WHEN BRUTE FORCE FAILS:

Strategic Thinking for Crime Control

By:

Mark A.R. Kleiman

School of Public Affairs
University of California at Los Angeles

March 31, 2005

This research was supported under award #2001-IJ-CX-0033 from the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. Findings and conclusions of the research reported here are those of the author and do not reflect the official position or policies of the U.S. Department of Justice, or of the University of California.

This report would not have been completed without the devoted assistance of Kenna Ackley, Elizabeth Dodd, Paul Fishbein, Karen Friedman, Lowry Heussler, Sherol Michaeli, Owen Paun, Nicole Saloun, and Jeannie Wilkinson.

Table of Contents

ABSTRACT

While the institutions of the criminal justice system serve many goals, crime control deserves a prominent place among them. Designers of crime control strategies need to consider the responses of potential offenders to the enforcement risks they face.

Insofar as potential offenders are selfishly rational – deciding to offend or not depending on the relationship between the cost of complying with the law and the risk of breaking it – and the resources needed to detect and punish any given crime are finite, the scarcity of punishment resources will tend to create a positive-feedback effect in the rate of offending. When the level of offending goes up, the result is that the punishment per offense goes down as the limited punishment capacity is stretched over a larger number of offenses. The lowered risk of punishment, in turn, will further stimulate offending. Thus crime rates under these assumptions will tend to exhibit "tipping" behavior, with both high-violation and low-violation equilibria possible given the same underlying causal situation.

If, on the other hand, the punishment per crime is fixed and the total amount of punishment allowed to vary, then, if the "demand curve" for offending is normal, the total amount of punishment assigned will form an inverted U as a function of the punishment per crime; great

lenity and great severity will both result in small amounts of actual punishment compared to moderation.

These considerations suggest the great importance of concentration of enforcement: by offense, by offender, and by time and place, and of the direct communication of deterrent threats in order to minimize the cost of "tipping" a high-violation equilibrium into a low-violation equilibrium. Sanctions credibility is vital. "Broken windows" policing, "Cease-Fire"-style gang interventions, and the "coerced abstinence" (testing-and-sanctions) approach to controlling illicit drug use among probationers can all be seen as applications of this simple logic. Since universal zero tolerance is never possible in a world of finite enforcement resources, targeted zero tolerance, with clear communication of precisely what will not be tolerated, is essential, and will far outperform "equal-opportunity" approaches to enforcement.

Relaxing the artificial assumption of perfect rationality among offenders to allow for such phenomena as hyperbolic discounting (excessive concern with the immediate over the distant future) and the risk-seeking behavior in losses associated with prospect theory provides additional insight into making crime control strategy, and helps explain the otherwise puzzling observation that swiftness and certainty are more important than severity in reducing offense rates. The interdependence of offense rates through enforcement swamping makes it even more

important to learn how to effectively communicate enforcement threats and to develop and maintain sanctions credibility.

EXECUTIVE SUMMARY

While the institutions of the criminal justice system serve many goals, crime control deserves a prominent place among them. Designers of crime control strategies need to consider the responses of potential offenders to the enforcement risks they face.

Insofar as potential offenders are selfishly rational – deciding to offend or not depending on the relationship between the cost of complying with the law and the risk of breaking it – and the resources needed to detect and punish any given crime are finite, the scarcity of punishment resources will tend to create a positive-feedback effect in the rate of offending. When the level of offending goes up, the result is that the punishment per offense goes down as the limited punishment capacity is stretched over a larger number of offenses. The lowered risk of punishment, in turn, will further stimulate offending. Thus crime rates under these assumptions will tend to exhibit "tipping" behavior, with both high-violation and low-violation equilibria possible given the same underlying causal situation.

If, on the other hand, the punishment per crime is fixed and the total amount of punishment allowed to vary, then, if the "demand curve" for offending is normal, the total amount of punishment assigned will form an inverted "U" as a function of the punishment per crime; great

lenity and great severity will both result in small amounts of actual punishment compared to moderation.

These considerations suggest the great importance of concentration of enforcement: by offense, by offender, and by time and place, and of the direct communication of deterrent threats in order to minimize the cost of "tipping" a high-violation equilibrium into a low-violation equilibrium. Sanctions credibility is vital. "Broken windows" policing, "Cease-Fire"-style gang interventions, and the "coerced abstinence" (testing-and-sanctions) approach to controlling illicit drug use among probationers can all be seen as applications of this simple logic. Since universal zero tolerance is never possible in a world of finite enforcement resources, targeted zero tolerance, with clear communication of precisely what will not be tolerated, is essential, and will far outperform "equal-opportunity" approaches to enforcement.

