Justice Quarterly
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rjqy20

Studying Discretion in the Processes that Generate Criminal Justice Sanctions
Shawn D. Bushway & Brian Forst
Published online: 24 May 2012.

To cite this article: Shawn D. Bushway & Brian Forst (2013) Studying Discretion in the Processes that Generate Criminal Justice Sanctions, Justice Quarterly, 30:2, 199-222, DOI: 10.1080/07418825.2012.682604
To link to this article: http://dx.doi.org/10.1080/07418825.2012.682604

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the “Content”) contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions
This paper presents a typology of discretion that differs from the standard typology of discretion in the legal literature: Type A discretion, which is the generally recognized discretion of individual actors to make decisions within a set of laws and rules, and Type B discretion, which is the crafting of laws and setting of rules in the first place. An analysis using aggregate data from 1992 to 2002 points to the prosecution stage as the primary source of Type A discretion that contributed to the massive increase in incarceration during this period. To understand how this could happen, this paper argues that actions need to be linked to outcomes. We identify three key outcomes of the sanctioning process—crime control, justice (guilt/innocence), racial disparity (fairness)—that deserve more attention from social scientists.

Keywords sentencing; discretion; criminal justice system; prosecutors; crime reduction

Shawn D. Bushway is a professor of criminal justice and Public Administration and Policy at the University at Albany (SUNY). He received his PhD in Public Policy Analysis and Political Economy in 1996 from the Heinz School of Public Policy and Management at Carnegie Mellon University. His current research focuses on the process of desistance, the impact of a criminal history on subsequent outcomes, and the distribution of discretion in the criminal justice sentencing process. Brian Forst is a professor of justice, law, and society at American University. He joined the AU faculty in 1992, following three years on the faculty at George Washington University. Before that, he was the director of research at the Institute for Law and Social Research (1977-1985) and the Police Foundation (1985-1989). Recent books include Terrorism, crime and public policy (Cambridge University Press, 2009), Errors of justice: nature, sources and remedies (Cambridge, 2004), and Criminologists on terrorism and homeland security, co-edited with Jack Greene and James Lynch (Cambridge, 2011). His primary research interests include prosecution and sentencing, terrorism and counterterrorism policy, discretion, and legitimacy. Correspondence to: S. Bushway, Professor of Criminal Justice, Professor of Public Administration and Policy, School of Criminal Justice, University at Albany (SUNY), Albany, NY, USA. E-mail: SBushway@uamail.albany.edu

A prior version of the paper was titled “Studying Discretion in Sentencing.” The authors subscribe to the notion, first described in National Academy of Sciences (1983), that sentencing is not just the stage at which judges assign punishment, but a product of all the steps along the way that lead to the ultimate punishment experienced by the offender, including indictment and parole decisions. Indeterminate sentencing, for example, includes the parole board by definition. However, this broad description of sentencing has not been adopted in the criminal justice literature, which has largely studied sentencing using conviction data-sets.
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.¹ The founding fathers gave us organizing principles and ideals, not detailed blueprints for governance and bright lines for policy. This lack of clear guidance has been a recurring source of tension in matters in the criminal and juvenile justice systems regarding the proper use of discretion. Lack of clear structures has also created problems for criminal justice/criminological researchers seeking to study the impact of discretion on the people being processed by the criminal justice system. This paper will focus on identifying productive new directions for research that can link discretion by key actors to important criminal justice outcomes, like crime levels, justice, and racial disparity in the distribution of punishment.

Our criminological interest in outcomes can clearly supplement, complement, and make use of the legal focus on process. Much legal scholarship on discretion focuses on the extent of discretion made available in any given case by a law or rule. Some legal scholarship also focuses on how rules affect discretion within the system itself, such as the research that looks at the hydraulic displacement between actors (Reitz, 1998). Although there are exceptions, the legal focus is on the law/rule itself (and the discretion it makes available) rather than the outcomes that the laws (and the incumbent discretion) are intended to produce. In criminology, there is a similar complementarity in evaluations of process vs. outcomes. Process evaluations focus on documenting the process involved in delivering a given treatment, while outcome evaluations focus almost solely on whether the treatment group differs from the control group. Both types of analyses are important—a significant and positive finding in an outcome analysis is more valuable, offering a clear understanding of the process involved in the treatment. In the same way, a process evaluation showing that the treatment was delivered as intended is more valuable in the presence of a specific outcome.

