Al Capone, the Sword of Damocles, and the Police–Corrections Budget Ratio

Afterword to the Special Issue

Lawrence W. Sherman
Cambridge University and University of Maryland

The Durlauf and Nagin (2011, this issue) proposal for less prison and more policing lies within our reach. Its feasibility can be increased by the following three strategies:

1. A public discussion of the present 1:1 ratio of spending on prisons and policing, unknown throughout U.S. history or in other modern nations.
2. A crime harm index (CHI) forecast at the point of arrest that would increase the number of “Al Capone” prosecutions of dangerous people for provable crimes, no matter how minor.
3. A “Sword of Damocles” approach to offender management for low-CHI forecast offenders who could be diverted from the path to prison at the point of arrest or prosecution.

In general, the academic commentators in this issue present a disappointingly short-sighted response to one of the best ideas criminology has seen in many decades. Other criminologists would be well advised to focus not on the commentators’ cautions but instead on the research agenda suggested by Durlauf and Nagin (2011), as well as on the strategic facts and approaches suggested in the subsequent discussion. Public criminology should not be politically naïve, as many of the commentaries seem to be. If criminologists agree that mass incarceration is a greater harm than malpractice of policing, they must also agree to accept a persuasive alternative as the key political compromise for reducing imprisonment. No other idea on the horizon can match the appeal of more police as the price of less prison (Sherman, 2010).

Direct correspondence to Lawrence W. Sherman (e-mail: ls434@cam.ac.uk).
The Police–Corrections Ratio

One day in 1990, the tax dollars spent on spent on corrections exceeded spending on local policing for the first time in U.S. history. (See Figure 1, computed from the Sourcebook of Criminal Justice Statistics.)

This watershed was reached after a decade of rising national investment in local policing, whereas investments in corrections at all levels rose even faster. In 1980, the cost of local, state, and federal corrections—all places that can process offenders caught by local police—was outpaced by 65% more funding going to local policing (Figure 2). But from 1990 to 2006, the criminal justice portfolio consistently had more funding invested in total corrections than in local policing.

Was the public—or “public criminology,” the kind of criminology that helps the public to understand key choices in crime policy (Loader and Sparks, 2011)—ever aware of American policing taking a back seat to American corrections? I doubt it. Nor did I ever know it had happened, until compiling the raw data in early 2010. This trend has survived for two decades without any discussion among criminologists, let alone journalists. It developed below any political radar screen, rather than by a clear democratic choice. How

---

2. This fact is not reported in Durlauf and Nagin (2011) because of the misleading inclusion of the national security functions in the federal policing budgets included in the 60% figure they do cite. Federal agencies contribute very few of the inmates in U.S. prisons and jails, with local policing contributing substantial numbers of cases to federal prisons and almost all of the cases in state prisons and local jails.
else can one account for such a decline in funding for the branch of criminal justice with the highest public confidence ratings (Sherman, 2000)?

The evidence suggests that at all prior times in U.S. history, and in all comparable G20 nations, far more money has been spent on policing than on corrections. Given the correlation between corrections budgets and prison population, the low rate of incarceration in the United States before 1980 (Blumstein, 2011, this issue) suggests that 1982 was typical of most previous periods in U.S. history. And although rising prison populations since then have received substantial press attention, their implications for the relative disinvestment in policing were never spelled out.

In England and Wales in 2010, the estimated budget for local (and some national) policing was £12 billion, whereas the estimated budget for all levels of corrections was £4 billion, a 3:1 ratio (Herbert, 2010); police accounted for 60% of the criminal justice budget, a proportion almost twice as high as the 33% in the recent United States (Figure 2). In the state of South Australia (the capital of which is Adelaide), the police budget of $668 million is 3.5 times higher than the $190 million correctional services budget. In Japan, the Prefectural (state and local) police budget of 3.4 billion yen is 1.7 times higher than the roughly 2 billion yen cost of the national prison system. In Hong Kong, the 2010 police

budget of HK$13 billion is 4.5 times higher than the correctional services budget (including community supervision) of HK $2.85 million. I can find no indication of any other nation even approaching parity between the costs of policing and the costs of corrections.