Relaxing the artificial assumption of perfect rationality among offenders to allow for such phenomena as hyperbolic discounting (excessive concern with the immediate over the distant future) and the risk-seeking behavior in losses associated with prospect theory provides additional insight into making crime control strategy, and helps explain the otherwise puzzling observation that swiftness and certainty are more important than severity in reducing offense rates. The interdependence of offense rates through enforcement swamping makes it even more

important to learn how to effectively communicate enforcement threats and to develop and maintain sanctions credibility.

INTRODUCTION

Engineers have a sardonic saying: "If brute force fails, you're not using enough." That just about describes the approach of the United States to crime control since the crime wave of the 1960s. In the mid-1970s, U.S. prisons and jails held about 300,000 prisoners. Today, they house more than two million people at any given moment: a seven-fold increase over thirty years (Hill and Paige 1998, Harrison and Karberg 2003). That brute-force approach has reduced crime partly through deterrence and partly through sheer incapacitation: the more people who would otherwise be criminally active are locked up, the lower the crime rate. However, the magnitude of the decrease in the crime rate hardly seems proportionate to the magnitude of the increase in incarceration.

The cost has been enormous: not so much in budget terms ($50 billion a year isn't actually big money in a $10 trillion economy, though the prison boom has played a part in causing a fiscal crunch at the state level, where most of the corrections spending takes place) as in the suffering inflicted on prisoners and those who care about them or depend on them and the social damage done to communities where a spell of incarceration no longer counts as an abnormal event in the life of a young man. At any one time, 8% of African-American men are incarcerated (Rafael 2003), and an astonishing 21% of African Americans between 24-44 years of age have been incarcerated at some time in their

lives. (Raphael 2003) The numbers for African-Americans living in high-poverty neighborhoods, and for those who have not completed high school, are even worse.

Incarceration is cruel; it damages prisoners and their intimates. But so too is crime, which damages not only its direct victims but also those who undertake costly efforts to avoid becoming victims, and those who lose out as a result of those efforts. That poverty is a cause of crime is a commonplace; but it is equally true, though less widely recognized, that crime causes and sustains poverty, by driving businesses and jobs out of poor neighborhoods.

The United States today combines an unprecedentedly high rate of incarceration with rates of crime that are lower than they were a decade or two ago but still twice what they were in the 1950s and early 1960s.

The obvious question is whether it might be possible to keep pushing down the frequency of victimization – and in particular the rate of violent crime – while also reducing the number of people in prison.

Many agencies, public and private, contribute in one way or another to reducing the incidence of crime. Not all of those agencies have any explicit crime-control responsibility. By the same token, agencies of criminal justice have many missions, of which crime control is only one. It would, therefore, be a mistake to imagine an entity called the "criminal

justice system" with a well-defined objective function and the capacity to formulate plans.

Still, it is possible to consider crime control as an objective, and to think about alternative means to that end. As Cook (1986) has pointed out, the incidence of completed crime is a very imperfect measure of the success of crime control, since a low level of completed crime can co-exist with a high level of victimization risk. That will happen when the environment is so risky, and potential victims engage in so much costly crime avoidance, that actual victimization is low. When a big city closes its parks at midnight to prevent muggings, the rate of actual muggings in the parks after midnight will be very low; but that reflects a defeat, not a victory, for the forces of law and order.

Crime control policy therefore ought to aim at minimizing, not completed crime, but the criminal threats faced by potential victims in various parts of the social environment. (Conceptually, we could operationalize the level of criminal threat as the victimization rate at some fixed level of crime avoidance or exposure: for example, the probability that a car parked overnight at a given location with a suitcase in the back seat is broken into, or that a given person walking down a given street at a given hour is assaulted or robbed.)

From this perspective, designing crime control policies requires making predications about the behavior of potential offenders under different circumstances, including different sorts of punitive threats.

The current American criminal justice system is best characterized as a process by which most offenses are never detected or are punished only trivially, but very severe sentences are handed out to some offenders, almost at random, and usually at a considerable distance in time from their offenses.

If offenders were perfectly rational, random delayed severity would be an effective approach to crime control. But perfect rationality is a poor model of offender behavior, at least for many offenders who are instead – like most people under some circumstances – reckless, impulsive, myopic and ill-informed. For such offenders, random delayed severity works poorly. Holding constant the overall level of punishment, there is reason to believe that it will usually (though not always) be the case that swifter and more predictable punishments will have greater total deterrent effect than slower and less predictable ones.

The risk of apprehension and punishment faced by an offender who commits a given crime in a given location depends on two sets of conditions outside that offender's immediate control: the level of enforcement activity and punishment capacity available to use against that class of crime in that location on the one hand, and on the other the number of other offenders doing similar things nearby. Thus offenders, like prey animals, find safety in numbers; that is the logical underpinning of a riot.

Conversely, from the law enforcement perspective, as long as enforcement is allocated among offenses on roughly an equal-opportunity basis the cost of creating any given level of risk for a particular class of offenses is proportional to the number of offenses being committed. Where offenses are common, they may "swamp" the enforcement system, creating a self-sustaining situation in which high offense rates lead to low punishment risks for offenders, and those low punishment risks in turn sustain high offense rates.