Because of our focus on outcomes in this paper, we conceptualize discretion differently from most legal scholarship. The most authoritative legal-theoretic account of discretion has been advanced by Dworkin (1977), who distinguishes 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 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

¹. The idea of rule of law precedes John Adams by more than two millennia. Plato wrote, in around 350BC, that “if law is the master ... and the government is its slave, then the situation is full of promise and men enjoy all the blessings that the gods shower on a state.”
uncontroversial business that comes with the job. This is the type of discretion that most scholars consider when discussing discretion in the criminal justice system.

In what follows, we talk about two kinds of discretion—the weak discretion (Type A) that individual actors have to make decisions with variation given a set of rules, and the discretion (Type B) that legislators and criminal justice policy-makers have to establish a set of rules. Given that this discretion itself is governed by rules, our Type B discretion differs from Dworkin’s “strong” discretion, which refers to an official’s making a decision that goes beyond the boundaries of ordinary discretion, often nullifying a law perceived as wrong if applied in a particular situation. The choice by legislators to impose sentencing guidelines is an act of Type B discretion—the legally allowable choice by judges of sentences within the sentencing guideline ranges is an act of “weak” or Type A discretion. Type B discretion, the ability to create rules and policies, can be used to limit and shape Type A discretion, and there is often a tension between the rules set by actors (Type B) and the discretion available to lower-level actors within those rules (Type A).

Most discussion in the literature deals with Type A discretion, but as we will discuss later, we believe Type B discretion deserves more direct study in criminal justice. The states in the USA clearly have a great deal of Type B discretion to set sentencing rules that fall within the general guidelines of constitutionality, and the exercise of this discretion by legislators has led to a huge variation across states in the basic structure of sentencing. This variation is potentially more influential than variation in Type A discretion of individual actors within any given set of rules.

In the next section, we identify the key decision nodes at which discretion is exercised, and we present an analysis that reveals the prosecution stage as the main place that accounts for the increase in incarceration levels. This analysis is based on an extraordinary decade, 1992-2002, when crime rates dropped precipitously as prison populations soared. We then look at the link between the use of rule-setting discretion and three outcomes: racial disparity, crime control, and justice. We conclude with concrete suggestions for further research.

**Identifying Sources of Change in the Decision to Incarcerate**

We start by asking where in the system discretion is exercised in a way that most affects outcomes that we as a society care about. One such outcome is the proportion of people in prison or jail, commonly referred to as the “incarceration rate.” During the last 30 years, the USA has experienced a massive

2. As we discuss below, we find it useful to distinguish the flow variable, proportion of convicted persons incarcerated, from the stock variable, the percentage of the population in prison or jail. It might make better sense to regard the former as the “incarceration rate” and the latter as the “proportion of the population incarcerated.”
increase in this statistic. In the sentencing symposium that gave rise to this article, Sabol (2010) presented a paper documenting a consensus that the increase in incarcerations is the result of changes in behavior by policy-makers, rather than an increase in crime or other exogenous factor. In all likelihood, Type B discretion was exercised to change the rules, which led directly to an increase in prison sentences. Ironically and problematically, however, researchers do not know which policy changes have contributed most to this increase in incarceration rates.

In what follows, we analyze the relative importance of discretion exercised at each stage in terms of its impact on the aggregate rate of incarceration. Although crude, this method can be used to highlight parts of the system that might account for changes in the incarceration rate.

Specifically, we show that the aggregate proportion of the population in prison or jail can be estimated as a multiplicative series of factors that reflect each of the following 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.  
- The rate at which offenses result in 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 & Lynch, 1997). 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

3. 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.

