

## Looking Beneath the Surface

### Reaction Essay to “Discretion, Rule of Law, and Rationality”

Kim Steven Hunt<sup>1</sup>

Forst and Bushway (2010) argue that criminal justice policy choices and tools (laws, regulations and programs) can create systematic departures from optimal choices (in this case, ineffective or unfair punishments), and much of this dysfunction is the result of failures to account for and control discretion. Failures to effectively manage discretion can create both random error and systematic bias with serious social consequences. The essay then presents a variety of ways of thinking about discretion and ideas for ways to study discretion in hopes that a better understanding will lead to better policy choices. The authors do an outstanding job of reviewing the literature on decision making as it relates to discretion and research programs that may advance understanding in this area.

Rules governing sentencing policy, as well as discretion, can increase social cost (Forst and Bushway, 2010: 9), and the authors discuss of the ebb and flow of rulemaking and discretion and consequent social costs of practices such as sentencing guidelines. In this brief essay, I wish to take the discussion of sentencing guidelines, surely one of the key rulemaking innovations in this field in recent decades, deeper. Forst and Bushway (2010) argue, correctly, that the Minnesota sentencing guidelines became the dominant model for creation of other U.S. guideline systems. They do little to explain how sentencing guidelines affects errors in social costs such as punishments, except to say that the Minnesota guidelines appear to have been more successful in

---

<sup>1</sup> The views expressed in this essay are the author’s alone and do not reflect the policies or goals of the United States Sentencing Commission.

reducing errors associated with overly severe punishments as compared to the U.S. federal sentencing guidelines (Forst and Bushway, 2010: 10). I would like to use this observation as a starting point. My goal is not to contest this conclusion, but to suggest a foundation for strengthening policy conclusions using policy evaluation techniques.

One aspect of looking beneath the surface of rulemaking and discretion in complex systems is to examine what constitutes a policy treatment or mechanism such as “sentencing guidelines” before drawing conclusions about outcomes of those policies. The authors note but do not attempt to separate the multiple policy “treatments” that went into effect simultaneously or in rapid succession with the passage of sentencing reform legislation in Minnesota, the federal system, and elsewhere; treatments which include not only sentencing guidelines, but also the change from indeterminate to determinate sentencing, the expansion of mandatory minimum sentences, legislative directives impacting guideline structure and function, etc.

The authors are to be commended for their strong support for a research program that focuses on a thorough understanding of processes, and formal theory-driven models that “can help generate useful insight about the tension between the multiple goals in the criminal justice system” (Forst & Bushway, 2010, 27) and the roles and motives of individual actors and institutions. In this essay I wish to add complementary thoughts also founded on process and theory-driven models. First, this reaction essay will catalogue some problems inherent in unpacking the outcomes of complex policy initiatives such as sentencing reforms, emphasizing the interplay of discretion with context, program mechanisms, and theories of change using the literature of policy evaluation, in particular theory-driven realist evaluation (Pawson, 2006). Next an aggregate data exercise is presented which attempts to begin to unpack the impact of

sentencing guidelines on policy outcomes by reanalyzing a recent national study of sentencing guidelines (Shepherd, 2007; Hunt & Sridharan, 2010) and suggestions for next steps.

### **Assessing Impacts of Complex Policies and Programs**

Forst and Bushway note that the last 30 years have seen a substantial increase in discretion exercised by at least some actors through policies such as determinate sentencing, sentencing guidelines, mandatory minimum statutes, minimum time served requirements in indeterminate sentencing venues, and three-strikes laws. Engen (2009) similarly notes that research must uncover how judges and prosecutors exercise discretion under various types of systems (policy tools such as presumptive v. voluntary, offense-based v. grid-based guidelines, etc.)

