emphasis added)
Still others see discretion as much more complex: sequential, multi-layered, and involving a mix of systemic and individual case influences (Reitz, 1998; Hawkins, 2001). Bjerk (2008) has added to this discussion by observing that negotiated pleas will fall considerably short of expected sentences in trial in a world of offenders who tend to be risk preferring and prosecutors who tend to be risk averse. These observations on discretion suggest a degree of complexity that does not lend itself readily to the sort of linear, static, deterministic models that are common to the analysis of public policy. They suggest models that simulate the key decision points of the criminal justice system at which discretion is most consequentially exercised.

Missing from conventional discussions of the exercise of discretion is consideration of social costs -- in particular, the effect of policies and individual decisions by the police and prosecutors on the social costs of failures to convict culpable offenders on the one hand and over-incarceration of others, some of whom are innocent, on the other. The reality is that any system of justice that we can imagine will have these types of errors even when it is functioning perfectly. It is fashionable to label all errors, particularly conviction of innocents, as "miscarriages of justice." Although some of these mistakes are clearly the result of malfeasance of agents, some if not most of them are the natural result of a system of justice that relies on probabilistic rules of conviction, as in a hypothesis test. Discussions about the system, and the discretion in the system, must be able explicitly to weight the relative costs of these different errors, often categorized as false positives (arrests and convictions of the innocent, incarcerating non-risky people), and false negatives (failing to convict culpable offenders, not incarcerating dangerous convicts). In any given system, these two types of errors are often hydraulically linked -- the way to decrease one is usually to accept an increase in the other. Although we continually search for the new technology and methods that will allow us to decrease both errors simultaneously (DNA testing, new methods of interrogation), we need to recognize explicitly that errors will exist in any system.

Part of this consideration involves explicit assessments of the costs of the different errors, both tangible and intangible. In the heated discussions about selective incapacitation (Monahan, 1978; Gottfredson and Gottfredson, 1994), scholars rightfully pointed out that the "rubber hit the road" when it came down to making the cutpoint that would establish some as dangerous and others as not. Although prediction often involves discussion of model fit, evaluation of these models ultimately rests on a comparison of error rates and costs. In this discussion, academics have ignored the respective costs of these errors and decided that these prediction exercises are too difficult (Gottfredson and Gottfredson, 1994; Auerhahn, 1999; Gottfredson and Moriarity 2006). But as Richard Berk (2009) makes clear, policy makers are comfortable with the notion that not all prediction errors are equal. He argues that cost functions in prediction exercises should recognize that false positives and false negatives are not equally costly. Although rarely done, it is relatively straightforward to implement non-equivalent cost functions in empirical prediction models.

This might be a very clear example of empirical researchers lagging behind policymakers. In most discussions of decisions about guilt and innocence, policymakers and
legal scholars routinely accept that the ratio of costs is not 1. Consider Blackstone's (1765) famous quote that it is "better that ten guilty persons escape than that one innocent suffer." Volokh's (1997) entertaining review of the relative weights assigned in state statute's reveals that at least in the laws of the United States, this ratio varies from 1 to 1, to 100 to 1 (in Oklahoma).

But we know of no empirical attempt to estimate what the ratios are in practice. Simple back-of-the-envelope calculations suggest that the actual ratio is well above even 100 to 1. Consider that well over 10 million felony offenses involving adult offenders occur annually that do not end in conviction. Estimates of wrongful convictions, by contrast, have been estimated at between 0.5 and one percent of all felony convictions, which amounts to some 5,000 to 10,000 annually (Huff et al., 1996; Forst, 2004). This gives a ratio of 1,500 to 3,000 false negatives to every innocent person convicted. Clearly, the social costs of crimes matter too. Therefore, we advocate for more simple descriptive research that identifies the actual ratio that exists in conviction and incarceration decisions, along with empirical research on the relative costs of these decisions. We also advocate against rhetoric and research that focuses only on wrongful conviction without consideration of the full range of other errors and their costs. Any system will make errors. Intelligent policymaking starts from an acceptance of this fact, but it also requires that policymakers have information about what is actually happening in the system, and the relative costs of these outcomes.

Another hole in the literature involves information about cases dropped by the prosecutor, information that is critical to an understanding of prosecutorial discretion. Since information about these cases -- and about all other case outcomes -- is not readily available to the public or to scholars, the prosecutor is insulated from systems of measured public accountability in most states (Forst, 2010). And the media, capable of enlightening the public on the folly of weak accountability for the vast majority of felony cases, have opted instead to feed the public's appetite for information about sensational cases. The exercise of discretion by prosecutors is thus mostly out of public view.

We know of only one national dataset that has readily available data on dismissed cases, the State Courts Processing Statistics (SCPS), and even that dataset starts with indicted cases, thereby missing cases that are rejected for prosecution or dismissed after arrest ("nolle prosequi"). The Bureau of Justice Statistics reports that 23% of all felony defendants in the 75 largest counties in the US have their cases rejected or "nolled" (Cohen and Kyckelhahn, 2010).

What little we do know suggests that prosecutors, like other public servants, have varying incentives, and this variation manifests as more transparency in some offices than in others, much lower ratios of pleas to trials in some metropolitan areas than in others, and opportunities for substantial disparity and inefficiencies in the exercise of discretion. The goals of prosecution have not been made sufficiently clear, and detailed information about the decisions made by prosecutors is not sufficiently accessible to allow anyone to know whether prosecutors tend to make decisions about individual cases that correspond closely or consistently to any particular standard of justice or efficiency.
Rules, Discretion and Social Costs. Both rules and discretion can increase social costs, as we have noted, and these costs are not evenly distributed across the population. Rules that are too tough concentrate costs on offenders and their families without sufficient compensation in crime reduction. Rules that are too lenient concentrate costs on victims, without sufficient compensation to others. The exercise of discretion can produce both types of social costs: with some defendants getting treated too harshly, others too leniently, and society at large questioning whether a system with unwarranted disparities is just and legitimate, worthy of support. The goal of sentencing policy should be to find a mix of rules and discretion that minimizes the total of these costs to society.

Robert Dawson's American Bar Foundation report on sentencing opens with a statement by Geoffrey C. Hazard, Jr., executive director of the ABF, which lays bare an insensitivity to the importance of social costs:

The verdict of contemporary common consent is in accord; the only question these days seems to be how severe we must become in order that the criminal law will work. Upon this premise, the problems of criminal law enforcement can be reduced to two questions of efficiency: How to get as many offenders as possible through the procedural mill to the point of sentencing; and, at the point of sentencing, how to distribute the short supply of punishments so that the maximum rectifying effects will be achieved ... It is inappropriate here to explore whether the premise is well founded. (Dawson, 1969, p. x)

Hazard's emphasis on efficiency is stunning for its lack of regard for either systematic errors created by misguided rules of criminal justice intervention and sanctioning or random errors associated with disparity in the exercise of discretion. Both types of error impose costs on society that may be profound (Polinsky and Shavell, 2000; Forst, 2004). Systematic errors are of two basic types: those that result in failure to bring culpable offenders to justice, which have been called "errors of impunity" (Bohm, 2005; Forst, 2004), and those that impose excessive costs either on innocent people wrongfully suspected or convicted of crime or on offenders who received harsher sanctions than warranted by basic considerations of justice and efficiency, if not by an ideal "optimal" sanction, one that minimizes the total social costs of crime and justice. Random errors of justice are those that result from a criminal justice agent's acting on an individual case in a manner that departs from the norm for that offense-offender category (if not from the optimal sentence) which makes for unwarranted disparities in sanctions for offenders charged with similar crimes and with similar criminal histories. These errors may be the product of a suspect's case getting handled by an agent who manifests either a lapse in competence or excessive toughness or leniency for the matter at hand.

A review of sentencing practices and policies over the past 50 years reveals strong ebbs and flows in the mix of rules and discretion in sentencing. This history is well known in the criminal justice community, but it is worth retelling in terms of systematic and random errors of justice. Mounting evidence of unwarranted disparity in sentencing -- via a mix of random and systematic departures from sentencing norms -- contributed to a marked shift away from
indeterminate sentencing systems toward systems of sentencing guidelines and mandatory sentences throughout the 1980s and '90s. Indeterminate sentencing involved the giving of a broad range of prison time after a finding or plea of guilt, followed by a parole board's using the rehabilitation model to determine when the offender had been "corrected" and was ready to be returned to the community. Other sources of variation in judges' sentences were found to be associated with their espousal of different sentencing goals: retribution (or its kinder sibling, "just desert"), general deterrence, special deterrence (the "stick" counterpart to the "carrot" of rehabilitation), and incapacitation (Forst and Wellford, 1981; National Academy of Sciences, 1983). The freedom of judges to choose among these sentencing goals has been found to contribute to unwarranted variation in the exercise of sentencing discretion, along with tendencies for some judges to be more predictable in their sentencing practices than others, regardless of their preference for one goal or another (Forst and Wellford, 1981). Others have found the exercise of judicial discretion to lean toward harsher sentences for federal judges appointed by Republican presidents and toward more lenient sentences for federal judges appointed by Democratic presidents (Schanzenbach and Tiller, 2007).

At the vanguard of the movement to reduce unwarranted variation in sentencing was the development of a system of sentencing guidelines for the state of Minnesota, implemented in 1980. This system aimed to reduce both systematic and random errors in sentencing -- the former by reducing the average sentence for each offense-offender category, and the latter by narrowing the range of sentences in each category and ending the parole system, which had introduced unwarranted variation in terms served, in the name of determining whether the offender had been rehabilitated (Frase, 2005; Knapp, 1982).