4. Consider the identity: Inmates/Population = (Victimizations/Population) × (Reported Crimes/Victimizations) × (Arrests/Reported Crimes) × (Convictions/Arrests) × (Incarcerations/Convictions) × (Average term/Incarcerations). In a steady state, the number of inmates will equal the number of incarcerations annually times the average term per incarceration.
criminal justice system to the change. Analysis of this sort can provide indicators of changes in the exercise of discretion by police, prosecutors, judges, and legislators.

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 recently entering tend to be serving longer terms than those already there). 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 available on an aggregate, but not on an original offense basis.

Recognizing these limitations, we 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 legislators to the increase in the aggregate incarceration rate. 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 US population (41%). 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 actually declined for drug offenses during this 10-year period, while the aggregate arrest and conviction rates remained fairly constant. The increase in the number of inmates is in fact due primarily to tougher sanctions and larger pools of offenders in the 1980s, with smaller pro-

5. For the period 1992-2002, 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 over the ten year period. 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%.
6. The results reported here are from the Bureau of Justice Statistics National Judicial Reporting Program, reported in its Felony Sentences in State Courts bulletins and the BJS State Court Processing Statistics, reported in its Felony Defendants in Large Urban Counties series.
7. 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 US population was 255 million in 1992 and 290 million in 2002.
8. 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 US 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.
portions of those cohorts exiting prison from 1992 to 2002 than in earlier periods, which more than offset the decline in new incarcerations during the period. Police and prosecutors made no attempt to adjust by reducing arrest and conviction rates. Policy-makers who wish to reduce the incarceration of drug offenders today need to find ways to change the behavior of police and prosecutors rather than judges, especially in urban areas and jurisdictions bound by severe drug laws. It would help as well for legislators to reform overly punitive drug sentencing laws, as we suggest below.

Evaluating Type B Discretion

In non-justice settings, the exercise of discretion is generally a by-product of a larger inquiry that assesses how the rules that shape discretion (Type B) and the exercise of discretion within that set of rules (Type A) allows organizations and individuals to obtain desired goals. Discretion is thus not analyzed in a vacuum, but relative to its ability to achieve the desired outcomes.

Accordingly, 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—sentences that achieve their desired goals—and regarding departures from those sentences as errors, and using principles of error management generally to inform the exercise of discretion (Bibas, Schanzenbach, & Tiller, 2009). The idea of misguided or misplaced discretion (both Type A and Type B) is a relative one—we must first have a sense of what "should" happen under the "preferred" system. Some examples of problem domains that call for a similar balancing of rules and Type A discretion to deal with conditions of complexity and uncertainty in matters that can impose significant harms on individuals are: quality control management (Evans & Lindsay, 2004), signal detection applications in clinical diagnosis and treatment (Akay, 1996), control of monetary policy (McCallum, 2007), and financial portfolio management (Fabozzi & Drake, 2010). In each case, the assessment of how well the rules work to accomplish the goals of the system is a key element of the assessment of discretion.

It is a truism in criminology that the goals of sentencing can lead to differences in how discretion is exercised. Indeed, one of the major sources of variation in judges' sentences is 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 & Wellford, 1981; Gottfredson, 1999; National Academy of Sciences, 1983). The freedom of judges to choose among these sentencing goals has been found to contribute to variation in the exercise of sentencing discretion, along with tendencies for some judges to be more consistent in their sentencing practices than others, regardless of their preference for one goal or another (Forst & Wellford, 1981).
Discretion thus cannot be evaluated without first deciding what goals are to be achieved. In a system without an agreed-upon set of goals, it is easier to keep the focus on the process rather than outcomes—but the social scientific insight is that discretion cannot be fully evaluated without some sense of what it is that the system is trying to achieve. Having agreed that the system has goals, we can evaluate the impact of the various rules to achieve these goals, which should guide policy-makers in their use of Type B discretion when choosing among rules and policies. We can also evaluate how the use of Type A discretion works either to help or hinder the accomplishment of these goals.