The proliferation of sentencing policy changes and attending changes in discretion exercised by various actors represents a problem for researchers attempting to understand the policy impact of a given policy or program, as Forst and Bushway emphasize. Approximately 25 state and federal policy initiatives have emerged under the label sentencing guidelines in the United States alone,<sup>2</sup> broadly conceived as providing substantial advantages over the prevailing status quo which was and continues to be indeterminate sentencing (American Law Institute, 2010). Underlying the term “sentencing guidelines” in the United States are systems primarily organized around grids,<sup>3</sup> with different rules regarding the application of mandatory minimums

---

<sup>2</sup> The term sentencing guidelines has also been used in England and Wales, Scotland, South Korea, China, New Zealand, and Australia to describe sentencing policy initiatives. These initiatives have generally taken a different format than the U.S. initiatives, further adding to the diversity and complexity.

<sup>3</sup> Ohio and Delaware guidelines represent notable exceptions.

within a guideline structure;<sup>4</sup> judicial discretion (presumptive or voluntary); how offenses and offenders are characterized with respect to things such as weighting the seriousness of offenses and offenders (criminal history categories and whether to count certain crimes such as juvenile adjudications); and whether the grids are regarded as prescriptive, in ways that intended to rationalize decisions with respect to punishment and other sentencing goals, or regarded as descriptive of prevailing goals, intended merely to better manage existing discretion.

Despite this variety, or perhaps because of it and its attending complexity, there have been few comparative studies assessing the pathways by which different sentencing policies impact key policy outcomes (Bushway and Piehl, 2007). In part, this scarcity may result from the fact that sentencing guidelines are relatively new (approximately 30 years old) as policy tools go. However, as a result it is “neither obvious nor easy to establish” a common frame of reference between these initiatives, given this diversity (Ostrom, et al., 2008, p. 5).

In this response essay I argue that the field of policy evaluation provides useful ideas for approaching these conceptual problems. The question of the impact of such policies falls within the realm of evaluation, and there have been few policy evaluations, either single venue or comparative, of the impact of sentencing reform efforts on optimal choices for public safety (Tonry, 1996; Bushway & Piehl, 2007; Engen, 2009).<sup>5</sup> The literature is particularly weak with respect to understanding the contexts and mechanisms that facilitate certain desired outcomes, which is a strength of the realist evaluation literature (Pawson and Tilley, 1997, Pawson, 2006, Pawson and Sridharan, 2009).

---

<sup>4</sup> Guideline ranges may be adjusted to accommodate mandatory minimums or not. If ranges are not adjusted and ranges and minimums conflict, then mandatory minimums trump the guidelines.

<sup>5</sup> An exception is the larger literature on the impact of sentencing reform on unwarranted disparity.

There is little discussion in the sentencing policy literature of the underlying program theories by which sentencing guidelines can impact long term outcomes such as incarceration levels, costs, or crime levels (Stemen, et al., 2005). Such a discussion would delineate the resources and motivation of actors, opportunities for choices between alternative courses of action, the possible pathways through which these choices could lead to key outcomes, or the role of contextual factors and the interaction of context and policy. This focus is the strength of the policy evaluation literature.

The literature of policy evaluation provides a well-developed means of discussing the route to unpacking the impacts of policy changes (sentencing guidelines):

- A theory-driven approach to evaluation (Chen, 1990) provides the program logic regarding how sentencing guidelines can be expected to affect outcomes. The program logic provides a model of the mechanisms that link program activities and outcomes, and then uses program theory to guide evaluation design – what actors and processes should be active in what way to produce certain outcomes (and what steps, if missed, are likely to lead to failure). For sentencing guidelines, this could mean examining how grids will be applied to criminal cases to produce a recommendation, how recommendations will affect the plea and in turn the sentence, etc.
- Program evaluators ask what has to happen for outcomes to be evaluated with integrity. For example, programs must be implemented with fidelity (Rhine, Mawhorr, & Parks, 2006; Welsh, 2006), and the treatment (sentencing guidelines) has to be separated into its key components and examined to insure program

fidelity. This step is often missing in the literature on sentencing guideline impact analysis. For example, one measure of program fidelity is guideline compliance rates. The first question to ask is whether compliance data are collected and made broadly available. Next, given diverse administrative rules for scoring noncompliant sentences, what does compliance mean in practice for similar cases, and for dissimilar cases. For a simple example, is sentence duration and disposition included in compliance? Are binding plea agreements treated as compliance?