The Minnesota sentencing guidelines became the dominant model for the creation of guidelines not only in other states, but for the federal system as well. In 1984, Congress passed the Sentencing Reform Act (as part of the larger Comprehensive Crime Control Act), which authorized a federal commission to design a system of mandatory guidelines and end the federal system of indeterminate sentencing with parole. The guidelines became effective in November 1987 and continued with little change for nearly 20 years. Unlike the Minnesota guidelines, which initially reduced average sentences for most categories of offense and offender, the federal guidelines generally increased both the rate of incarceration and the term of sentence for most categories. The federal guidelines may thus have reduced the social costs of errors associated with unwarranted disparity in the exercise of discretion, but appears to have produced more than offsetting increases in the social costs associated with overly severe punishments, especially for nonviolent offenders.

The federal guidelines went from mandatory to advisory under the Supreme Court's 2005 ruling in United States v. Booker (543 U.S. 220 [2005]), which held that sentences imposed under guidelines were based on data that did not pass the evidentiary test needed to convict in trial, thus violated the Sixth Amendment right to trial by jury.

In time, the Minnesota guidelines became tougher, thanks largely to legislation fueled by the increased politicization of crime and punishment in the 1980s and '90s.
Unwarranted disparity in the exercise of discretion has been found even under regimes of sentencing guidelines. These are, after all, guidelines: they allow for some exercise of judicial discretion, including departures from guidelines. And even within a narrower range of discretion under guidelines, evidence of disparity has been uncovered. Bushway and Piehl (2001), for example, found that under a system of sentencing guidelines in Maryland, judges tended to give longer sentences than recommended by the guidelines to defendants at the severe end of the guidelines grid, and these defendants were disproportionately African Americans.

Still, the change from unfettered discretion in sentencing to mandatory guidelines in many jurisdictions during this period was as sharp a shift in the balance of discretion and rules as we have seen in the history of the United States system of justice, a sea change. And as sea tides ebb and wane, so too has this balance begun to move back in the direction of discretion. Three landmark Supreme Court decisions did much to return sentencing discretion to judges: *Apprendi v. New Jersey* (530 U.S. 466 [2000]), *Blakely v. Washington* (542 U.S. 296 [2004]), and *United States v. Booker* (543 U.S. 219 [2005]) -- largely by eliminating the practice of basing sentences on information about the offense and offender that had not been subjected to courtroom scrutiny. Prior to these decisions, defendants could be sentenced under guidelines to crimes for which they had been acquitted in court.

It remains to be seen whether sentencing legislation will become more lenient and whether the boundaries of judicial discretion in sentencing will continue to widen over the coming years. The latter seems more likely, at least to the extent that legislators who vote for greater leniency are more vulnerable to political attacks from the right than legislators who vote to widen the boundaries of discretion.

Departures from optimal sentences are influenced also by the exercise of prosecutorial discretion. Mischief in the exercise of prosecutorial discretion is an obvious source of such departures, as we discuss more fully in the section to follow on abuses of discretion. A potentially stronger source of departure from optimal sentences receives little attention: whether the prosecutor should aim for more convictions or for stronger convictions. Conventional wisdom holds that plea-to-trial ratios are driven by caseload pressures, but jurisdictions with huge caseloads often have low ratios of pleas to trials (as few as four pleas per trial), while those with much smaller caseloads often have high ratios (as many as twenty or more pleas per trial).

---

9 In *Apprendi*, the Court ruled, five to four, that the Sixth Amendment right to a jury trial prohibited judges from enhancing criminal sentences beyond statutory maximums based on facts other than those decided by the jury beyond a reasonable doubt.

10 In *Blakely*, the Court ruled, five to four, that the Sixth Amendment right to a jury trial prohibited judges from enhancing criminal sentences based on facts other than those decided by the jury or confessed to by the defendant.

11 In *Booker*, the Court ruled, six to three, that the federal sentencing guidelines violate the Sixth Amendment right to a jury trial by requiring judges to determine facts for criminal sentencing, and so should be made advisory rather than mandatory.
Prosecution policy -- whether to aim for quality convictions by being more selective at the screening stage and then putting more effort into bringing cases to trial, or to aim for more convictions by accepting more marginal cases and putting greater effort into negotiating guilty pleas in return for charge reductions -- appears to be no less important a determinant of the ratio of pleas to trials than the press of large caseloads. This policy decision can also strongly influence the number of failures to convict culpable offenders and associated tendencies to over-incarcerate some offenders, under-incarcerate others, and wrongfully convict innocent people (Forst, 2004).

Rethinking Discretion. One of the more incisive and useful observations on the exercise of discretion -- and one that helps to motivate the current inquiry -- is that by Frank Remington (1969, p. xix): "The importance of the exercise of discretionary power is so obvious that it would not warrant emphasis were it not for the fact that it has commonly been ignored." This is surely less true today than in 1969, but there remains considerable opportunity for the discovery of new insights about the exercise of discretion in the criminal justice system.

Which critical aspects of discretion are still commonly ignored? What is missing from the collective assessments that have been made to date on the exercise of discretion in the criminal justice system? We think that a useful way to begin addressing these questions is to examine how the basic problem of discretionary decision making is common to other domains, and how it has been analyzed and managed in those domains. We are struck by the extent to which the criminal justice and criminology literature tends to ignore the influence of major developments in society at large on the criminal justice system -- community-centered solutions to service, new systems of accountability, notions of transparency and legitimacy, and the replacement of hierarchies with networks are prominent examples. Discretion is exercised extensively outside the criminal justice system, and frameworks of analysis provide principles and tools that have been used productively to assess and manage discretion in those domains. They can help to inform the current inquiry as well. In other settings, the exercise of discretion is generally a by-product of a larger inquiry that assesses both the rules that shape discretion and the exercise of discretion within that set of rules. In the criminal justice setting, as in other settings, discretion can be useful not only when the rules are effective and just yet in need of interpretation, but also as an antidote to rules that violate fundamental notions of effectiveness and fairness.

We can begin to better understand and resolve tensions between rules or laws and the exercise of discretion, by introducing the idea of optimal sentences, regarding departures from those sentences as errors, and using principles of error management generally to inform the exercise of discretion (Bibas et al., 2009). The idea of misguided or misplaced rules is a relative one -- we must first have a sense of what "should" happen under the "preferred" system. Here are some examples of problem domains that call for a similar balancing of rules and discretion to deal with conditions of complexity and uncertainty in matters that can impose significant harms on individuals: quality control management (Evans and Lindsay, 2004), signal detection applications in clinical diagnosis and treatment (Akay, 1996), control of monetary policy (McCallum, 2007), and financial portfolio management (Fabozzi and Drake, 2010).
A central idea of this essay is that criminal justice sanctions and interventions can create systematic departures from optimal sanction levels, akin to statistical bias, and failures to effectively manage the exercise of discretion can create both systematic and random departures, with potentially grave consequences in terms of both direct and indirect social costs. While the idea of an optimal sanction for any particular offender is more a useful conceptual tool than it is an entity that can be precisely calculated to everyone's satisfaction, the setting of sanctions and rules for the practice of justice may nonetheless benefit from an awareness of the ways that systematic and random departures from optimal sentences and the associated costs are managed in other settings. One might reasonably expect that corresponding tools can be crafted to establish rules and boundaries and manage the exercise of discretion in the criminal justice system in such a way as to reduce the social costs associated with gross departures, both systematic and random, from optimal sanction levels.

From an empirical point of view, we cannot have a discussion about "misplaced" or "misguided" discretion until the "optimal" or "normative" discretion is first identified. All too often in current research, the "optimal" or "expected" is not defined, which makes it difficult to understand the observed effect. Alternatively, the "null" is defined naively as no errors or according to some poorly defined "legalistic" model. Before we can identify deviations, we need to know what "should" happen.

The rational choice model could be a useful "default" or null model. Observed behavior could be compared against what we would expect based on a rational model, and then observed deviations are explained using psychological models of behavior or other systematic observations of non rationality, like discrimination.

A prominent example of this type of work in criminology involves Alfred Blumstein's simple thought experiment involving racial discrimination in incarceration. Typically researchers try to control for all observable variables that could explain the incarceration sentence, and then include race as an outcome. If race is significant, there is evidence of extra legal variation in sentencing. But, this is true only if the model is fully specified, clearly not a reasonable assumption. Blumstein proposed a simple alternative model, in which he imagined a world in which people were processed from arrest to conviction via a simple function where a constant proportion of people make it from arrest to incarceration, conditional only on the nature of the crime. Under this model, the proportion of blacks in the arrested population will be equal to the proportion of blacks in the incarcerated population, controlling for the differential rates of incarceration per crime. He finds he can explain the vast majority of racial disparity in incarceration for serious crimes using this simple model, but a relatively small amount of the disparity for drug crimes. The power of this model comes from its "counterfactual", which can be compared with the observed outcomes. Garland et al. (2008) have recently argued that this approach should be used to analyze differences in practices over time and across places as a way of benchmarking progress.
Another example of this type of modeling comes in the context of plea bargaining. There is much modern research on the existence of the trial penalty, and most scholars reach the conclusion that the measured trial penalty is large and difficult to account for (Ulmer and Bradley 2006). But, papers in criminology do not typically propose a model that could provide a reasonable estimate of how big it “should” be. A simple rational choice model of expected utility suggests that a risk neutral defendant will take no plea deal that is greater than the expected value of the sentence at trial, which is simply the probability of a conviction at trial times the expected sentence at trial. If the probability of conviction at trial is 75%, then this model predicts that the plea sentences should by 75% of the trial sentences. To our knowledge, Smith (1986) is the only researcher who tests this claim. Smith’s model could account for the observed trial penalty on average, but could not explain all sentences at the individual level.