In those circumstances, equal-opportunity enforcement will fail to control crime. But if enforcement is concentrated in ways that offenders can perceive, then the perceived risk of committing a particular offense at a given place (though not of all offenses at all places) can be made very high. A high perceived punishment risk will tend to drive the offense rate within the area of concentration down, and with it the cost of maintaining a high rate of apprehension and punishment per offense. Once that process is complete, the resources no longer needed in the initial area of concentration can be moved to another, creating in stages a crime control effect that could not have been created all at once.

Since that process of concentration depends for its efficacy on the perceptions of offenders, perception management sometimes deserves a central rather than a peripheral place in enforcement planning; an unperceived threat has no deterrent power, and there is no assurance that offenders will, on their own, notice even substantial changes in the

probability of apprehension and punishment. So the designers of enforcement policies need to consider how best to communicate deterrent threats, including communicating them directly and personally to high-risk individuals and groups.

Predictability, swiftness, concentration, and communication: these four principles, substituted for brute force, might be able to greatly reduce both the level of crime and the number of Americans behind bars.

References for Introduction

Harrison, Paige M. and Jennifer C. Karberg. 2004. "Prison and Jail Inmates at Midyear 2003." *Bureau of Justice Statistics Bulletin* NCJ 203947 (May).

Hill, George and Paige Harrison. 1998. "Prisoners Under State or Federal Jurisdiction." *Bureau of Justice Statistics, National Prisoner Statistics Data Series* NPS-1. Available from http://www.ojp.usdoj.gov/bjs/prisons.htm January 31, 2005.

Holzer, Harry J., Steven Raphael and Michael A. Stoll. 2003. "Employment Dimensions of Reentry: Understanding the Nexus between Prisoner Reentry and Work." Paper presented at Urban Institute Reentry Roundtable, May, New York City, NY.

Raphael, Steven. 2003. "The Socioeconomic Status of Black Males: the Increasing Importance of Incarceration." Working Paper. Goldman School of Public Policy, Berkeley, CA.

CHAPTER 1

DESIGNING ENFORCEMENT REGIMES

What combination of laws and enforcement activities would result in the least total damage from crime and crime control? That straightforward (albeit hard–to–answer question) is not the only question that might be asked about the design of a crime control regime, or even the question most frequently asked.

Any discussion of crime control strategy, and especially the part of the crime-control effort that involves investigation, prosecution, and punishment, runs into the irreducible fact that the problem has moral aspects as well as purely operational ones. We care about the accuracy and proportionality of punishments in general, and we care about whether this particular offender gets his just deserts, which means also whether this particular victim and the victim's intimates get to have the wrong done them publicly acknowledged through the punishment of the wrongdoer.

Still, crime can also be considered merely as a social problem, like air pollution. The rates of fraud, theft, and assault in all their varieties, and the violation rates for various regulatory regimes including the regulations about abusable drugs, are facts about the world, as the

ozone count is a fact about the world, and one of the purposes of criminal justice operations is to reduce these rates.

Putting aside the question of justice, we could set up the crime-control problem as a cost minimization. There are costs of crime, and there are costs of crime control, and the optimal policy (ignoring justice) would be to find the set of enforcement strategies that minimized the sum of those costs over time, subject to the constraints on the system such as the budget, the organizational realities, and the procedural requirement of a fair trial.

To each offense we could assign a cost, based on the willingness-to-pay of those damaged by it to avoid that damage. (Or, to be more precise, the relevant quantities are the willingness-to-pay of those at risk of being damaged by that sort of offense for reductions in those risks; that formulation avoids the problem of asking someone the unanswerable question "How much is it worth to you not to be murdered?" and replaces it with the still difficult, but in principle answerable, question "How much would it be worth to you to reduce your probability of being murdered this year from 1 in 10,000 to 1 in 20,000 without changing your activities?" (Howard, 1989; Viscusi, 1993).

Those willingness-to-pay figures would include not only the costs of victimization but the costs of precaution – all the things people do at some cost because they fear victimization, and might do less of if the

victimization risk were lower – and residual anxiety about being victimized. To that figure, summed up over offenses, we would have to add the indirect costs of victimization and especially of avoidance: the lost job opportunities in poor urban neighborhoods, for example, because shoppers and shopkeepers don't want to do business there. There is good reason to think that avoidance costs, indirect costs, and residual anxiety add much more to the total than completed victimizations. (That doesn't mean that the precautions are irrational at an individual level; it may be that taking lots of precaution is the crime-cost-minimizing strategy for any individual, though due to external effects we would all be better off if everyone took fewer precautions.)