While we are aware that this approach will be problematic to someone who focuses solely on process, we note that others have begun to introduce the idea that a focus on outcomes is needed when discussing aspects of sentencing such as incarceration. For example, Weisberg and Petersilia (2010) have suggested that discussions about mass incarceration should consider the idea of “unnecessary incarceration.” This idea carries with it the idea that some incarceration is “necessary” or desirable. In their essay, they warn that a movement against incarceration that does not focus on the outcomes that sentencing is intended to achieve could, and probably would, have undesirable outcomes. Outcomes and goals give meaningful and important shape to a discussion of discretion.

The last 30 years have seen dramatic changes in sentencing policy intent on achieving a variety of goals, including disparity reduction and increases in punishment. Examples of policy changes include determinate sentencing guidelines, mandatory minimum statutes for some crimes, minimum time-served requirements, and three-strike laws (Stemen, Rengifo, & Wilson, 2005), and the increased use of formal risk assessment at the parole and probation stages (Harcourt, 2007).

However, research on these discretionary policy changes is limited. 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 to answer. 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

9. In the most recent American Society of Criminology meetings in Washington DC, Richard Rosenfeld spoke on a Presidential Panel Discussion entitled “Innovations in Criminal Justice Policy and Practice” about the need to evaluate policies based on what they were trying to accomplish, rather than some external objective goal unrelated to the goals of the policy. One example of this type of disconnect includes assertions that the federal sentencing guidelines fail to reduce crime. Given that the policy was not created to reduce crime, this seems like an odd criticism. We endorse this linkage of evaluation to the goals of the policy.
threat) that still emphasize subjective decision-making as the central causal mechanism” (p. 333).

Outcome 1: Racial Disparity in Sentencing Outcomes

Perhaps the most productive area of policy research involves the study of guidelines and their ability to reduce racial disparity. Some scholars believe that the presumptive sentencing guidelines as practiced in states like Minnesota and Washington have accomplished their goal of reducing disparity (Frase, 2005; Reitz, 2005; Tonry, 1996). 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 & 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). Moreover, the few before-and-after studies that do exist show either no differences or small differences after guidelines, in part because the disparity was not large before the guidelines were implemented (Miethe & Moore, 1985; Wooldredge, 2009). One possible exception is the recent update by the US 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 sentencing rules.

Because we are interested in disparity system-wide, 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.

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.

Theory can also be useful in guiding research into disparity by providing a hypothesis about what should happen under a reasonable set of assumptions. Alfred Blumstein’s simple thought experiment involving racial discrimination in incarceration is an excellent example. Blumstein (1982, 1993) 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 equal the proportion of blacks in the incarcerated population, controlling for the differential rates of incarceration per crime. He can explain the vast majority of racial disparity in incarceration for serious crimes using this simple model, but can explain only 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, Spohn, and Wodahl (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.10

Another productive avenue for research on disparity involves looking at guidelines or statutes themselves as a source of disparity. The “crack cocaine” enhancement in the federal system, which treats crack cocaine more harshly than powder cocaine (with dramatic implications for race), is an excellent example of how this can happen (The Sentencing Project, 2009). Researchers now routinely use the presumptive sentence of the guidelines as a control (Engen & Gainey, 2000; US Sentencing Commission, 2010) in models of disparity, but do not study the guidelines themselves. Yet, guidelines can create disparity, especially since the impact of any given aggravating or 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.

Bushway and Piehl (2011) present the first detailed analysis of a sentencing guideline grid, focusing not on the actual sentences imposed, but instead on the recommended sentences. They identify the impact of each case factor underlying the scores that produce the recommended sentence. Their most basic insight is that an additional point from criminal history or crime severity results in a different increase in the recommended sentence length depending on the other factors that are present. Because the distribution of factors is not random, and some factors tend to only be present when a number of other

---

10. 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 (Ayres & Waldfogel, 1994; Knowles, Persico, & Todd, 2001).
factors are present, the average impact of each factor on the recommended sentence varies also. This can impact racial disparity, because blacks and whites tend to have different “packages” of factors, which implies that blacks and whites tend to “live” in different areas of the grid.

One implication is that seemingly modest rule changes could lead to observable shifts in the overrepresentation of African-Americans in prison. In Maryland, factors that are correlated with race such as weapons use and the presence of a vulnerable victim contribute the most to racial disparity in punishment.