More recent examples where scholars create an explicit counterfactual for theory and empirical testing include the application of outcome analysis to the question of racial discrimination in police searches and bail setting (Knowles et al., 2001; Ayres and Waldfogel, 1994). These scholars start with a reasonable model of agent behavior, such as the fact that police officers want to maximize hits on searches. This behavior will not minimize crime, so it is not strictly speaking "optimal", but it seems to be a reasonable approximation of what police officers might want to do on the job. Then, the researchers derive what the "hit rates" should look like if this was what the police were in fact doing (i.e., the hit rates should be the same for all observable groups). Then, the observed behavior is simply compared with this predicted outcome. Even if one does not agree with this supposition of the model, the resulting conversation will be a step forward if it results in a new counterfactual about behavior (Engel, 2008).

3. What Are the Key Decision Nodes?

While more suitable frameworks for inquiry can lend coherence to our thinking about the problem of discretion, this problem cannot be assessed fully in the abstract. The exercise of discretion varies from setting to setting, and it varies more fundamentally at each stage in the criminal justice process from victimization to confinement and return to the community. We can see this by focusing on the decisions made by legislatures, police, prosecutors, judges and probation/parole officers.

The Legislature. Although sometimes overlooked as a decision-maker in the criminal justice system, the legislature has the power to make laws, and organize the judicial system. Mandatory minimums, determinate sentences, "truth in sentencing", and sentencing guidelines are all sentencing initiatives that translate directly into sentencing outcomes for defendants. Absent federal oversight, Reitz (2005), among others, has noted how the states have become laboratories for different sentencing structures that emanate directly from the legislatures. Blumstein and Beck (1999) and Raphael and Stoll (2009) have outlined how the rise in prison population has been caused by a change in policy rather than an increase in crime, a change that has started with legislatures.
The Police. The ABF survey revealed several significant decisions made routinely by low-level police officers (Goldstein, 1993): whether to stop and question, whether to initiate an investigation, whether to use undercover methods, whether to arrest, whether to physically detain the arrestee, and what charges to cite in the police report (p. 36). To this list can be added other decisions routinely made by front-line police: whether to respond to a call for service, how to initiate contact with a suspect, how much time to put into a particular matter, whether to conduct a street sweep, crack down, or other "special initiative" order maintenance operation (Kennedy, 2009, p. 126), when to pursue and when to abort a vehicle chase, how much force to apply to induce compliance (Skolnick and Fyfe, 1993, pp. 38-40), which witnesses to interview and suspects to interrogate, how much and which items of evidence to collect and process, when to use a lineup, and how much follow-up investigation to engage in after arrest.

Of particular salience to the question of police discretion was the ABF survey finding of pervasive police tendencies for selective enforcement and non-enforcement in the routine conduct of their operations (Goldstein, p. 37). This finding is troubling to the extent that the officer's decision is arbitrary; it is especially troubling when the arbitrariness manifests as discriminatory decisions all along the way from the decision to stop and question to the arrest decision.

Goldstein pointed out also that the exercise of discretion changes with significant changes in policing, including the community policing development, which gave increased autonomy to front-line officers; new systems of police accountability and transparency, which have altered the incentives of police at all levels; and re-definitions of the police function, to make it more service-oriented and less hierarchical, bureaucratic, and secretive (p. 61).

David Kennedy (2009) has extended Goldstein's perspective of discretion from the front-line officer to the organizational level. He observes that discretion can be exercised as policy by the precinct or district commander or department-wide by the chief, and that the use of such discretionary policy can have significant deterrent impacts on crime. (We amplify Kennedy's distinction between discretion exercised at the policy-setting level and the front-line officer in the next section.) He distinguishes between "predominantly ordinary" police activities and the selective targeting of resources on problems in the community: selected offenses and behaviors, misbehaving individuals, groups, gangs or networks. He includes the clarification and communication of such targeting practices as important discretionary tools in policing for the prevention of crime (pp. 109-110).

Prosecutors. The discretionary authority of the prosecutor has been called the most influential in the criminal justice system.12 This influence begins at the screening stage, where

12 According to former Supreme Court Justice Robert H. Jackson (1940), "The prosecutor has more control over life, liberty, and reputation than any other person in America." (p. 18)

Similarly, according to Albert Reiss (1974), the district attorney exercises "the greatest discretion
an assistant district attorney decides whether to file charges or decline prosecution based on the strength of the evidence in a case, the seriousness of the offense, the perceived dangerousness of the offender, and a host of situational factors, such as current problems in certain areas of the community. If the decision is made to file charges, subsequent decisions must be made regarding the amount of time and effort to put into each case, how to prepare and present the case to the grand jury and arraignment hearings, whether to divert the case to a special program such as drug treatment, whether and how to negotiate and whether accept a guilty plea, how to prepare a case for trial, and what sentence to recommend to the judge. The influence of the prosecutor, relative to the judge, is reflected in a statistic that surprises many: for every felony case that a judge presides over in trial, the prosecutor decides the fate of 14 adult felony cases brought by the police. (Forst, 2010)

The breadth of discretion available to the prosecutor is largely the product of opportunities to draw from a range of standards of evidence from the case entry stage, where arrests must meet the "probable cause" standard of evidence, to the trial stage, where a guilty verdict must meet the "beyond a reasonable doubt" standard. The prosecutor exercises discretion at several points along the way: to advise the police in seeking further evidence in a case or to interrogate other suspects; to decide whether to decline or proceed with prosecution in a given case; to recommend whether the judge should release or hold the defendant on bail; to decide whether to negotiate a guilty plea or to induce one defendant to testify or provide information against another defendant or suspect; to decide which items of evidence to reveal at the grand jury stage, whom to select from the jury pool, which witnesses to call to the stand, how much tangible evidence to present, and what sentence to recommend to the judge or jury.

While individual prosecutors make these determinations, the chief prosecutor, usually a district attorney, must decide how to allocate scarce resources across various stages of prosecution and how best to proceed generally for various offense categories: street crime, domestic violence, drug violations, child abuse, white collar offenses, and repeat offenders in violation of court orders and warrants. Joan Jacoby (1980) has identified four distinct strategic perspectives that drive the decisions made in any prosecutorial office: legal sufficiency, system efficiency, defendant rehabilitation, or trial sufficiency. These archetypes shape the exercise of discretion in case screening and processing. Chief prosecutors often issue unpublished guidelines to govern the decisions made at the screening stage: the decision about whether and how to charge the suspect must take into account the often conflicting goals of crime prevention, fairness, reform of the offender and reintegration into the community, and resource conservation.

Discretionary practice revolves largely around the prosecutor's views of the primary goals of prosecution. A prosecutor who sees the victim primarily as a witness for the state may be less inclined to regard restitution as important as one who sees the victim as an aggrieved party deserving of restorative justice. Such differences were observed by the ABF survey over 50 years ago. Worrall (2008) argues that the typical prosecutor in the United States has greater discretionary authority than in any other country. See also Jacoby (1980) and Kress (1976).
years ago (Remington, 1993, pp. 86-91). These differences may be a matter of unwritten policy at the office level or a matter of the exercise of discretion by an assistant prosecutor, or both.

Prosecutors tend to downplay the extent of their discretionary authority, but the differences from one prosecutor's office to the next -- even operating within the same jurisdiction -- can be stunning. Sharp differences have been found, for example, in the tendency for some district attorney offices to aim for high-quality convictions by rejecting arrests at a higher rate at the screening stage and then taking more cases to trial, while other district attorney offices tend to aim for more convictions, accepting cases that would be rejected in other jurisdictions and then relying more heavily on convictions by plea rather than by trial (Boland and Forst, 1985). These differences exist in the federal system even in an age of uniform sentencing guidelines, often working through the way in which prosecutors offer substantial assistance departures (Spohn 1995; Nagel and Schulhofer, 1992; Johnson, Ulmer and Kramer 2008) and statutes that carry mandatory minimums (Ulmer et al., 2007). One recent study found only small but significant differences at the prosecutor level within federal jurisdictions after taking into account jurisdictional differences and case characteristics (Spohn and Fornango, 2009). These results seem to support the claim that inter-office differences are due to variations in tradition and in the exercise of discretion at the policy-setting stage, not because the rules of evidence vary from one jurisdiction to the next or because of caseload differences.

Judge Charles Breitel (1960) and others have proposed more transparent guidelines to manage the prosecutor's broad discretionary authority, but such proposals rarely take hold in a meaningful way. Prosecutors prefer not to be second-guessed, and the Supreme Court has endorsed this view of limited accountability, under the general view that such intrusions into prosecutorial discretion would violate the separation of powers and could seriously retard the effectiveness of the prosecutor to protect the community.

Judges. Robert Dawson's 1969 treatment of discretion in sentencing was the first major inquiry into the problem of unwarranted disparity that results, and how this undermines respect for the law. His study emphasized that the problem is the product of the exercise of discretion not only by judges, but by other actors in the formal system of justice, especially police, prosecutors, and correctional officials. A more stinging critique of the exercise of judicial

13 Differences between offices of the Los Angeles County district attorney have been reported by Petersilia, Abrahamse, and Wilson (1990a), and earlier by Greenwood, et al. (1976). Differences between United States Attorney offices in adjacent federal districts were reported by Boland and Forst (1985), p. 11.