Simplifying slightly, we can model an overall crime control strategy, or crime control regime as specifying for each type of offense a degree of enforcement effort and a punishment, or probability distribution over punishments, for each detected offense of that type. Given any proposed regime, and a prediction about the level of criminal activity and precaution that would result of we chose to put that regime in place, summing up the costs would give the total cost of the crime not prevented by that crime control regime, or, again more precisely, the cost of the victimization risks the regime fails to eliminate.

On the other side of the coin, we would sum up the costs of crime control: the budget costs, of course, but also the damage inflicted by the enforcement and punishment process, which we can think of in terms of

offenders' (and innocent people's) willingness to pay to avoid the risk of various sorts of enforcement interaction, from a street frisk to a wiretap to a prison sentence, either for themselves or for the people they care about. (It has been argued that the suffering resulting from an appropriate sentence isn't a cost, but that approach is inconsistent with the principles of benefit-cost analysis, which doesn't generally consider the moral or legal status of a valuing subject in deciding how to count that subject's preferences.)

Note that the total cost of punishment in this sense is distinct from the budgetary cost of law enforcement. That distinction is an important one. The budget (determining the number of officers, courtrooms, and prison cells or other corrections capacity) is not just part of the minimand, it is also a constraint. A solution to the crime-control problem that requires a bigger budget than the political process will make available must be rejected as infeasible, and much of what follows will depend on the fact that enforcement agencies need to develop strategies that control crime while economizing their own resources.

For each regime, then, we could in principle calculate a projected cost of crime and a projected cost of enforcement, encompassing not just police work but prosecution and corrections as well. We can ask what policy would be optimal in the sense of minimizing that sum of costs. (Having done so, we might then choose some other policy for reasons not captured in the cost-minimization exercise, but the calculation would tell

us how much it costs us on the crime-control side to serve those other purposes or obey those constraints.)

Of course actually doing the requisite measurements and making the necessary predictions would be absurdly beyond all capacities. Still, one might ask about the characteristics of such hypothetical abstract and simplified total-cost-minimizing enforcement regimes as a heuristic for designing actual policies for the actual complicated and messy world.

In that sense, this project attempts a theory of rational crime control: it does not assume that actual decisions about crime control efforts are in fact consistent with economic rationality, but it asks what policies would be adopted if that were the case. Despite the necessary abstractions and simplifications, the conclusions may be heuristically useful in pointing toward improvements in actual policy.

Unlike earlier work in this vein going back to Bentham and Beccaria and forward to Becker and his school, however, the current project does not assume that potential offenders are acting in an economically rational fashion, an assumption that would conflict both with the findings of ethnographic studies of offender decision-making and with the observed relationship between the gains from some kinds of offending and the associated risks of punishment (Bentham, 1830; Beccaria, 1764; Becker, 1974).

But rationality isn't binary; it is possible – more than possible, it is virtually universal – to be a little bit irrational, or somewhat irrational, or even a whole bunch irrational, without having one's decisions utterly uninfluenced by their likely consequences. As long as potential offenders are, on average, somewhat less likely to commit a given offense when either the probability or the severity of the punishment increases, then the enforcement system can influence their behavior in the direction economists would predict, even if the offenders aren't perfectly maximizing their subjective expected utility subject to the constraints of the situation.

Recent work in what is now called behavioral economics has begun to discern some patterns in imperfectly rational behavior. Undervaluing the distant future compared to the immediate future – the phenomenon that explains, among other things, procrastination and most Americans' failure to save as much for their retirement as they think they should – is among the most important, along with the tendency to undervalue small risks of large disasters (Kahneman and Tversky, 1979; Ainslie, 1991). But the sheer difficulty of acquiring, and appropriately processing, the relevant information also looms large; actual offenders, like actual people of other types, experience both costs of gathering and processing information and limits on their capacity to do so.

That will be the working assumption here: that offenders are imperfectly rational, but not utterly insensitive to consequences. (The

validity of whatever tentative conclusions are drawn is therefore limited to circumstances in which that assumption holds good.) The application of that set of ideas to the problem of crime control is one of the central themes of this essay. If we assume that offenders shy away from punishment, and shy away somewhat more from more likely and/or more severe punishment, but suffer – perhaps in aggravated form – from some of the same decision-making imperfections studied by the behavioral economists, how would that modify our thinking about optimal crime control?

This is, therefore, an essay in what Raiffa, in *The Art and Science of Negotiation,* calls the asymmetric prescriptive/descriptive mode, *describing* the behavior of one set of actors in a strategic system in order to *prescribe* good choices for the other set of actors (Raiffa, 1982). The imagined reader of this document is someone interested in controlling crime, and thus someone who identifies with the problem faced by the designers of enforcement regimes. (And, therefore, second person pronouns will refer to the enforcement decision-maker.)

Two sets of ideas dominate the analysis. The first has just been mentioned: imperfect rationality (including imperfect information) among offenders, and its consequences. It is not enough to make sure that crime doesn't pay; it is necessary to ensure that crime doesn't even seem to pay, and that offenders won't offend despite "knowing" at a purely cognitive level that crime doesn't pay.