The police–corrections ratio is a separate issue from the very important data supplied by Marie Gottschalk’s policy essay (2011, this issue) about the absolute ratios of police per capita across the U.S. states:

Historical and cross-national evidence appears to buttress claims that more police means less crime, and that the United States is underpoliced. Jurisdictions with the most police officers today tend to have the lowest imprisonment rates and the smallest rates of increase in imprisonment (Stuntz, 2008: 1993). This relationship has held, more or less, since the Gilded Age. In his analysis of cross-regional variations in police per capita, murder rates, and imprisonment, Stuntz found that the South, where police and other government services were historically underfunded, has had a much lower number of police per capita and generally much higher imprisonment levels and murder rates.

Similarly, we can observe that England, Australia, Japan, and Hong Kong, all with much higher police–corrections budget ratios than the United States, also have much lower rates of homicide.

Historically, we know that the “preventive” police were developed in England as a more effective alternative to the most visible punishment: hanging. At a time when Home Secretary Robert Peel abolished the death penalty for more than 100 offense categories (Hurd, 2007), the prison population at that time was tiny, although transportation to Australia was still flourishing. Political opposition to the development and expansion of the police remained intense for decades (see, e.g., Bagehot, 1867), even as hanging dropped sharply. Meanwhile, the homicide rate in England fell steadily (Monkkonen, 2006) as police numbers rose and spread across the countryside. Yet neither comparative nor historical analysis of broad shifts in crime prevention strategy have been visible in our public criminology.

Instead, Michael H. Tonry (2011, this issue) confidently assures us that more police will only cause more racial profiling, with little benefit likely at all. Eric P. Baumer (2011, this issue) assures us that “[a]lthough we do not have good evidence about the effects on certainty of increases in police force size and/or strategic allocations of police forces, presumably, the perception of certainty is already high in these areas”—despite Anderson’s (1999) analysis, which he cites, but not for Anderson’s clear argument that poor Black people in Philadelphia have given up on the police as unlikely to do much to protect them from violence. Baumer’s comment that crime reductions in microplaces may not translate into

---

macrocrime reductions is a useful analytic caution but certainly no reason not to proceed with testing the hypothesis.

John S. Goldcamp (2011, this issue) more readily concedes that targeted policing can have some benefit but then misleads the reader into thinking that the evidence is that such effects have only short-term benefits, not lasting ones. Although Sherman (1990) and Sherman and Rogan (1995) do show early termination of local tactical initiatives, there simply is no evidence in experimental criminology that even tests for effects of increased preventive patrol for longer than a year. Sherman and Weisburd (1995) showed deterrent effects that lasted as long as the experimental dosage lasted in Minneapolis: 9 months. Had the dosage been sustained for 9 years, it may well have sustained the crime-reduction benefits for just as long. Thus, an absence of evidence on long-term police effects is hardly evidence of absence of such effects.

As of this writing, local police are being laid off across the United States, in states facing bankruptcy as a result of prison costs. Absent the elephantine costs of mass incarceration, more state funding might be provided to local education; that, in turn, could provide relief for local taxpayers to maintain the size of their police agencies. These facts complicate the commentators’ objections that Durlauf and Nagin (2011) are too narrow in their focus on only the police–corrections trade-off. Anyone who focuses on prison costs necessarily focuses on education, early childhood interventions, health, and welfare. The analytic and political focus on a “criminal justice portfolio” is merely an indicator of how voters may think about issues more clearly, despite the separation of powers between (state-level) spending on prisons and (local-level) spending on police.

The police–corrections budget ratio is an opportunity for state governors and legislators, as well as mayors, district attorneys, and city council members. The opportunity is to do just what Durlauf and Nagin (2011) suggest: Explore ways to trade expensive prisons for effective policing. Taking Baum’s (2011) point that shorter sentences alone may not reduce prison populations, city-level decisions can take the more direct approach of front-end reductions in offenders arrested or prosecuted for imprisonable crime. Although Marc Mauer’s (2011, this issue) point about closing entire prisons is the key to saving states money, big city governments can challenge rural communities’ needs for jobs with urban communities’ need for public safety. More city police might then justify strategies for diverting inmates from state prisons. State governments can incentivize these strategies by promising to split the resulting cost-savings with the cities. Doing this in the form of state subsidies for hiring more local police may well be politically viable, as well as financially sound. The political viability of this opportunity, however, will require at least two more elements to keep voters at ease: Al Capone and the Sword of Damocles.