14 Boland and Forst (1985) found low rates of case rejection at screening in some urban settings, like Manhattan, and low plea-to-trial ratios in other urban settings, like Washington, DC, and New Orleans. They found also that jurisdictions with high case rejection rates at the screening stage tended to have lower plea-to-trial ratios.

discretion came three years later, from federal judge Marvin Frankel: "The almost wholly
unchecked and sweeping powers we give to judges in the fashioning of sentences are terrifying
and intolerable for a society that professes devotion to the rule of law." (p. 5)

A related concern has been expressed by Richard Posner (2008): judges often issue
rulings and make decisions based on ideology and perspective rather than on strictly legal
considerations. Especially in appellate review, their decisions can be driven by attitudinal or
political inclination, by strategic interests, by psychological and cognitive illusions, by hyper-
rational choice illusions, by excessive attachment to organizational precedent, or by rigid
legalistic interpretation. Posner argues that while ideology inevitably shapes the exercise of
discretion at the margins, judges should exercise discretion and use a balanced approach that
effectively combines pragmatic and legalistic considerations.

Probation and Parole Officers. The probation and parole functions in the United States
were designed principally from the model of rehabilitation, but today they serve as well the
shorter term goals of community protection and reintegration (Morris and Tonry, 1990; Dickey,
1993; Clear and Dammer, 2002; Petersilia, 2009). Probation and parole officers are the agents
responsible for achieving these diverse policing and social service aims.

While the discretionary authority of these officers is often severely hampered by
impossibly large caseloads, it is nonetheless considerable. In 2008, probation and parole officers
were responsible for over four million probationers and nearly one million parolees, over twice
the numbers, combined, as the populations in prison and jail (Glaze, Minton and West, 2009).
Probation and parole officers must make judicious use of carrots and sticks to induce
probationers and parolees to behave while under their authority in the community. The stick is
formidable: nearly half of all new prison admissions are probationers and parolees who violated
terms of probation or parole (Clear and Braga, 2002).

Joan Petersilia (1993) has identified several ways in which community correctional
officers exercise discretion. First, they conduct presentence investigations and prepare the
reports that shape sentencing decisions. Second, they enforce court-ordered sanctions and
provide security to the community by testing for drug or alcohol abuse, checking police arrest
reports, and monitoring community service and victim restitution requirements. These functions
have been informed and enhanced substantially in recent years through the expansion of
electronic monitoring and frequent drug testing. Third, they support the goals of rehabilitation
and reintegration through counseling and encouraging participation in education, training, and
employment programs. Fourth, they monitor restitution programs to ensure compliance with
court orders. Discretion is exercised in each of these activities; it is exercised most powerfully in
taking action to end the probationer's or parolee's community residence and move the offender
back to prison.

Discretionary Authority of Each Agent/Agency. Concern is frequently expressed about
the fact that the United States has the highest rate of incarceration in the world. This problem
can serve to motivate an inquiry into the effect of discretionary decision making throughout the
criminal justice system, by logically integrating information from each major stage at which discretion is exercised, and thus provide a way to understand the relative importance of discretion at each stage. The aggregate rate of incarceration -- prisoners as a proportion of the population -- can be estimated as a multiplicative series of factors that reflect each of these stages of discretionary decision and policy making:

* The crime rate -- offenses as a percentage of the population, expressed usually as "offenses per 100,000 residents" -- reflects the decision-making patterns of the citizenry to commit and then others to report offenses, for non-consensual crimes.\(^{16}\)

* The rate at which offenses result in an arrest reflects police effectiveness and discretion.

* The rate at which arrests result in conviction reflects the quality of police arrests and prosecutorial discretion and effectiveness.

* The rate at which convictions end in incarceration reflects sentencing laws and the exercise of discretion by prosecutors to recommend and judges to impose in-or-out terms.

* The average term of incarceration reflects sentencing laws and the exercise of discretion by prosecutors to recommend and judges to impose terms of incarceration.

In a steady state, the product of these five factors is inmates as a proportion of the population (Forst and Lynch, 1998). Leaving aside for the moment the prospective biases associated with a steady-state assumption, when the aggregate incarceration rate – inmates as a proportion of the resident population -- changes substantially for any crime category over a particular time period, one can examine the contribution of each of these major sectors of the criminal justice system to the change. Although one cannot easily disentangle discretion from effectiveness at each stage, analysis of this sort can provide indicators of changes in the exercise of discretion by police, prosecutors, and judges.

The steady-state assumption can bias these estimates in several ways. Because of lags from offense to arrest, arrest to conviction, conviction to incarceration, and the fulfillment of a sentence term, the steady-state assumption will be erroneous when any of the rates change over time, as each component will refer to somewhat different pools of cases in a particular period. In a time of falling crime rates the estimates of inmates per population will be biased downward (people recently entering prison are from smaller pools of offenders than were those currently incarcerated), and in a time of growing punitiveness, the estimates will be biased upward (people

---

\(^{16}\) These two important elements can be analyzed separately and are likely to be influenced by criminal justice agents; we can simplify the analysis for current purposes by regarding those elements as exogenous to the problem of discretion by criminal justice agents and agencies.
recently entering tend to be serving longer terms than those already there).\textsuperscript{17} Moreover, a given offense can give rise to more than one arrest, and the data to adjust for this unit-of-observation switch were not available for the current analysis. The estimates can be biased as well by incarcerations associated with probation and parole failures, about which data are not available on an original offense basis.

Recognizing these limitations, we can nonetheless examine a recent decade of considerable flux in crime -- from 1992 to 2002 -- to get a sense of the respective contributions of police, prosecutors, judges and sentencing laws to the increase in the aggregate incarceration rate. We choose this decade because it is fairly recent, a period of unprecedented decline in crime rates,\textsuperscript{18} and for which data on the key stages from reported offense to incarceration are available by major crime category for the nation's largest urban counties.\textsuperscript{19}

While the reported offense rates of major categories of crime -- homicide, rape, robbery, burglary, and assault -- declined by about half from 1992 to 2002, the prison population increased substantially, both absolutely (by 61\%) and as a percentage of the resident U.S. population (41\%).\textsuperscript{20} The aggregate increase has been attributed to recent increases in drug incarcerations, but for the urban counties sampled, both the incarceration rate and average term of sentence declined for drug offenses during this ten-year period, while the aggregate arrest and conviction rates remained fairly constant.\textsuperscript{21} The increase is in fact due primarily to tougher sanctions and larger pools of offenders in the 1980s, with smaller proportions of those cohorts exiting prison in the 1990s than in earlier periods. If we wish to do more about the problem of

\textsuperscript{17}For the period 1992 to 2002, the bias associated with declining crime was substantial, while that for punitiveness was, on the whole, negligible. This is evident in comparisons of the actual change in prisoners per capita with that estimated by the components of the steady-state model for the two years. For murder, the gross incarceration rate increased by 53\% from 1992 to 2002, while the steady-state estimate of the rate declined by 20\%. For aggravated assault the rate more than doubled, while the steady-state estimate of the rate increased by just 20\%.

\textsuperscript{18}The increase in imprisonments undoubtedly contributed to the decline in crime during this decade, by way of deterrence and incapacitation, although it is impossible to say by how much precisely. The question of optimal laws and policies should consider this effect. We discuss the exercise of discretion here without regard to its effect on crime rates.

\textsuperscript{19}The results reported here are from the Bureau of Justice Statistics National Judicial Reporting Program, reported in its \textit{Felony Sentences in State Courts} bulletins and the BJS State Court Processing Statistics, reported in its \textit{Felony Defendants in Large Urban Counties} series.

\textsuperscript{20}The prison population was 850,566 in 1992 and 1,367,547 in 2002; the combined prison-jail population was 1,295,150 in 1992 and 2,033,022 in 2002, which yield absolute and per capita increases in the gross incarceration rate of 57\% and 38\%, respectively. Source: Glaze, Minton and West (2009). The U.S. population was 255 million in 1992 and 290 million in 2002.

\textsuperscript{21}This finding shows up in both the BJS State Court Processing Statistics series for 75 large urban counties, which accounted for about one third of the U.S. population in 2002 and over half of the crime, and the BJS National Judicial Reporting Program series for 300 counties, which accounted for even more of both population and crime.
over-incarceration of drug offenders today, it might be more effective to focus on discretion
eexercised by police and prosecutors rather than by judges, especially in urban areas and
jurisdictions bound by draconian drug laws. Rachel Barkow (2008) has made a compelling
case for the expansion of such "merciful" uses of discretion prior to trial in situations where the
law has encroached excessively on commonsense notions of justice.

How do these findings vary from county to county? Focusing again on drug crimes, the
picture varies substantially from place to place. At the extremes, the number of inmates
generated per drug arrest increased fourfold in Montgomery County, MD, and Salt Lake County,
UT (from lenient to more punitive than average), while the ratio decreased by over 75% in
Fulton County, GA, Dallas County, TX, and Harris County, TX (from extremely punitive to
moderately punitive). The increases in punitiveness in Montgomery and Salt Lake Counties
were attributable in both cases to moderate increases in the conviction rate and approximately
threefold increases in average terms of incarceration, countering trends in most other U.S.
counties analyzed. The declines in punitiveness, as reflected by declines in inmates generated
per drug arrest, were due principally to sharp declines in the conviction and incarceration rates in
Fulton County, and in the two Texas counties due principally to a reduction in average sentence
terms from the 10-to-15 year range to the 2-to-3 year range.

Clearly, discretion is exercised quite differently from place to place, from one crime
category to another, and over time. Scholarly conversations about discretion rarely acknowledge
these substantial variations. We expand on this theme in the next section.

4. Discretion in Setting Rules and Policies

In thinking about discretion and decision making, it is tempting to focus exclusively on
the discretion of individual actors to deviate from the "rules". Judicial discretion, for example,
can easily be interpreted narrowly as the freedom of the judge to make decisions that differ from
the decision of another judge in the same situation. Clearly, this is judicial discretion, but so is
the decision of judges to adopt a guideline system, as they did in Maryland, or any other of a
number of other possible "systematic" decisions about how the courts should operate. The
immense variation identified in the above section at each of these stages is not simply the result

22 Unfortunately, the data did not include changes in the rate at which drug probationers
and parolees were incarcerated due to violations of terms of probation or parole. An analysis of
similar data for 300 large U.S. Counties produced similar results.