Bushway and Piehl (2011) also highlight the value of moving to a study of guidelines in addition to sentencing outcomes. Using the same data, Bushway and Piehl (2001) found that the judges appeared to be introducing additional racial disparity when they deviated from the guidelines. Judges are going far above guidelines on factors like juvenile history that are particularly prevalent for African-Americans, and they are going far below the guidelines on factors like vulnerable victims, which were more prevalent for whites. A more explicit understanding of what the guidelines are doing directly (Type B discretion) helps shed light on what other actors are doing in response (Type A discretion).

Outcome 2: Crime

It would be a mistake to focus only on the punishment levels themselves; policy-makers desire to influence other outcomes, such as crime. Legal scholars can argue about the validity of such goals; however, as long as policy-makers consider crime control as a legitimate goal of sentencing, it bears exploring empirically.

The statistical analysis of laws intended to reduce crime, such as three-strike laws or gun control laws, look at net effects (after the system actors use Type A discretion to respond) of policies on desired behavior. In these types of models, the impact of the law on crime is examined 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 & Miles, 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 (Webster, Doob, & Zimring, 2006). 11

11. We applaud those who hold up a high standard of proof for causal research on the impact of sentencing policy on crime. However, we caution researchers to use the same standards when evaluating all causal work on sentencing policies, including the impact of incarceration on future life outcomes for the individual and his/her family. Individuals who are opposed to the idea of using sentencing policies to prevent crime on philosophical grounds should state this opposition up front—empirical work on the causal impact of the policies on crime is clearly irrelevant for these individuals.
Another potentially powerful approach to 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. Lee and McCrary (2009) and Hjalmarsson (2009) both use individual data to examine the impact on behavior of moving from the juvenile justice system to the adult system. 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 focus on the differential behavior of those who are subject to three strikes laws in California (Helland & Tabarrok, 2007) offer what is probably the best available evidence on the ability of these laws to produce desired outcomes. More effort should be focused on collecting individual-level data that link sentencing data with data on criminal behavior (e.g. arrests).

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 integrate the state police arrest data repository with a correctional data-set that includes all convictions. This setup allows researchers to analyze the impact of laws such as those that defer conviction (and avoid labeling) on subsequent criminal behavior (Chiricos, Barrick, Bales, & Bontrager, 2007). Chiricos et al. (2007) study is the rare study in criminology which explicitly links a particular sentencing statute to crime/recidivism. While some researchers may be opposed to this linkage (which we acknowledge is novel in criminology, although not in economics), we believe that papers such as the Chiricos et al. (2007) study highlight both the feasibility and desirability of bringing the study of crime together with the study of Type B discretion in criminology.

Before closing this section, it is important to acknowledge that not all Type B 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 formal and informal policy about how to handle certain types of cases. Moreover, these actors 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, Ulmer, & Kramer, 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 & Johnson, 2004). Although there is undoubtedly meaningful variation at this level in sentencing.

---

12. 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 discussed above.
data-sets, efforts to explain this variation with observable contextual variables have been largely unsuccessful.

We are also uncomfortable relying on empirical models to differentiate between individual discretion and discretion at the group or policy level. Tests for how a particular set of actors are acting require a theoretically relevant model for that behavior. Only then will the empirical model capture interesting variation that can be explained. We believe that it is useful to develop theoretical models, even simple ones that can create accounts of what should happen at this level under reasonable assumptions. If the observed means approximately equal the predicted or expected means, we will 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) to the study of 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). 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 interesting theory-based model. In this model, the researcher in effect argues that judges are deviating from the recommendations—meaning that the recommendations become the rule from which individuals deviate. The contribution comes from the willingness of Engen and Gainey to specify a model of behavior that allows the researcher to distinguish the racial impact of the guidelines from the impact of the judges as a group.

Outcome 3: Justice

Articles on discretion rarely if ever consider the impacts of sentencing rules and process on errors in the decisions about innocence and guilt. Yet clearly, policies and individual decisions by the police, prosecutors, defense attorney, and judges can affect the ability to convict culpable offenders and acquit innocent defendants. The reality is that any system of justice that we can imagine will have errors, even when it is functioning perfectly. In fact, many introductory statistics text books use the jury trial model as a way of communicating the two types of errors in a hypothesis test.