The second is "enforcement swamping": the tendency for the risk of punishment for any given offense type to fall if the frequency of the offense rises (unless, as generally is not the case, enforcement resources rise as quickly as offense levels do). As growing numbers of offenses meet relatively fixed enforcement resources, the risk associated with committing an offense fall. That makes the choices of potential offenders, insofar as they are acting more or less rationally, partially interdependent: since the risk of being punished for a crime tends to fall as the number of crimes committed by others rises, the attractiveness of offending will rise along with offense rates. This builds a positive feedback loop into the offending process: If offenders react in the expected direction to changes in the risk of punishment, and insofar as they correctly perceive what that risk is and how it changes, then both increases and decreases in offending will, to some extent, be self-reinforcing. By the same token, offences will tend to cluster: by crime type, by geography, and by time.

Taken together, these two ideas suggest the vital importance of *the communication of credible deterrent threats*.

Consider a perfectly rational, foresighted, and self-controlled, but un-conscientious, individual who has to make a series of choices about observing or breaking some law, where the cost of observance or the benefit from violation is X, measured in dollars. ("Un-conscientious" in the sense of not having any scuples about breaking the law, and simply

reacting to the reward-and-punishment situation.) And consider the problem of (again, radically simplifying) an integrated enforcement authority that gets to decide two things: how much effort to put into detecting violations and how large a penalty to exact for each detected violation.

To start with, for simplicity let the violation benefit X be a constant.

Let the proportion of instances in which he violates the law be V, where V is bounded by zero and one.

Call the total amount of punishment inflicted P.

Let the probability that a violation is detected be D, again bounded by zero and one.

Let the penalty for a detected violation (say, the fine, but generally the sentence) be S, bounded from below by zero but unbounded above.

Again simplifying, let D and S be constants across offending opportunities.

Then the expected value of punishment per violation committed $E[P_v]$ is D•S, and the expected value of punishment per offending opportunity $E[P_o]$ is V • D • S (Kleiman, 1993).

Again for simplicity, assume that punishment is sufficiently close in time to offending that discounting can be ignored.

Given the simplifying assumption of homogeneity among opportunities, a perfectly rational offender will either offend every time (if $E[P_v] < X$) or never offend, if $E[P_v] > X$.

Thus, in this deliberately unrealistic situation, $E[P_o]$, the expected punishment per opportunity to offend, is an increasing function of D and S up to the point where D • S = X, and then abruptly drops to zero as the violation rate V drops to zero. (If offending opportunities are homogeneous, then either the kind of offense in question is worth committing, from the offender's viewpoint, because the gain from doing so is greater than the expected loss from being caught and punished, or it isn't; either the offender will take every violation opportunity, or none of them.)

Increasing the certainty and severity of punishment will therefore be futile and expensive just up to the point where it becomes perfectly efficacious and (aside from the costs of monitoring) costless. The strictest regime will also be the least punitive, in the sense of inflicting the least actual punishment, while a regime only somewhat less strict will involve the maximum amount of actual punishment.

That little piece of reasoning does not, of course demonstrate anything about any actual situation or decision. But it does establish an important logical possibility: that beyond some point increasing the certainty and/or severity of punishment might lead, not just to less crime, but to less total punishment. In a country suffering both from much higher crime than it had forty years ago and from the expense and cruelty of keeping two million of its people locked up at any given moment, that logical possibility is worth pondering.

If opportunities for this hypothetical individual are allowed to be heterogeneous – if one or more of X, D, and S vary from opportunity to opportunity in ways the potential offender can accurately perceive, or if instead of one offender we have many – then matters are different. If opportunities are heterogeneous, then the turnaround from much punishment and much offending to little punishment and little offending will not occur at a knife-edge, but gradually. Instead of a rising curve that suddenly drops to zero, the curve of expected (or aggregate) punishment per opportunity will be an inverted U.

In this framework, the expected punishment for an offense could be thought of as a kind of "price," in the economic sense of obeying the Law of Demand: the higher the price of a good, the less of it will be demanded. For these purposes, and again for perfectly rational offenders, it is s not the nominal punishment that counts, but rather the combination of the actual punishment likely to be handed out if an offense is detected and the probability of getting caught.

Thinking then, of offending as a sort of commodity purchased at some price in the form of expected punishment, the inverted -U result will hold true whenever the absolute value of the punishment-elasticity of criminal activity is below unity up to some point but rises to above unity beyond that point: that is, if offending behaves as what economists call a "normal good." (The intuition in economics is that, as price rises, a given percentage change in price has a bigger impact on the consumer's budget, and therefore makes a bigger impact on his behavior. There is no formal reason why the analogous relationship should hold in the crime-and-punishment case, but the inverted U can be observed in the behavioral psychology laboratory.)