23 The results reported here are for the 24 counties for which data on offenses, arrests,
convictions, incarcerations, and average terms of sentence were available for both 1992 and
2002. These 24 counties have slightly higher crime rates, generally, than do the other 16
counties in the BJS State Court Processing Statistics data set, which is a two-stage stratified
sample, with 40 of the 75 most populous counties selected at stage one and a systematic sample
of state court felony cases selected at stage two.
of individual discretion by the actors in the system, but also a result of policies and rules at each stage that can, by design, affect outcomes in the system. Target outcomes may include a reduction in disparity, a reduction in crime, or an increase in justice (fewer errors).

It is indisputable that the last 30 years have seen a dramatic amount of discretion exercised by policymakers, with movements to determinate sentencing policy, guidelines, mandatory minimum statutes for some crimes, minimum time served requirements, and three-strike laws (Stemen, Rengifo, and Wilson 2005), not to mention the increased use of formal risk assessment at the parole and probation stages (Harcourt 2007). At the same time, some policymakers sought to limit the size of their prison populations (Marvell and Moody, 1996).

Rodney Engen (2009) summarized efforts to describe the impact of these policies as follows: "Regrettably, seemingly straightforward policy questions such as whether these sentencing reforms achieved any of these objectives . . ., are difficult answers. There is, simply, little research (p. 324)." This lack of empirical research is also reflected in the lack of theoretical development, where, according to Engen, the "preoccupation with detecting and explaining unwarranted disparity in sentencing research has been accompanied by a near-exclusive emphasis on individual level social psychological theories of decision making and to a lesser extent on contextual theories (e.g. racial threat) that still emphasize subjective decision making as the central causal mechanism.(p. 333)."

The most striking example of this lack of research involves the lack of real understanding of the causes of the massive increases in incarceration during the last 30 years. While post hoc analysis (Blumstein and Beck 1999, Raphael and Stoll 2009) concludes that the increase is due to policies rather than changes in crime, there is little consensus about which policies have led to these changes. Research on specific policies, such as determinate sentencing, seems to show no consistent effect on incarceration rates (Engen 2009).

The most productive area of policy research involves the study of guidelines and their ability to reduce disparity. Legal scholars such as Tonry (1996), Frase (2005), and Reitz (2005) believe that the presumptive sentencing as practiced in states like Minnesota and Washington have accomplished their goal of reducing disparity. These conclusions are based on analyses that show that there are small or non-significant disparities associated with race and ethnicity in states with presumptive guidelines, and states with less strict guidelines often have stronger effects of race (Bushway and Piehl, 2007). However, systematic comparisons across jurisdictions do not appear to show large differences in disparities in non-guideline systems relative to guideline systems (Engen 2009, Mitchell 2005). Moreover, the few before-and-after studies that do exist show either no differences or a small difference after guidelines, in part because there is not large disparity before the guidelines were implemented (Miethe and Moore, 1985; Wooldredge 2009). One possible exception is the recent update by the U.S. Sentencing Commission (2010) on racial disparity post Booker. The new report appears to show a fairly sizeable increase in racial disparity after the Booker decision, which made the federal sentencing guidelines advisory.
These types of before-and-after studies are rare and hard to conduct, at least partially because the available data changes after guideline systems are implemented. But, the larger problem comes from the fact that these analyses are conducted on conviction data, which cannot take into account the changes that occur in the processing of cases from arrest to conviction by actors in the system in response to the changes in the rules. At the same time that sentencing practices change, charging practices, and, even conviction probabilities, also change in meaningful ways, and disparity that used to be present at the conviction stage might reappear at another stage (Alschuler 1978; Piehl and Bushway 2007; Miethe 1987; Engen 2009).

This leads directly to an observation about the needs of empirical research about the impact of systematic discretion on disparity in the criminal justice system. Because we are interested in disparity systemwide, the analysis of rules and policy changes should be conducted on the population of people at whom the policy is directed. Guideline rules are directed at reducing disparity for everyone in the criminal justice system, not just those convicted of a certain crime. Moreover, punishment is not over at sentencing. This means, in practical terms, that we need to collect data on those who are arrested, and follow them through their final release from the system. Then, we can assess more completely how the policy affected disparity after taking into account the actions (and counteractions) of the other actors in the system.

This recommendation is directly related to the well known problem of selection that has haunted research on sentencing since it was first observed in 1983 by Richard Berk. Studies of race or other factors that may be prone to discrimination typically want to know the overall, unconditional impact of that factor on the sentence. Yet, we often have only a conditional or selected sample. For example, we have only those people who are convicted or those who are incarcerated. We need to conduct some type of statistical technique, usually some version of the Heckman correction for selection effects, to create inference on the unconditional estimate from the conditional sample. The empirical demands of these techniques are strenuous (Bushway et al., 2007), yet we do not need to do corrections for selection if we use unselected data.

Very few data systems exist that allow for this type of analysis, but where possible, this type of analysis will lead to a more accurate description of what is actually happening in the system. A reasonable benchmark for disparity might include an analysis of who served at least one year in prison among those arrested for crimes. Discrete dependent variables will allow us to avoid the thorny problem of censoring typical in incarceration analyses. This might not be the ideal benchmark, but considerable progress could be made in understanding disparity if easily measurable benchmarks were available across jurisdictions, and over time, for an unconditional population.

We also agree with Engen (2009) that the emphasis on disparity has taken away from direct research on the policies themselves. Researchers look to see if race matters more or less after the guidelines, but spend little time looking to see if sentences change in response to the
Guidelines are updated regularly, but researchers rarely look to see if these changes show up in the outcomes of the system for individuals affected by the system. Emily Owens (2009) examined a change in the Maryland guidelines in how criminal history record scores are calculated. The sentencing commission decided to lower the age at which the juvenile record no longer mattered for the calculation of the criminal record score from 26 to 23. This meant that 23-to-25 year olds would be treated differently after the change, while everyone else would be treated identically, creating a meaningful comparison group. Owens showed that 23-to-25 year olds with juvenile records receive sentences about 7 months shorter after the change in guideline policy. This result shows clearly that the sentencing commission has the ability to affect punishment in Maryland.

Another strategy to study the impact of guidelines is based on human error. The worksheets to compute guideline recommendations are complicated, and errors are common. If the system is largely ignoring guidelines, then the cases with errors in the calculation of guideline sentences should be treated the same as similar cases without errors. But if guideline recommendations matter, then cases with errors should be treated differently than those without. This approach has been applied to the Maryland guidelines (Bushway et al. 2010). The guidelines appear to constrain judges from giving lower sentences, an important finding that speaks again to the ability of legislatures to implement policy even in a world where downstream actors have considerable discretion.

Researchers also do not tend to study the guidelines or statutes themselves when looking at disparity. The "crack cocaine" enhancement in the federal system, which treats crack cocaine differently from powder cocaine (with dramatic implications for race) is a rare exception (The Sentencing Project, 2009). Researchers now routinely use the presumptive sentence of the guidelines as a control (Engen and Gainey 2000, U.S. Sentencing Commission 2010) in models of disparity, but do not study the guidelines themselves. Yet, guidelines themselves can create disparity, especially since the impact of any given aggravating/mitigating factor depends on the location on the grid, which is determined by the other factors. For example, weapon possession can carry a stiffer penalty for those with longer criminal histories. If blacks tend to have longer criminal histories than whites, the guideline can itself create disparity, a fact that will be missed in standard analyses. In the Owens paper (2009) mentioned above, the guideline change reducing the penalty for juvenile records differentially advantaged blacks, who are more likely to have juvenile records.

King et al. (2005) explicitly compare an outcome (in this case plea discounts) across guideline systems. They found substantial variation that was hard to explain by the guidelines themselves. More explicit comparisons are also possible. For example, researchers could take the average African-American and white offenders in Pennsylvania, and calculate their recommended sentence in Pennsylvania and in Maryland. Then, researchers could do the same thing for the average Maryland offender. Although this will be an oversimplification, this type

\[\text{For exceptions, see research on whether guidelines that aimed to restrict prison population were able to accomplish this task. See Marvell (1995) for a review.}\]
of simple comparison of the guidelines (rather than the outcome of the entire system) could serve as a useful starting point for a discussion about how different guidelines mete out punishment. This simple comparison could provide a systematic comparison of the two guideline systems themselves. This relatively simple exercise might be a first step forward towards more complicated inter-jurisdictional analyses using datasets like the State Courts Processing Statistics.

We should be clear that such a comparison should focus on both mean differences in outcomes across states – after all, it is interesting to know how different states view the same crime – but also in differences in disparity/dispersion across states. At this point, we do not have a strong understanding of how disparity varies across states. From a policy perspective, it is important to understand how different policies can affect both the mean and the variance.

It is difficult to compare administrative data across jurisdictions because even the basic facts of the cases, like the type of offense, are determined idiosyncratically across jurisdictions. However, the Bureau of Justice Statistics may be amenable to changes in how they collect the data if researchers begin struggling directly with these kinds of issues. Interstate research is clearly the holy grail for researchers interested in studying intentional discretionary differences in the process, given the existence of 50 very different state regimes as independent laboratories for sentencing in the United States. Concerted efforts, perhaps sponsored by funding agencies like NIJ, in conjunction with data collection agencies like BJS, are needed to get us out of the "talking" stage and into real analyses. The good news is that the State Courts Processing Statistics, which has data from indictment to conviction has considerable promise for this type of inter-jurisdictional analysis, and is currently underexploited by researchers.