- Realist evaluation follows a theory-driven approach. Realist evaluators see outcomes such as reductions in crime or unwarranted disparity as a result of both favorable contexts and mechanisms, and these processes must be uncovered and understood to understand policy impacts. Successful programs interventions occur when context, mechanisms and outcomes are aligned by practitioners and policy makers. “As far as evidence based policy goes it means breaking with the lazy linguistic habit of supposing it is programs that work, and resting content with how often they work. The prerequisite is to look beneath the surface in order to inspect how they work (Pawson, 2006: 24).” In the absence of such scrutiny, and given the diversity of guidelines previously mentioned, why would we expect the impact of something called sentencing guidelines to be the same across these layers of complexity?

## **Aggregate Data Exercise**

With a focus on process and theory-driven models, Hunt and Sridharan (2010) use a realist evaluation framework to examine the reported impact of sentencing guidelines on crime rates (Shepherd, 2007). Shepherd analyzed state-level crime data on the fifty States and the District of Columbia between 1960 and 2000, each of the 50 States for each year of this period. The analysis controls for state-specific economic and demographic variables which may also affect crime rates, better isolating the effect of sentencing policy on seven separate crime rates respond to both the implementation of sentencing guidelines and economic and demographic factors, while also accounting for national trends and unique state-specific features with dummy variables. Shepherd finds that guideline adoption is associated with significant increases in both violent crime and property crime.

The reanalysis of Shepherd's data began with a successful replication of her findings (Hunt and Sridharan, 2010: Table 7). The reanalysis also demonstrates that Shepherd's findings appeared relatively resilient to modifications including recoding of the independent variable of interest, implementation of some form of guidelines, dates of implementation, and model specifications (Hunt and Sridharan, 2010: Table 8), and continue to show guideline adoption associated with higher crime rates. Using a realist evaluation focus, the dataset was reanalyzed using a simple distinction dealing with the context and mechanism of guideline implementation. The authors posited that jurisdictions that expressly link prison use to resources in the presence of sentencing guidelines will have less crime, at least for the more serious crime types (e.g. murder, rape, aggravated assault). This is predicted because these systems reserve scarce prison space for those with the longest and most serious prior criminal records and the most serious crimes, while attempting to divert less serious offenders. Other guideline systems (not expressly

linked to prison resources) do not ignore prior record or serious crimes, of course, but may be more indiscriminate in who receives a sentence to prison, leading to a relatively higher proportion of less serious offenders imprisoned and higher aggregate levels of imprisonment overall. Conversely, the authors hypothesized that for the least severe crimes, this linkage may not hold as property offenders may return to the community sooner in states with guidelines linked to prison resources, and deterrence and incapacitation effects would be minimal for these relatively lower level offenders who soon reappear in the community.

The results of this realist approach are intriguing but certainly not definitive given the limits of aggregate data and the available research design. For jurisdictions with guidelines linked to prison resources, violent crimes are lower in such jurisdictions after controlling for date of guideline implementation and various social and political factors (Hunt and Sridharan, 2010: Table 9).<sup>6</sup> Other guideline systems are associated with more crime. The effect is statistically significant for aggravated assault, robbery, burglary, and auto theft.<sup>7</sup> For example, aggravated assault rates are similar for the two groups of States in the early 1960s, but begin to diverge and the growth in aggravated assault rates is significantly slower for the states that have linked their sentencing guidelines to prison resources (Hunt and Sridharan, 2010: Figure 1). Given that some of the differences in growth rates happen before the introduction of sentencing guidelines, the growth likely cannot entirely be attributed to the introduction of sentencing guidelines. Larceny on the other hand shows a different pattern – both groups of States show a similar pattern until the 1980s, when States that have linked guidelines to resources seem to be associated with

---

<sup>6</sup> That is, the coefficient for guideline systems that prioritize prison resources is negative, indicating lower crime rates.