If we make our hypothetical offender not just perfectly rational but perfectly informed and risk-averse, even a low-probability threat of a truly drastic punishment would act as a virtually complete deterrent, meaning that the level of surveillance, and thus its cost, could be kept low. Keeping the cost of surveillance small means that almost all the

cost of keeping crime too "expensive" to be attractive to offenders would be the cost of actually delivering punishment, and in that hypothesized world a convincing threat would be enough, leading to a very-low-crime, very-low-punishment, very-low-total-enforcement-cost world.

If actual offenders have less foresight, less risk-aversion, and less self-control than the rational actors of the economics textbooks, getting past what ought to be the peak of the inverted U may not be enough, since they may continue to offend even when offending no longer pays.

Thus heterogeneity, imperfect rationality, and imperfect information all complicate the analysis. Still, the question of how to make the actual world approximate the ideal one is worth pondering. There may be important cases in which it turns out to be possible to have less crime and less enforcement than current policies produce by convincingly communicating deterrent threats.

One consequence of costly and imperfect information-gathering and information-processing is the presence of perceptual lags. These are costly to the decision-maker (in this case, the offender), but they also complicate the problem of anyone trying to influence his decisions: in this case, the designers of our hypothetical crime-control regime. If there is a lag between a change in enforcement and punishment practices and offenders' observing that change and adjusting their behavior accordingly, then moving to a higher-probability or higher-severity

punishment regime will incur transition costs, even if, in equilibrium, it pushes the system past the peak of the inverted U and thus in the long run would produce a low level of offending and consequently of punishment. That means that casual empiricism – trying things to see how they work and abandoning them after a while if they work badly – won't be a good meta-strategy for law enforcement, since the short-run results of a policy on the total amount of punishment required may have the opposite sign from the longer-run results: A strategy that looks disastrous when first tried might (or, of course, might not) pay off with persistence.

In the presence of budget constraints, and in the absence of capital-market-like mechanisms for converting potential future savings into currently spendable resources, the existence of transition costs will sometimes make otherwise optimal strategies infeasible. That points our attention toward means of reducing those transition costs.

Again, a highly simplified model provides an illustration.

Imagine 1000 perfectly rational commuters who have to choose each day between parking legally and paying $10 for the privilege and parking illegally in what is supposed to be a shopper-parking area and risking a $30 ticket if they get caught staying more than four hours. That makes X, the cost of compliance with the law, $10, and S, the penalty for a detected violation, $30. Assume again that they are not conscientious

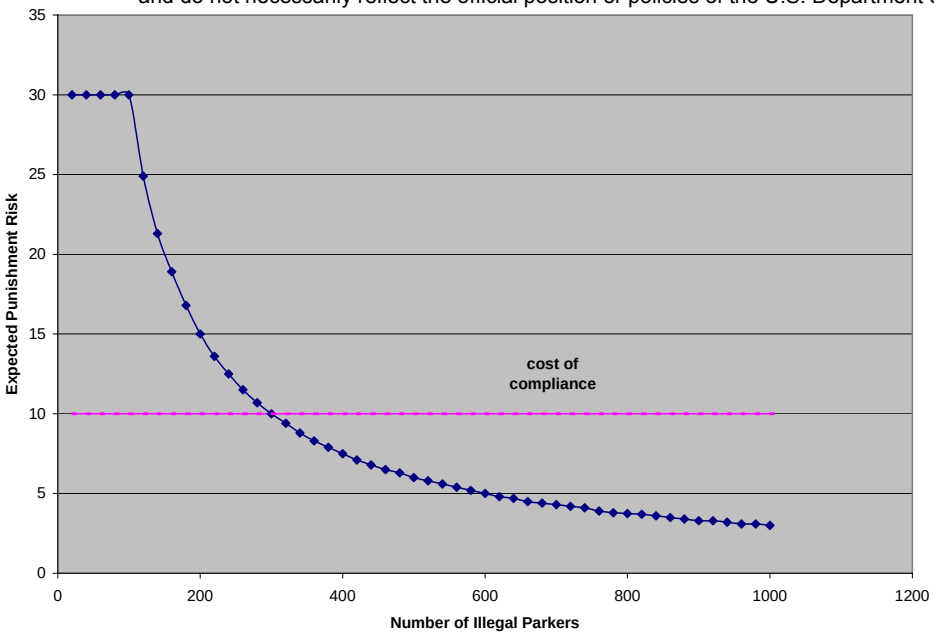
about parking, but simply try to park as cheaply as possible. Assume further that there is only a single ticket-writer, capable of writing no more than 100 tickets per day.

(Parking fines are different from most actual punishments in that the loss to the violator is a gain to the state. To make the analogy to criminal justice operations closer, imagine that the penalty for misparking is that the attendant uses a key to scratch the paint on the vehicle, and that the $30 is the repair cost, or alternatively that it costs more than $30 to collect $30 in parking fines. The point is that most punishment is deadweight loss rather than transfer.)