Of course, it would be a mistake to focus only on the punishment levels themselves: policymakers desire to influence other outcomes as well. The econometric analysis of unconditional effects of laws intended to reduce crime, such as three-strike laws or gun control laws, are examples of issues that call for empirical analyses that look at net effects (after the system responds) of policies on desired behavior. In these types of models, the impact of the law is examined on crime in a place, usually a state or county. The first step of these types of analyses is to show that the laws were implemented. The second step is to conduct a before/after analysis with aggregate data on crime (Levitt and Miles, 2007). A similar strategy has been applied in political science to the study of different election practices on average sentences at the county level (Gordon and Huber 2007). The causal power of these analyses is often open to question, because of the inherit limitations of aggregate, state or county-level analyses (Chen, 2008; Committee to Improve Research Information and Data on Firearms, 2004). The challenge is to find a reasonable counterfactual in the absence of an experimental trial.

---

25 To be clear, we are not comparing outcomes, but simply guideline recommendations. It will be even more difficult to compare sentencing outcomes across jurisdictions, because many more factors are involved in difficult to define ways. Piehl and Bushway (2007) do conduct this exercise with outcomes, but not everyone agrees that this exercise was useful.
Another potentially powerful approach to the study the impact of laws on behavior involves a comparison of individuals who are subject to the law with those who are not. The advantage here is that it is easier to create a believable counterfactual. For example, Carpenter and Dobkin (2009) look at changes in mortality as people age past the drinking age to produce credible evidence that the drinking age laws have the desired impact on drinking behavior. Individuals essentially serve as their own control in this approach. McCrary (2008) and Hjalmarsson (2009) both use individual data to examine the impact of moving from the juvenile justice system to the adult system on behavior. These papers are noticeably more compelling than the standard cross sectional comparison using aggregate data for comparisons between systems such as the juvenile and adult systems (Levitt, 1998). Similar analyses that focuses on the differential behavior of those who are subject to Three Strikes laws in California (Helland and Tabarrok; 2008) offer what is probably the best available evidence on the ability of these laws to affect desired outcomes. More effort should be focused on collecting individual level data that links sentencing data and data on criminal behavior like arrest.

The state of Florida stands as an exemplar for this kind of data. The Florida Statistical Analysis Center in the Florida Department of Law Enforcement makes data available to researchers that integrates the state police arrest repository with a correctional dataset that includes all convictions. This set-up allows researchers to analyze the impact of laws such as those that defer conviction (and avoid labeling) on subsequent criminal behavior (Chiricos et al. 2007). This dataset could be improved even further if the arrest data could be linked with dispositions, so that analyses could be conducted on a sample of arrested individuals, rather than simply those who are convicted, in order to deal with the selection problem we discussed above.

Before closing this section, it is important to acknowledge that not all policy discretion is identified as laws or explicit rules. Prosecutors, judges, defense attorneys, and probation/parole officers all work in group settings where there is undoubtedly formal and informal policy about how to handle certain types of cases. Moreover, these actors undoubtedly interact to create workgroup norms that represent a certain kind of non-individualized discretion (type of multi-level models that capture the "average" behavior of district work groups (Steffensmeier et al. 1998). Researchers have used multi-level models to identify the average effect at district level, which is then identified as the result of unobserved policies at the district or court level (Ulmer and Johnson 2004). Although there is undoubtedly meaningful variation at this level in sentencing datasets, efforts to explain this variation with observable contextual variables have been largely unsuccessful. We share Engen's (2009) skepticism that social contextual theories such as racial threat will ultimately be able to explain the type of systematic discretion being captured by these models.

---

26 See also compelling examples that look at the impact of state laws on work behavior for juveniles by Rothstein (2007) and Apel et al. (2008). In these cases, different state laws on working for 16 year olds have dramatic impacts on changes in work effort as people age from 15 to 16.
We are also uncomfortable relying on empirical models to differentiate between individual discretion and that at the group or policy level. In these increasingly common models, researchers first estimate a case-level multiple regression model with legal and extra legal factors included in a reduced form linear model. Then, variation on the coefficients at the district level is then explained using context specific factors at this level (Ulmer and Johnson 2004). The case level model is theoretically uninteresting (researchers do not argue that this linear model captures the decision-making process), so variation in the coefficient is also less interesting. If we want to test how a particular set of actors are acting, we should propose a theoretically interesting model for that behavior. Then, and only then, will the multi-level capture interesting variation that can be explained by researchers. As discussed in Section II, we believe that it is useful to develop theoretical models, even simple ones, that can create accounts of what "should" happen at this level. If the observed means equal the predicted or expected mean, we have created insight into the processes that are happening at this level. If they do not, we have a basis for developing new models.

This type of theory-based modeling has been usefully applied by Smith (1986) with respect to pleas. We believe this approach is more promising than the dominant, but uninteresting, null model that claims that legally relevant factors such as criminal history and type of crime, entered in a linear model, should explain all observed variation (e.g., Steffensmeier, et al. 1998). After all, guideline recommendations, which by definition rely only on legally relevant factors, cannot be explained by a linear model of the legally relevant factors used to create the recommendations (Mustard 2001). The introduction by Engen and Gainey (2000) of the presumptive sentence as the starting point for judicial decision making in guideline jurisdictions represents one such model. In this model, deviations from these recommendations become the key variable, and models explaining this deviation yield different results from models that rely on a "legalistic" model as the alternative (Bushway and Piehl 2001).

Formal theoretical models can also help generate useful insight about the tension between multiple goals in the criminal justice system. David Bjerk (2007) created a simple model that demonstrates the inescapable tension between punishment and fairness in a world with guilty pleas. This model focuses attention on the plea bargain as the key place where our current system of justice can create errors, especially for severe sentences (see also Ulmer, Eisenstein and Johnson, 2010). While not easily estimable, these models might generate useful stylized facts which can then be compared with observed outcomes (Landes 1971, 1973, 1974; Forst and Brosi 1977).

5. Discretion Beyond Rational Choice

There is no need to stay within the world of rational choice when building these models of decision making. One particularly promising area involves the use of heuristics in lieu of
decision models that assume rationality. Amos Tversky and Daniel Kahneman's (1974) research on the use of such models under conditions of uncertainty and complexity gave rise to the behavioral economics revolution, with rich implications for policy in a broad variety of settings, including the criminal justice system.

Guthrie, Rachlinsky and Wistrich (2005) applied the logic of behavioral economics and heuristics to judicial thinking and found biases in the way judges think when presented with hypothetical cases. They surveyed 167 federal magistrate judges anonymously (out of 519 on the bench), using twelve variations of a five-page questionnaire to test for each of five cognitive biases in civil cases. They found that even experienced judges tended to rely on cognitive decision-making processes that can produce systematic errors in judgment.

The selection of rules and use of discretion by judges and other criminal justice agents is especially ripe for the use of knowledge of how rational and non-rational modes of thinking blend in shaping the routine behaviors and decision processes of prospective offenders, victims, and key actors in the justice system. For example, to the extent that the rational model of behavior applies to particular categories of offending, prospective offenders will be deterred either by tougher sanctions or by the exercise of discretion in ways that increase the likelihood of punishment, exploiting tendencies for offenders to understate the risks of arrest, prosecution and punishment (Polinsky and Shavell, 2000).

Conventional ("neoclassical") economic theory assumes a world of rational behavior under perfect information, and this model often predicts fairly well even when the assumptions are only approximately true. But several factors may tend to invalidate these assumptions in criminal justice settings. Here are two that could be especially prominent: 1. The decisions confronting the police, prosecutors, judges, and others who influence sentencing outcomes are often too complex to lend themselves readily to the agent's knowing which of the available options actually maximizes utility; 2. The utility functions of the officials acting on behalf of the agency often depart from the utility functions that correspond to their personal self-interested goals. It is common to use frameworks that have been created in the field of behavioral economics to deal with these departures from the neoclassical norm, to build a theoretical framework that corresponds more closely to the actual thought processes of these agents.

---

27 Here are three prominent types of heuristics and the associated biases: 1. anchoring and adjustment -- assessing new problems using facts known about similar instances and making adjustments (bias: new problems may be different, and the adjustments may be incorrect); 2. availability -- exaggerating more dangerous, familiar events (bias: exaggerations often lead to over-reaction to new cases); 3. representativeness -- inference based on anecdote rather than valid empirically derived probability (bias: anecdotes are often atypical).

28 The researchers found the judges to be susceptible especially to five "cognitive illusions": anchoring (basing assessments on irrelevant premises), framing (treating identical effects differently through language), hindsight bias (ignoring difficulties in predicting before knowing outcomes), representativeness (replacing valid base rate evidence with anecdote or inverted conditional probabilities), and egocentric biases (overestimating one's own abilities).
An approach along this line of inquiry was considered by Celeste Albonetti in 1991, where she assumed that people use stereotypes as shortcuts to make decisions when under time or workload constraints. This idea was adopted by Steffensmeier et al. (1998) in their focal concerns theory. The problem with this approach is that the test of this theoretical idea has been reduced to looking for race after controlling for legal factors in a linear model. This linear model is not in fact a reasonable presentation of the “rational choice” model. Moreover, the behavioral economics model will generate richer predictions that will allow for stronger tests of the model. We believe that a more explicit focus on the models from behavioral economics could generate important advances in sentencing.