**Al Capone and the Crime Harm Index**

The name of Al Capone can be invoked easily in the politics of punishment for an emotionally intelligent argument: Very dangerous people should be put in prison and
kept there, even if they can only be convicted of relatively minor crimes (in volumes high enough for long-term incapacitation). Blamed for the St. Patrick’s Day Massacre in Chicago, Capone could not be convicted for those murders. But he did receive a federal 17-year sentence for repeated income tax misdemeanors and felonies.

At a late 2010 meeting with 20 of the most powerful prosecutors in the United States, I invoked the name of Capone to illustrate one half of a front-end strategy for prosecutors to bring down the imprisonment rate. The other half of that strategy would use diversion, or undercharging, to minimize the use of incarceration (within sentencing guidelines) for people who are forecast to cause very little harm in the near future. But the “Capone” half could use advanced nonlinear “data mining” to identify defendants who might be charged with minor crimes, and yet pose major risks of harm (Berk et al., 2009). Not just any recidivism. The criterion can be horrible, extremely harmful crimes, from homicide to crimes against children. They would be the kinds of cases that, if committed by a defendant who was diverted from maximum prosecution, could cause the end of a prosecutor’s political career.

The meeting was an opportunity to explain my National Institute of Justice (NIJ) lecture (Sherman, 2010) on the uses of a CHI. This measure can be calibrated by the median number of days in prison provided by (first offender) sentencing guidelines for each conviction an offender would be forecast to experience in the future, by type of crime. The forecast can be based, as in our earlier work (Berk et al., 2009), on criminal history, age, local crime rate, and other objective, nonprotected characteristics (thus excluding gender, race, religion, and national origin). The CHI also could be used at a mezzo- and macrolevel to track the amount of harm committed by offenders who have been presented to a prosecutor’s office, regardless of the decisions they make to prosecute or not. Using illustrations from our probation work in Philadelphia, I showed that the very high overall rate of violent crime in that offender population was highly concentrated among a tiny “power few”: The 10% of offenders forecast to be most dangerous were charged with more than 75% of the homicides and attempted homicides, for example.

This suggestion was tied to the idea that prosecutors are first and foremost “crime busters,” and not just “criminal busters.” The idea of evidence-based prosecution would embrace any strategy that would result in less crime and, possibly, at less cost to the taxpayer. Although this is (for some prosecutors) a new idea that may be disorienting on first hearing, there have already been big-city prosecutors from Brooklyn to San Francisco who recently campaigned (and won) on platforms with similar principles.

The prosecutors cautiously welcomed my suggestion that an advanced forecasting tool could allow prosecutors to identify very dangerous people with more precision than previous forecasting methods. Many in the room, echoing James Q. Wilson’s (2011, this issue) reminder not to forget about retribution, were predictably reluctant to give up the retributive (and general deterrent) goals of prison sentences in the case of low-harm offenders. What some found appealing, however, was the use of diversion in high-volume
cases where penalties were already low, and the probability of recidivism was high. What they would want, they said, was to have a viable community supervision plan for anyone who was diverted from prison. What they meant, in effect, was just what Alfred Blumstein (2011), John S. Goldkamp (2011), and Laurie O. Robinson (2011, this issue) discussed in their policy essays. The prosecutors wanted a high capacity for celerity in case a nonimprisoned offender became dangerous. Let us call that capacity a “Sword of Damocles.”

The Sword of Damocles, Celerity, and the Regulatory Pyramid
The Sword of Damocles is an allusion to the ancient tale of a braggart who was ordered by a king to sit under a sharp, heavy sword hanging by a thin thread over the braggart’s head, all throughout a long dinner at the king’s palace. If at any point the braggart should start to brag, the king would nod to a servant who would cut the thread. The sword would then fall and instantly kill the braggart. As the story goes, celerity + severity = no bragging at dinner.