However, in both criminology and legal scholarship, it is fashionable to label all errors, particularly conviction of innocents, as "miscarriages of justice," as if something went wrong to produce this result. Although some of these mistakes are the result of malfeasance of agents, most of them are the natural result of a system of justice that relies on probabilistic rules of conviction. "Beyond a reasonable doubt" is a probabilistic, not a deterministic, standard. Any system will have natural variation—in the case of the criminal justice system, this variation shows up in the discrete decision of guilt or innocence.
Criminologists have long recognized that the evaluation of any prediction exercise requires an explicit recognition of both false positives and false negatives that will occur under any given decision rule (Gottfredson & Moriarty, 2006). The guilt and innocence outcomes of criminal justice decision processes follow the same basic framework of other decisions, and those processes should be exposed to the same standard of evaluation.

In our view, discussions about the system and discretion in the system must be able to weigh the relative costs of these different errors, often categorized as false positives (arrests and convictions of the innocent), and false negatives (failures to convict culpable offenders). In any given system, these two types of errors tend to be hydraulically linked—the way to decrease one is usually to accept an increase in the other. Although policy-makers continually search for the new technology and methods that will allow them to decrease both errors simultaneously (e.g. through DNA testing, improved methods of interrogation), both types of errors will exist in any system. Social science can and should be used to study how different rules can affect both the baseline levels and the tradeoffs between these errors.

Part of this consideration can involve explicit assessments of the costs of the different errors, both tangible and intangible. In the heated discussions about selective incapacitation (Gottfredson & Gottfredson, 1994; Monahan, 1978), scholars rightfully pointed out that the "rubber hit the road" when it came down to specifying the model and setting the cut-point 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 (Gottfredson & Gottfredson, 1994; Gottfredson & Moriarty, 2006).

As Berk (2009) makes clear, policy-makers are comfortable with the notion that not all prediction errors are equal. And, in most discussions about decisions about guilt and innocence, policy-makers and legal scholars routinely accept that the ratio of the respective costs of false positives to false negatives is not one. Consider Blackstone’s (1769/1979) 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 USA, this ratio varies from 1 to 1, to 100 to 1 (in Oklahoma).

As a first approximation, we start with the observation that well over 10 million felony offenses involving adult offenders occur annually that do not end in conviction (false negatives). Wrongful convictions (false positives) have been estimated at between 0.5 and 1% of all felony convictions, which amounts to between 5,000 and 10,000 annually (Forst, 2004; Huff, Rattner, & Sagarin, 1996). This gives a ratio of 1,500-3,000 false negatives to every innocent person convicted. Bushway (2011) has conducted a more systematic analysis of homicide cases using case-level data in Chicago. He reports a ratio of 66 false negatives for every false positive for homicide cases. It is not surprising that this ratio is lower for homicides, but it is worth asking if either ratio
is normatively acceptable. A careful cost-benefit analysis of the relative costs and benefits of false positives and false negatives by offense category might help policy-makers decide if the current laws are producing outcomes that we as a society find consistent with our a priori beliefs about what should happen. If the cost of a single false positive is infinite, then clearly we can tolerate more false negatives to eliminate some false positives. Indeed, in this case, we would actually not tolerate convicting anyone at all, because of the extreme costs of falsely convicting anyone.

False positives, while costly, are not infinitely costly under any reasonable cost calculus, and the social costs of false negatives require consideration too. We advocate for simpler 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 believe that a social scientific approach to this question will avoid focusing only on wrongful convictions without consideration of the full range of other errors (including wrongful arrests) and their costs. Any practical system of justice will make legitimate errors even when everything works perfectly. Intelligent policy-making starts from an acceptance of this fact, but it also requires that policy-makers know what is actually happening in the system and the relative costs of the outcomes.