6. Abuses of Discretion

Perhaps the most serious concern about discretion is the potential for agents of the justice system to abuse it. Police corruption and brutality and sensational acts of prosecutorial misbehavior are especially toxic and costly. They impose immediate costs on the people at the receiving end of these abuses, and they can impose even more serious long-term costs on society by undermining the legitimacy of the criminal justice system. Examples of such abuses are legion. In policing, corruption often manifests as solicitation or acceptance of bribes in exchange for failure to either report consensual crimes or make arrests in such cases, and in the falsification of evidence to make arrests and secure convictions of suspects. Prominent cases of such abuses were revealed by the Knapp Commission, which investigated corruption within the New York City Police Department in the early 1970s (Skolnick and Fyfe, 1993) and the Rampart Scandal, involving brutality and corruption by 70 police officers of the anti-gang unit of the Los Angeles Police Department's Rampart Division in the late 1990s (Chemerinsky, 2000). The abuses included unprovoked shootings and beatings, planting of evidence, framing of suspects, stealing and dealing narcotics, perjury, and covering up evidence of these activities.

Abuses of prosecutorial discretion are less well known, but not less important than those committed by the police. The importance of these abuses derives from the position of the prosecutor, considered by many authorities to be the most powerful of all criminal justice agents. Their power resides in the authority they have to file cases in court and determine how to charge each case accepted; to determine how much time and effort to give to each, driven largely by the decision whether to prepare for trial or settle with a plea; and for those convicted, to determine what sentence to recommend to the judge or jury. As elected officials in most jurisdictions, prosecutors are publicly accountable for these decisions, but available evidence provides virtually no basis for knowing how well the prosecutor in any community fares in exercising all this authority.

A prominent example of gross abuse of prosecutorial exercise of discretion came to light in the 2006 case of Durham County District Attorney Mike Nifong's aggressive and dishonest charging of members of the Duke University lacrosse team for allegedly raping a woman hired
as entertainment at a party. (Taylor and Johnson 2008). Observers have also noted a federal judge's 2009 acceptance of Attorney General Eric Holder's motion to set aside the verdict and dismiss the indictment of Alaska Senator Ted Stevens with prejudice, based on what the judge called "the worst case of prosecutorial misconduct" he'd ever seen (Wilber, 2009). Reports of such abuses are less common than reports of the celebrity cases prosecutors exploit now and then. More remarkable is the lack of scrutiny given to abuses in run-of-the-mill felonies that go unnoticed. Abuses in sensational cases that come to light because of intense public scrutiny make the prospect of abuses in these other cases all but inevitable and worthy of more attention.

An obvious solution to the problem of abuses of discretionary authority is increased transparency. Police operations have become substantially more transparent over the past several decades, thanks largely to the development and use of more effective systems of accountability and a more intrusive media, but the operations of the prosecutor have not.

Transparency is also the first step in identifying abuses in discretion empirically. Data on individual actors collected as part of routine administrative procedure raises the potential to identify prospectively individuals who appear to be using their discretion abusively, or at the very least in ways that make them outliers worthy of future investigation. There are very few examples of this type of research, in part because the criminal justice system is generally loathe to release data on individual actors. For example, very few states with sentencing guidelines release data with judge identifiers – Pennsylvania and the District of Columbia are notable exceptions. Although confidentiality concerns are often cited, these are publicly paid actors performing public roles; we argue for more movement towards the release of individual identifiers in these administrative datasets.29

Once this data is released, we need to understand how to identify individual discretion in the data. The biggest problem, as mentioned in Section 4, is how to distinguish between actions of the individual and actions of the group. The best possible option is to have actions of the group codified in a way that is available for research. We can then study the deviations from the rule as individual discretion. Generally, however, we do not have the policy rule for judges, prosecutors or police, so we need another way to distinguish individual action from group norms, or case characteristics.

One paper that explicitly tries to identify potential discretionary abuse in individual police actions is by Greg Ridgeway and John MacDonald (2009). They use brute force statistical methods that try to benchmark "standard police stopping behavior" for comparison purposes so as to identify possible outliers. Police who were flagged as outliers were then investigated by the New York City police to determine if they were in fact engaged in questionable discretionary stopping behavior. Forst et al. (1982) conducted a similar analysis using predicted conviction rates for each officer's case mix in the cities of New York and Washington, and then followed up with in-person double-blind interviews with the outlier officers at both ends of the conviction rate scale, finding that the officers at the high end were significantly more likely to produce

29See also Schanzenbach and Tiller (2008) for a similar call.
multiple witnesses and conduct follow-up investigations. Many systems in the criminal justice system have the kind of large datasets that would support this style of analysis.

Although researchers do not often use this approach as a way of aiding investigations of discretionary abuse, researchers do look at variations from the mean, or average actions of the group in standard academic work. This is usually done with multi-level or fixed effect models, where individual observable case characteristics and the average actions of the district are first isolated, leaving only variation that is at the individual level.

Spohn and Fornango (2009) used this approach to study prosecutors. They found a non-trivial level of unexplained variation at this level, which could potentially be misused discretion, although they found no evidence that this discretion was correlated with characteristics of the individual prosecutor. This finding is consistent with aggregate analysis at the judge level that does not find that aggregate characteristics of the judges at the district level can explain “judge effects” on average (Johnson 2006, Schanzenbach 2005), although they might explain race and sex-disparities (Schanzenbach 2005). Schanzenbach and Tiller (2008) perform a unique individual level analysis of federal judges. They find non-trivial variation in sentence length, and they find that it is correlated with political affiliation. Equally interesting, they find that the discretion is modified or constrained by the nature of appellate review. This type of analysis using individual identifiers, rather than average behaviors, demonstrates the importance of these individual level identifiers.

The multi-level approaches that focus on average individual behavior are limited: we can control only for case characteristics that are observed. It is at least plausible that some actors get different cases than others. In the case of police discretion, controlling for group means at very low levels of aggregation can help solve this problem (Ridgeway and MacDonald 2009). An alternative approach is to use the random assignment of cases to identify the impact of individual actors.

This approach has been used by a small group of researchers as part of an instrumental variable approach. In these studies, random assignment is used to identify differences in sentence lengths, which is then used to study specific deterrence (Green and Wink 2010) or the impact of sentence length on employment (Kling, 2006). Unfortunately, the individual variation due to the judges themselves is not studied for potential abuse, and is used only to identify subsequent effects.

The problem with this latter use is that the variation itself is not modeled. Judges can sentencing people to different sentences for many reasons, including their own judicial philosophy of punishment. Don Gottfredson collected data on judicial preferences regarding punishment, and showed how they mapped to actual sentencing practice (Gottfredson, 1999). The fact that different actors are doing different things makes it hard to interpret the findings from these types of reduced-form instrumental variable analyses, which do not identify the mechanism by which sentences are handed down by these individuals. This point was made
clear in an important but often overlooked paper by Charles Manski and Daniel Nagin (1998), which identifies the limitations of these types of reduced-form treatment effect models. Unless you can model the decision-making process, it is hard to identify precisely the treatment impact of sentence length (or any other outcome) on subsequent behavior.30

Modeling decision-making behavior can also help identify possible abuses of discretion. Steven Levitt and Mark Duggan (2002) famously presented an analysis of corruption in sumo wrestling that is based on reasonable models of wrestler incentives. They look at matches where there might be incentives to cheat and try to rule out alternative explanations. Brian Jacob and Steven Levitt (2003) identified plausible cheating patterns on a high stakes test for elementary school students, and then used these patterns to flag teachers who were potentially cheating. In a similar vein, political scientists Greg Huber and Sanford Gordon (2004) used Pennsylvania data to show that sentences by the same judge over time varies with proximity to elections, with judges becoming more harsh when they were facing reelection.

Papers that generate interesting and plausible models for how individuals can (ab)use discretion have tremendous potential to add to the understanding of discretion at the individual level. This discussion brings us back full circle to the need to model behavior directly. Research on discretion and decision making must propose models of behavior, either explicitly or implicitly, and then work hard to identify data that can be used to identify this behavior with as much structure as possible.

7. The Future of Empirical Research on Discretion

Discretion is inherently neither good nor bad. It can be used skillfully to counter ill-conceived or vague laws and policies, or to minimize wrongful arrests and convictions, excessive punishments, and failures to bring culpable offenders to justice. But it can be misused as well, with immediate harm to the victims of the abuse and long-lasting harm to the legitimacy of our system of criminal justice. Research on discretion is vital to helping policymakers manage and control discretion. In this paper we have tried to review what we think are key areas of research in discretion. In this section, we try to clarify what we think are key tasks for making progress on the study of discretion. We hope this presents a useful starting point for a discussion about a productive research agenda.

First, researchers should identify the total amount of discretion that is exercised in the system. Al Blumstein’s (1982, 1993) landmark papers looking at discretion with respect to racial differences present a key way forward in this area. The goal here is not to model the discretion of any given actor, but rather to identify variation that is introduced by the entire system after arrest. This research demands access to data on people from arrest to final disposition. Conviction-only datasets miss large numbers of people, and therefore do not provide us with an

30 For a more general treatment of this problem, with a compelling and interesting example, see a paper on peer effects by Carrell et al. (2009).
overall assessment of discretion in the system. This research also requires some type of counterfactual model of what “should” happen.

Blumstein claims that racial disproportionality at arrest should be reflected in the disproportionality at incarceration if the system does not use its discretion in a racially disparate way. Discretion is then represented as deviation from this norm. Blumstein’s finding that drug cases have the most discretion is an important finding that can support more exploration. We join with others (Garland et al. 2008) in calling for more of this type of research, and present some statistics that could be used to monitor discretion across jurisdictions and over time.