Assume still further that, to start with, all 1000 commuters have decided to park illegally. (If we set the problem in Boston that assumption becomes plausible.)

Now imagine that you are the parking authority, and that your mission is to get the commuters to start parking in the pay lot.

Your first step should be what Schelling calls "vicarious problem solving": analyzing the problem from the commuters' viewpoint (Schelling, 2003). That leads to a discouraging finding: they are behaving rationally. Since your ticket-writer can only write 100 tickets a day, the risk of a ticket for each commuter is only one in 10 (A in the analysis above), which makes his expected punishment per violation $E[P_v]$ only 10% of $30, or $3. That's cheaper than parking legally -- $E[P_v]$



< X – so commuters have no reason (in the absence of conscience) to change their unlawful ways. Despite the imposition of $3000 a day in fines, the shoppers' lot stays full of commuters' cars. The volume of offenses has swamped the capacity of the enforcement system, reducing the expected punishment per offense to the point where offending is rational. Thus a high violation rate is self-sustaining, just as a low violation rate would be; if we started with all the commuters parking legally (imagine Minneapolis rather than Boston), that too would be rational behavior, since any one person parking illegally would be sure to get a ticket.

Both "everyone parking illegally" and "everyone parking legally" are what the game theorists call "Nash equilibria": given the actual behavior of the players, no player regrets his decision. In this simplified model, those are the only two Nash equilibria. If almost everyone complies, those who fail to comply are making a mistake, and some of them will figure that out and behave differently next time. If almost no one

complies, those who comply are making a mistake, and the same principle applies. A Nash equilibrium is an equilibrium – a stable state of the system – precisely because no one (rationally) wants to change what he's doing. The system described has two Nash equilibria, with an unstable Schelling "tipping point" between them, a level of violation where the advantages and disadvantages of violation are closely balanced but where any deviation away from the point of unstable balance – toward more compliance, or towards less – feeds on itself and "tips" the system all the way to the extreme in which it started to move, like a coin balanced on edge falling in whatever direction the wind nudges it (Schelling, 1978; Gladwell, 2002).

(This analysis may illuminate the well-understood but undertheorized fact that offending tends to be highly concentrated; if the expected punishment per offense is lower where offending is common, then offenders will rationally seek safety from enforcement in numbers, as prey animals do from predators. It might also explain why "broken windows" policing – cracking down on minor offenses – sometimes succeeds in reducing the rate of serious crime; burglars and robbers might take disorderly behavior in a neighborhood as an indicator that the neighborhood is at a low-compliance equilibrium and therefore a relatively safe place to commit serious crime (Kelling and Wilson, 1982).

Returning to the parking-lot example: your problem is to move from a high-violation-rate equilibrium, where the rate of offenses has

swamped the enforcement mechanism, to a low-violation-rate equilibrium.

Raising fines wouldn't help the parking situation unless you could raise them to nearly $100 (or perhaps a little bit less if the commuters are risk-averse). Assume that option is out, for example because the fines are fixed by law.

You could try hiring more ticket-writers. Four ticket-writers would be enough to raise the risk of getting a ticket from 10% to 40%, making the new $E[P_v]$ 40% of $30, or $12: enough to make rational commuters park legally, since $E[P_v] > X$. If the commuters were not only perfectly rational but also perfectly informed, as soon as the extra ticket-writers were hired they would all begin parking legally. In a slightly more realistic situation, it would take time for the commuters to perceive and act on the increased detection risk. Eventually, though, some of them would start to notice it, and adjust their parking strategy accordingly.

As they did so, the expected punishment per violation $E[P_v]$ would rise for all of the remaining commuters. If, for example, 200 commuters park legally, then the remaining 800 miscreants face a 50% chance of a ticket rather than a 40% chance, pushing $E[P_v]$ up to $15 and thus making the fact that misparking is now a suboptimal strategy easier to notice. Each additional dropout from the practice of misparking increases the pressure on the remainder, by allowing the fixed

enforcement resource to be concentrated more and more narrowly. Eventually, given a workable level of rational behavior among the commuters, you will have succeeded in switching from the Boston equilibrium to the Minneapolis equilibrium.

Once everyone is parking legally – indeed, once there are fewer than 400 violators per day – you no longer need four ticket writers. At a low enough level of violation, you might not even need one full-time ticket-writer to maintain virtual certainty that misparking will lead to a ticket and thus maintain the new, more desirable equilibrium. In a world of high compliance, there will be less need of surveillance, and less use of actual punishment, than was true in the world of low compliance. Having solved the problem in the lot in question, you can move the three excess ticket-writers to the next problem lot.