What Durlauf and Nagin (2011) imply is a slightly different, perhaps more enlightened, recommendation: more certainty + more celerity + less severity = more compliance with law. The so-called problem-solving courts for drugs, mental health, guns, and other high-volume offenders have relied, in part, on the celerity of their sanctions to achieve compliance. But they have done so in a way that embraces the graduated sanction severity of the regulatory pyramid described by Ayres and Braithwaite (1992). The “tit-for-tat” responses of mild sanctions for small infractions depends on having a finely graduated scale of escalation in sanction severity, starting with very low severity levels. At the same time, Ayres and Braithwaite argue, there must be a very big stick of severity waiting in the background. Keeping it in the background may increase compliance by minimizing defiance (Sherman, 1993), especially because its presence is well known by both the regulated and the regulators.

The Sword of Damocles is also what the drug court judge in Hawaii wields in Project Hope. But probation officers in other jurisdictions say they cannot replicate these powers. Legal delays and processes make instant consequences for violations almost impossible in those places, at least using court as the venue for regulation. Thus, a key part of the research agenda for the Durlauf and Nagin (2011) proposal needs to be inventions for achieving greater celerity from more responsive agencies (Sherman, 2011).

One model is police-level diversion, of the kind widely tested with great success at reducing repeat offending among juveniles (Petrosino, Turpin-Petrosino, and Guckenbarg, 2010). As Peter W. Neyroud (2011, this issue) notes, and will soon test himself, such diversion can be part of what an increased (or even decreased) police budget is spent on. It could include police-led programs like restorative justice (Sherman and Strang, 2007). But it also could include far more active engagement on a daily basis with known offenders of low risk who have been diverted from imprisonment.

The founders of modern police, like Peel, envisioned that police would prevent crime, in part, by managing offenders as well as places. David Weisburd’s (2011, this issue) evidence-based argument for police focusing on places is helpful, but it is not necessary for police to
give up on monitoring offenders as well. In fact, police long ago gave up, in large part, the kind of offender monitoring they had done historically (Steffens, 1931). What could make “less prison, more police” more successful in causing less crime is a more systematic means of police monitoring low-harm offenders in the community. Unlike probation, police are in the streets 24 hours a day. In principle, they are far better suited to monitor the compliance of diverted or noncustodial supervisees than any other institution in criminal justice.

The key point, as William Bratton (2011, this issue) points out, is that an increase in policing does not necessarily require increase in arrest rates; it could even lead to fewer arrests. Durlauf and Nagin (2011) clearly leave that issue up to the results of far more research. Despite some commentators’ skepticism, having more police would create more capacity for regulatory strategies that could include far more alternatives to arrest. The pessimism that police always will impose coercion, or even misuse their powers, is contradicted by impressive strides in police agency management in the past four decades. Police culture certainly has a long way to go to adopting an evidence-based perspective. But policing may well be on a rising curve of obeying the rule of law in the experience of most crime victims and offenders. It also is clearly moving toward a far greater concern with crime outcomes, and not just punishment outputs.

The fact that racial profiling and other police disasters are still reported should not mislead us into thinking that long-term improvement has not happened and cannot happen. There are many reasons that law enforcement agencies make disastrous mistakes (O’Hara, 2005), quite independent of their number of officers—or even their structures of accountability. One reason, indeed, may be not having enough officers for the challenge they face. Less prison and more police would be an opportunity to test a wide range of innovations in management, as well as strategy, to implement evidence-based policing.

Rather than opposing a “police surge” on the grounds that it would do nothing to increase celerity, criminologists should help invent new ways that policing would be all about celerity. No less than celerity, celerity is a variable that can be manipulated independently of severity. The crisis of mass incarceration is all about severity. Breaking our national addiction to severity will probably require ample “methadone” in the form of impressive increases in both celerity and severity. Such positively labeled innovations as “citizenship testing” (by police checking offender compliance with various requirements of avoiding prison) would be another way to accomplish Philip J. Cook’s (2011, this issue) suggestion to limit the opportunities for offending. So would a substantial increase in hot spots policing. Taken together, they might actually form a solution to the police legitimacy crisis Marie Gottschalk (2011) describes in high-crime neighborhoods.

**Public Criminology and Emotional Intelligence**

This journal, and especially this issue, is a strong example of “public criminology.” It is arguable whether such criminology should entertain all ideas, no matter how implausible they may seem, or should focus only on ideas that are within public comprehension. My
suggestion is that when a window of opportunity is opened wide, it is no time to be offering politically unsalable recommendations—even for research and development.