Second, researchers should analyze the ability of policymakers at all levels to affect outcomes like levels of punishment, crime and errors. We recognize that the action of any given actor will be affected by other actors, but we believe it is important to identify whether the purposeful use of discretion by policymakers can have the desired impact on sentencing outcomes. Key questions of this type include asking whether guidelines reduce system-wide disparity, or whether mandatory sentencing policies reduce crime. As in the first point, this is an aggregate or system-wide question, and should be examined using data on all people being acted on by the system (i.e., not conviction data), if possible. Showing that guidelines reduced disparity at the conviction stage, without considering other stages, is inherently insufficient for analyses on the overall impact of a policy. Simulated comparisons of what “should have” happened if the policy was fully implemented with what actually happened will provide a sense of the level of discretion available to downstream actors. Analyses using selected samples should make use of sample selection techniques in an effort to estimate the “unconditional” coefficient that is the ultimate objective of this type of model (Berk 1983, Bushway et al. 2007).

Third, researchers should continue to try to study the action of a collective set of actors, like prosecutors or judges, using empirical methods that allow the researcher to isolate the actions of that set of actors over and above the actions of other actors. This can sometimes be done with survey techniques or hypothetical exercises like those used by psychologists to study trial outcomes (Rachlinski et al. 2009), but it can also be done using administrative datasets, provided researchers use sentencing structure to identify the actions of a given set of actors. For example, certain types of departures in the federal system can be offered only by prosecutors, so a study of those departures is inherently a study of prosecutorial action (Spohn and Fornango 2009). Researchers can also impose structure on the problem by making assumptions about the actions of individuals in the system. For example, Engen and Gainey (2000) and Bushway and Piehl (2001) argue that judges use the recommended sentences as a starting point for decision-making. In this framework, judicial discretion is the deviation between the sentencing outcome and recommended sentence. Economists have also started applying formal models of behavior to criminal justice actors, which sometimes are accompanied by estimable models (i.e., outcome analysis.) These models allow for formal tests, and can generate useful discussions about how exactly classes of actors apply their discretion.

When choosing which sets of actors to study, we also caution that the effort should not
be driven exclusively by the availability of data. We can and should solve specific problems because the data and tools are readily available to do so, but to restrict our analyses to issues for which data are readily available is to look for our lost keys only under lampposts, whether we lost them there or not. Areas in which we have good reason to believe that the social costs associated with our lack of knowledge of pertinent facts are greatest are the areas in which we should work most diligently to acquire the needed data and estimate the effectiveness of alternative uses of discretionary decision making and policy.

Our vote for the area most in need of illumination goes to prosecutorial discretion. While the reported offense rates of major categories of crime declined by half from 1992 to 2002, the prison population increased substantially, by 61% absolutely and by 41% as a percentage of the U.S. population. The increase has been widely attributed to increases in drug incarcerations, the area that Blumstein has defined as having the most discretion. For the urban counties we sampled, both the incarceration rate and average term of sentence declined for drug offenses during this ten-year period, while the aggregate arrest and conviction rates remained fairly constant. This suggests to us that a true understanding of this problem needs to focus on the discretion exercised by police and prosecutors rather than by judges. Rachel Barkow has recently (2008) made a strong case for the expanded use of prosecutorial discretion -- because of its relative unreviewability -- in circumstances in which the law has encroached excessively on commonsense notions of justice as mercy, and researchers can help to explore these prospects.31 More generally, most cases are resolved at the stages of arrest screening and by plea bargains, over which prosecutors have substantial discretion, yet research on those two critical stages is limited in the criminology literature.

Not surprisingly, information about prosecutorial operations is hard to come by. The operations of prosecutors throughout the land are today less transparent than they were throughout the twenty years following 1973, when aggregate information about arrest rejections and case dismissals, pleas and trials were reported by jurisdictions across the United States -- large, medium, and small, and by major crime category. This ended in the early 1990s, when the Bureau of Justice Statistics discontinued the collection and reporting of its cross-jurisdictional series on the prosecution of felony arrests. New ways need to be developed to study prosecutorial discretion, possibly with focused new data collection efforts. The State Courts Processing Statistics, which has data from indictment forward, is one place that researchers could start. The methods of researchers in the field of psychology and law, which so far have been focused mostly on trial outcomes, might also be usefully applied to the actions of prosecutors and others involved in the plea bargain decision.

Fourth, analyses should explicitly test for the interaction between actors in the system. Upstream actors can constrain downstream actors, but downstream actors can also temper

31 Barkow observes as well that this may be more easily said than done, as prosecutors face strong incentives to win convictions and weaker incentives to show mercy. Many are elected officials who run for office as protectors of the community against crime: "prosecutors tend to focus solely on law enforcement concerns. Error! Main Document Only. (p. 1363)
upstream actors. For example, the main theory of plea bargaining is referred to as “bargaining in the shadow of the trial”. Stephanos Bibas (2004) has raised important concerns about this theory, but there are little direct tests of this type of model. But the logic provides a useful framework. Models of trial outcomes (and psychologists have done this quite well) offer a set of expectations that we can bring to plea bargaining behavior.

Fifth, wherever possible, data on individual actors should be collected and analyzed. While we can study judges or prosecutors as a class, we can also study variation within this class if we can identify the individual actors who are responsible for a given case. This is particularly valuable if we have multiple cases per actor. Such data can help researchers identify actors who are either misusing or abusing their discretion. Systematic interrogation of data to identify outliers prospectively would be a positive use of data for policymakers. Criminal justice agents must be held fully accountable for their exercise of discretion: rewarded for using it well, corrected for using it poorly, and sanctioned for abusing their authority to exercise it prudently. Researchers can help in setting up systems of accountability that achieve all of these basic elements.

Individual level data can also help identify actors who appear to be doing their job particularly well, something we can hopefully model. Theoretical models of individual decision makers, based on differences in philosophy or experience, have much to teach us about the way actors behave in a system. Random variation in the assignment of actors can also be used to identify exogenous variation in sentencing outcomes, but caution should be exercised in interpreting these types of models. The variation in sentencing outcomes is itself not random, and this variation could affect interpretations of the results of a model.

Finally, increased efforts to model discretion and decision making at the individual level can help researchers build normative models for what policymakers “should do”. If we understand the use of discretion in the criminal justice system, we can then propose models for ways to shape and change this discretion. Essentially, researchers can build simulation models that will allow policymakers to evaluate different outcomes. The setting of rules and laws and the exercise of discretion stand to benefit substantially from an appreciation for the importance of social costs, the risks of justice errors, and the incentives of the key actors in the criminal justice system -- and how they relate to one another.

The prospects for reducing social costs by reshaping rules and the management of discretion appear to be considerable within three major sectors of the criminal justice system: policing, prosecution, and adjudication. This type of cost-benefit analysis is seen most often with respect to levels of incarceration (Donohue, 2009), but it can also be used effectively in other areas of decision-making, including the development of guidelines and risk assessment tools. Policymakers are trying to minimize overall risk, but they might be more afraid of some types of risk than others. Careful consideration of relative costs of different risks could lead to a better understanding of behavior – and more constructive efforts to constrain or manage
discretion and to encourage criminal justice officials to exercise it with greater awareness of the consequences.

In addressing these questions, we must remain realistic about the limits of our ability to estimate the pertinent variables -- especially the risks and social costs associated with alternative policies and boundaries of discretionary authority -- and respectful of the democratic processes that supersede our analytic discoveries. Those processes make many of these analytic ventures possible, by providing data on variation that can be studied. Moreover, the competing motivations of the actors in the system must be a central feature of models of discretion. Better understanding of these processes will allow researchers to better understand discretion, and allow empirical research (and researchers) to more fully participate in the ongoing debate about the management of discretion in the criminal justice system.
References


Barbara Boland and Brian Forst, "Prosecutors Don't Always Aim to Pleas," Federal Probation, Volume 49 (June 1985)


Shawn Bushway and Anne Piehl. “The Social Science Contribution to the Policy Debate


Terence Miethe and Charles Moore, "Socioeconomic Disparities under Determinate Sentencing

Frank W. Miller, *Prosecution: The Decision to Charge a Suspect with a Crime*, edited by Frank 

Ojmarrh Mitchell, "A Meta-Analysis of Race and Sentencing Research: Explaining the 

John Monahan, "The Prediction of Violent Criminal Behavior: A Methodological Critique and 
Prospectus," in *Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on 
Crime Rates*, edited by Alfred Blumstein, Jacqueline Cohen, and Daniel Nagin (Washington, 
DC: National Academy of Sciences, 1978)

Norval Morris and Michael Tonry, *Between Prison and Probation: Intermediate Punishments in 


Ilene Nagel and Stephen Schulhofer, “A Tale of Three Cities: An Empirical Study of Charging 

edited by Alfred Blumstein, Jacqueline Cohen, Susan E. Martin, and Michael H. Tonry 

Donald J. Newman, *Conviction: The Determination of Guilt or Innocence without Trial*, edited 

Lloyd E. Ohlin, "Surveying Discretion by Criminal Justice Decision Makers," in Lloyd E. Ohlin 
and Frank J. Remington, *Discretion in Criminal Justice: The Tension Between Individualization 
and Uniformity* (State University of New York Press, 1993)

Lloyd E. Ohlin and Frank J. Remington, *Discretion in Criminal Justice: The Tension Between 
Individualization and Uniformity* (State University of New York Press, 1993)


Joan Petersilia, "Measuring the Performance of Community Corrections," in *Performance 
Measures for the Criminal Justice System* (Washington, DC: Bureau of Justice Statistics, 1993),
NCJ-143505


Joan Petersilia, Allan Abrahamse, and James Q. Wilson, *Police Performance and Case Attrition* (Santa Monica: Rand Corporation, 1990)


Frank J. Remington, "The Decision to Charge, the Decision to Convict on a Plea of Guilty, and


Max Weber, *Politics as a Vocation* (lecture delivered in Munich, January 1919)