Formal theoretical models can help generate useful insight about the tension between multiple goals in the criminal justice system. Bjerk (2007) created a simple model that showed that false positives will increase in a world where people threatened with severe sentences can plead guilty to lesser sentences (see also Ulmer, Eisenstein, & Johnson, 2010). While not easily estimable, these models might generate useful stylized facts that can be compared with observed outcomes (Forst & Brosi, 1977; Landes, 1971).

Identifying Abuses of Type A Discretion

Perhaps the most serious concern about Type A discretion is the potential for agents of the justice system to abuse it. 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 & 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).

13. An anonymous reviewer raised reasonable questions about our approach to generating these estimates. We are unaware of other estimates of false negatives, and we encourage an active discussion about the best way to estimate false negatives as a productive line of future social science research.
Abuses of prosecutorial discretion are less well known, but equally important as abuses by police, because of the inherent power of the prosecutor. The more common problem in prosecution is the lack of transparency that renders the routine exercise of discretion unknown and unreviewable. As elected officials in most jurisdictions, prosecutors are publicly accountable for their routine decisions, but available evidence provides virtually no basis for knowing how well the prosecutor in any community fares in exercising this authority. The lack of scrutiny given to abuses in run-of-the-mill felonies is a major shortcoming in the study of discretion.

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 individuals who appear to be using their discretion abusively, or at the 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, 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 public sector agents performing public roles. The risk of abuse cannot be ignored; still, basic standards of public accountability call for movement towards the release of individual identifiers in these administrative datasets.14

Once this data is released, we need to understand how to identify individual discretion in the data. The biggest problem is distinguishing between actions of the individual and actions of the group, as prescribed by rules. 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 Ridgeway and MacDonald (2009). They use brute force statistical methods that benchmark “standard police stopping behavior” for comparison purposes so as to identify outliers. Police who were flagged as outliers were then investigated by New York City to determine if they were in fact engaged in questionable discretionary stopping behavior. Forst, Leahy, Shirhall, Bartolomeo, and Wish (1982) conducted a similar analysis using predicted conviction rates for each officer’s case mix in the cities of New York

14. See also Schanzenbach and Tiller (2008) for a similar call.
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 multiple witnesses and conduct follow-up investigations. Many components of the criminal justice system have the kind of large data-sets that would support this style of analysis.15

Although researchers do not often use this approach as a way of studying 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 at the individual level.

Spohn and Fornango (2009) used this approach to study prosecutors. They found a non-trivial amount of unexplained variation at this level, which could 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 it correlated with political affiliation, consistent with Posner’s (2008) assertion that the individual beliefs are the key driver of judge discretion. Equally interesting, they find that the discretion is modified or constrained by the nature of appellate review.

An alternative approach is to use the random assignment of cases to identify the impact of individual actors on subsequent outcomes. In these studies, random assignment is used to identify differences in sentence lengths, which is then used to study specific deterrence (Green & Winik, 2010) or the impact of sentence length on employment (Kling, 2006). With this framework, researchers can simultaneously describe the impact of Type A discretion while also looking at its impact on outcomes. While there is substantial evidence to show that individual judges exercise meaningful discretion in the assignment of penalties, there is little evidence that the variation due to judges has any impact on the criminal behavior of defendants.

One problem with the approach of using random assignment to the judges as instrumental variation is that the variation itself is not modeled. Judges can sentence people to different sanctions for many reasons, including their own judicial philosophy of punishment. Gottfredson (1999) collected data on

15. An anonymous reviewer raised the issue that researchers are beholden to the system for access to the data. This makes them understandably unwilling to criticize the very actors who granted access to the data. We agree that this is a real issue, but hope that over time situations are created where access to the data is not limited to those who must necessarily curry favor with the actors who are to be analyzed in the data.
judicial preferences regarding punishment, and showed how they mapped to actual sentencing practice (see also Forst & Wellford, 1981). 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 mechanisms by which sentences are handed down by these individuals. This point was made clear in an important but often overlooked paper by Manski and Nagin (1998), which identifies the limitations of reduced-form treatment effect models. Unless the researcher can model the decision-making process, they will not identify the treatment impact of sentence length (or any other outcome) on subsequent behavior.¹⁶

Modeling decision-making behavior can also help identify possible abuses of discretion. Levitt and Duggan (2002) presented an analysis of corruption in sumo wrestling that is based on reasonable models of wrestler incentives. Jacob and 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 more likely to be cheating. In a similar vein, political scientists Huber and 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 use and abuse 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. Researchers of discretion and decision-making must propose models of behavior, and then work to identify data that can be used to identify this behavior with as much structure as possible.