As a forum for engaging with the public owners of our “laboratory,” the public sphere is a vital place for developing good ideas. But what makes a good idea? In a new treatise on “the natural history of innovation,” Steven Johnson (2010) argued that good ideas come from several sources, many of which connect several independent ideas to make one big idea work. A key criterion for making innovations successful, he wrote, is what is called “the adjacent possible.” Citing Charles Babbage as the inventor of the core elements of modern computers in what he called an “analytical engine” in the mid-19th century, Johnson observed that the innovation went nowhere because it lacked the “adjacent possible”—the modern electronics that could implement his conceptual framework. The Google search engine, as another example, is based on an old idea that could go nowhere until the Internet created a new universe to search.

In developing new ideas, public criminology must be equally mindful of the adjacent possible. It is little use suggesting things that make political sense to academics but that strike most voters—or at least most politicians—as inconceivable. In any political innovation, from health care to climate change initiatives, the idea of an “adjacent possible” in the political environment is a foremost consideration. Success in one initiative—such as health care—may even create a backlash that will destroy the adjacent possible political space for other initiatives (such as on climate change).

Another way to describe this strategy for public criminology is “emotional intelligence.” Just as Dick Thornburgh (2011, this issue) chides Durlauf and Nagin (2011) for using complex math symbols that the public would not understand, we must all be more emotionally intelligent about how our words and hypotheses are perceived. My references to the “Willie Horton problem” of an unexpected serious crime leading to more imprisonment led to one prosecutor telling me that it bordered on being politically offensive. “Stick to Al Capone,” I was advised. Such feedback can be taken seriously as long as we actually engage with practitioners and the public to gain feedback in the first place.

Many academic criminologists may prefer to avoid such engagement, which is clearly their right to choose. Science should be open to all styles of work. But when we engage in public criminology—as we do in responding to the Durlauf and Nagin (2011) proposals—we must raise our game to the level of skill required to serve the public well.

It is not very skillful for criminologists to say “let’s just reduce prison, but not increase police. Instead, let’s increase education and child care.” Although Elliott Currie (2011, this issue) and other commentators make a strong case for broadening our policy discussions about crime prevention beyond criminal justice, a proposal to reduce imprisonment is not a viable forum for such a discussion. Logically, perhaps, it might be. But empirically, proposals to reduce imprisonment will unleash intense emotions of insecurity that anyone proposing a radical new idea must address. Few people who fear crime will be assured by the promise of a home visitation program, no matter how effective such a program may
be in the long term. If fewer people are to be sent to prisons tomorrow, many voters will need the reassurance of knowing that there also will be more police tomorrow. Perhaps they want to see more police even before the prison population declines—which is exactly what happened in New York City after 1991.

In my ASC Presidential Address (Sherman, 2003), I called on criminology to help make the criminal justice system more emotionally intelligent (Goleman, 1995). The financial crisis has created an unprecedented opportunity for criminology to affect a major policy debate. Now is the time for criminology itself to become more emotionally intelligent.

References
Baumer, Eric P. 2011. Uncertainty about reduced severity, concerns about increased certainty, and alternative paths to lower rates of crime and imprisonment. *Criminology & Public Policy*. This issue.
Durlauf, Steven N. and Daniel S. Nagin. 2011. Imprisonment and crime: Can both be reduced? *Criminology & Public Policy*. This issue.
Gottschalk, Marie. 2011. Lengthy sentences, the police surge, and the forgotten men, women, and communities. *Criminology & Public Policy*. This issue.
Mauer, Marc. 2011. The challenges of implementing research-based policies. *Criminology & Public Policy*. This issue.

Neyroud, Peter W. 2011. More police, less prison, less crime? From Peel to Popper: The case for more scientific policing. Criminology & Public Policy. This issue.


Tonry, Michael H. 2011. Less imprisonment is no doubt a good thing: More policing probably is not. Criminology & Public Policy. This issue.

Weisburd, David. 2011. Shifting crime and justice resources from prisons to police: Shifting police from people to places. Criminology & Public Policy. This issue.

Wilson, James Q. 2011. Comment on Durlauf and Nagin. Criminology & Public Policy. This issue.