<sup>7</sup> Of course, auto theft is by no standard a violent crime, so the coefficients are not perfectly matched to predictions.

greater growth in larcenies (Hunt and Sridharan, 2010: Figure 2), as predicted. Following the language of Forst and Bushway (2010), there is at least some suggestion that guidelines do appear to raise or lower social costs based on their stated purpose, at least under some contexts.

The conclusions flow, too, from a realist evaluation perspective. The authors do not make the case that one form of sentencing guideline is “effective” and the other is not. The analysis appears to demonstrate that universal proclamations, in this case that sentencing guidelines are associated with crime increases, are overly general and divorced from important contextual and procedural realities, and from the role of human actors.

Future research along these lines would follow many of the prescriptions laid out by Forst and Bushway (2010). They urge that future research identify key decision nodes at which discretion is exercised. These decision points can be carefully expressed in a theory-based logic model (Wholey, 1994) that provides a visual depiction of how a program is supposed to work and details the linkages between inputs, activities, outputs, and outcomes. The logic model suggests key decision nodes and emphasizes the role of rules and discretion. The above aggregate data exercise has only begun to break the surface of this work, and a better understanding of the key decision nodes and discretionary choices of individual agents is needed. Forst and Bushway also urge a better understanding of theoretical pathways that link system rules and discretion. The above exercise is simply a precursor to this work, but this prescription is very much within the spirit of realist evaluation. The aggregate data exercise detailed above does suggest that conclusions about complex policy initiatives require looking beneath the surface at program goals, context and mechanisms in order to inspect how such programs work (Pawson, 2006:24).

## References

- American Law Institute. (2010). Model Penal Code: Sentencing Discussion Draft No. 3, Philadelphia, PA.
- Bushway, S., & Piehl, A. (2007). Social science research and the legal threat to presumptive sentencing guidelines. *Criminology & Public Policy*, 6, 461-482.
- Chen, H-t. 1990. *Theory Driven Evaluations*. Thousand Oaks, CA: Sage.
- Engen, Rodney. Assessing determinate and presumptive sentencing - making research relevant. *Criminology & Public Policy* 8:2: 323-336. (2009).
- Forst, Brian and Shawn Bushway. 2010. Discretion, rule of law, and rationality. In *Criminology & Public Policy*, forthcoming.
- Hunt, Kim S. and Sanjeev Sridharan. 2010. A realist evaluation approach to unpacking the impacts of the sentencing guidelines. *American Journal of Evaluation*, 31: first published on July 8, 2010 as doi:10.1177/1098214010371817.
- Ostrom, B. J., Ostrom, C. W., Hanson, R. A., & Kleiman, M. (2008). *Assessing Consistency and Fairness in Sentencing: A Comparative Study in Three States*. Washington, DC: National Institute of Justice.
- Pawson, R. (2006). *Evidence-based policy: A realist perspective*. London, England: Sage.
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London, England: SAGE.
- Rhine, E., Mawhorr, T., & Parks, E. (2006). Implementation: The bane of effective correctional programs. *Criminology & Public Policy*, 5, 347-358.
- Shepherd, J. (2007). Blakely's silver lining: Sentencing guidelines, judicial discretion, and crime. *Hastings Law Journal*, 58, 533-590.

- Stemen, D., Rengifo, A., & Wilson, J. (2005). Of fragmentation and ferment: The impact of state sentencing policies on incarceration rates, 1975-2002. Final Report to the National Institute of Justice Grant No. NIJ 2002-IJ-CX-0027, August 2005.
- Tonry, M. (1996). *Sentencing matters*. New York, NY: Oxford University Press.
- Welsh, W. (2006). The need for a comprehensive approach to program planning, development, and evaluation. *Criminology & Public Policy*, 5, 603-614.
- Wholey, J.S. (1994). Assessing the feasibility and likely usefulness of evaluation. In J.S. Wholey, H.P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation* (15-39). San Francisco: Jossey-Bass.