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 policy-makers manage and control discretion. In this final section, we try to clarify what we think are key tasks for making progress on the study of both Type A and Type B discretion.

First, researchers should analyze the ability of policy-makers at all levels to affect outcomes like levels of punishment, crime, and errors. The action of any given actor will be affected by other actors, but we believe it is important

¹⁶. For a more general treatment of this problem, with a compelling and interesting example, see a paper on peer effects by Carrell, Fullerton, and West (2009).
to identify whether the purposeful use of Type B discretion by policy-makers 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. This starts as an aggregate or system-wide question, and it should be examined using data on all people being acted on by the system (i.e. not conviction data), if possible. Showing that guidelines reduce disparity at the conviction stage, without considering other stages, is insufficient for analyses of the overall impact of a policy. Simulated comparisons of what should have happened if the policy were 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, Johnson, & Slocum, 2007).

Second, 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, Johnson, Wistrich, & Guthrie, 2009), but it can also be done using administrative data-sets, 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 (Spoth & 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 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 restricting 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.

Our vote for the area most in need of illumination goes to prosecutorial discretion (see also Ulmer, 2012). 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
US population. The increase has been widely attributed to increases in drug incarcerations, the cases that Blumstein (1993) has identified as having the most discretion. For the urban counties we sampled, both the incarceration rate and average term of sentence actually declined for drug offenses during this 10-year period, while the aggregate arrest and conviction rates remained fairly constant. This suggests that a true understanding of this problem needs to focus on the discretion exercised by police and prosecutors rather than by judges. Barkow (2008) has 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 and mercy. 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 20 years following 1973, when aggregate information about arrest rejections and case dismissals, pleas and trials were reported by jurisdictions across the US. This practice 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 should 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 where researchers could start. The methods of researchers in the field of psychology and law, which so far have focused mostly on trial outcomes, might also be usefully applied to the actions of prosecutors and others involved in the plea bargain decision.

Third, analyses should explicitly test for the interactions among actors in the system. Upstream actors can constrain downstream actors, but downstream actors can also temper those upstream. For example, the main theory of plea bargaining is referred to as “bargaining in the shadow of the trial.” Bibas (2004) has raised important concerns about this theory, but there are few direct tests of this type of model. Models of trial outcomes offer a set of expectations that we can bring to plea bargaining behavior.

Fourth, 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 responsible for given cases. This is particularly valuable when we have multiple cases per actor. Such data can help researchers identify actors who are either systematically misusing or abusing their discretion. Interrogation of data to identify outliers prospectively would be a positive use of data for policy-makers. Criminal justice agents should be held fully accountable for their exercise of discretion: rewarded for using it well, corrected for using it poorly, and sanc-

17. See Note 1, supra.
tioned for abusing their authority to exercise it prudently. Researchers can help practitioners to set 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 with respect to outcomes of interest. 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.

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 set boundaries on the possible set of choices available to policy-makers. Those processes make many of these analytic ventures possible by providing data on variation that can be studied.

The often competing motivations of the actors in the system must be a central feature of models of discretion. Better understanding of these motivations and processes will allow researchers to better understand discretion, and participate more fully in the ongoing debate about the management of discretion in the criminal justice system.

Acknowledgements

We are grateful to Gail Hsiao and Brendan Lantz for research assistance, and two anonymous reviewers for helpful suggestions. This paper was originally prepared for the Conference on The Past and Future of Empirical Sentencing Research, 23-24 September 2010, Albany, supported by a grant from the National Science Foundation (SES-0939099). We are also grateful for the three discussants, Kim Hunt, Nancy King, and Max Schanzenbach, at the Symposium who provided useful comments.

References