(Yes, once the extra ticket-writers have moved on, the new equilibrium is vulnerable to co-ordinated violation: against an enforcement capacity of only 100 tickets a day, if on any given day all 1,000 commuters agreed to mispark, they could swamp the system once again. But characteristically the coordination among offenders that creates crime "hot spots" is tacit, or even unconscious, rather than explicit. Unless the commuters hold a meeting or form a conspiracy, their attempts to test the new system will be uncoordinated, and therefore the new system will pass the test.)

How long you need to keep four ticket-writers working depends on how fast commuters adjust to changes in the risk of detection. Since the ticket-writers are a scarce resource, you want to economize on their use, which means you want to speed up the adaptation process. Since the commuters, being somewhat rational, also want to act on accurate rather than outdated information, *your interest as the designer of an enforcement regime runs parallel to the interests of the offenders whose behavior you are trying to shape*.

So what techniques are available to help speed the adaptation process?

One option is to hire ten ticket-writers instead of four, allowing you to ticket every single misparked car even if all 1000 commuters continue to violate the rules. That seems wasteful, since four ticket-writers will suffice to make misparking disadvantageous. In the short run, it is also unnecessarily punitive, since the total penalties imposed on the first day of the crackdown will be $30,000, ten times the previous rate. If only four ticket-writers are hired, the daily penalty rate will never rise above $12,000.

But what seems superfluously drastic in the short run may be economical and merciful in the only-slightly-longer run if, as seems plausible based on what happens in the behavioral-psychology laboratory, people adapt much more quickly to moves toward certainty

than they do to intermediate changes in probability. Even when the risk of getting a ticket is only 10%, getting one is not so unusual an event as to force a reasonable commuter to reconsider his parking strategy. Being ticketed on two successive days, though, ought to be a very big surprise, since its probability should only be 1%. And anyone ticketed three days running who doesn't suspect something odd is going on simply isn't paying attention: it's much more plausible that something has changed in the world than that one was the victim of a one-in-a thousand streak of bad luck. Either, a reasonable commuter will reason, one of the ticket-writers dislikes me or my car and is singling me out for attention, or the level of enforcement is no longer anywhere close to 10%. Moreover, commuters are likely to learn of one another's experiences, either through conversations or by observing how many cars have tickets on them.

So while having four ticket-writers working might take weeks or months to cause a drastic change in parking behavior, having ten ticket-writers working for a week should do the trick; if so, you will economize over time on both enforcement expenditure and punishment imposition by going in heavy rather than light. (The same principle sometimes applies in warfare and medicine; as the old saying has it, "If you have to cut off a dog's tail, it isn't really merciful to do it a little bit at a time.")

Other than going in with overwhelming force, how could you make the process happen faster? The entire process depends on the

expectations of the commuters; you want to change them as quickly as possible.

One answer is *concentration*. If you decide to use only four ticket-writers, why not tell them to work from the front of the lot, producing the certainty of a ticket for the 400 cars closest the front rather than 40% chance of a ticket for each misparked car? This makes the front of the lot a zero-tolerance zone, as opposed to having ten ticket-writers, which creates zero tolerance for the entire lot. If commuters fail to notice the pattern, those who tend to park in front will learn quickly that the enforcement regime has changed, because they get ticketed every time; if commuters notice the pattern, they will all try to crowd toward the back, allowing your ticket-writers to push the zero tolerance zone further and further back until the whole lot is covered. This approach might make the transition time with four ticket-writers nearly as short as that with ten ticket-writers, saving resources and avoiding the imposition of a great deal of punishment.

Another answer is *communication*. Instead of making your enforcement efforts easier to notice through the design of the efforts themselves (massive force or concentration), you could make them easier to notice by pointing them out. If your four ticket-writers are going to work from the front of the lot, why not post a sign designating the first 400 spaces as a zero-tolerance zone, to save days in the process of having the commuters notice? Posting the threat in advance makes the

delivery of the threat that much more impressive. Once commuters have learned to respect that zone, there are only 600 spaces left to mispark in; if no more than 600 cars mispark, the risk of a ticket (D) for any misparked car is two out of three. (Or you could expand the posted zero-tolerance zone to the first 800 spaces on the way to a blanket zero-tolerance policy.)

But as long as you're going to announce the zero-tolerance zone, why not do it in advance? Put up posters or hand out flyers a day before the three additional ticket-writers show up, stating that fact, pointing out that the risk of getting a ticket is about to quadruple (which will almost certainly be an understatement) and declaring the first 400 spaces zero-tolerance.

Even if no one believes the notices, you're in no way worse off than you would have been without them (except for the cost of the communications program) and you're actually ahead in two ways: the warning will make the actual enforcement effort more noticeable, and your next warning will be more likely to be believed.

If, on the other hand, some commuters believe the warnings, they will park legally, increasing the pressure on the rest and making for a quicker and lower-total-punishment transition to a high-compliance equilibrium. If 90% or more of the commuters believe the threat the first time out, you won't even need any actual extra ticket-writers; your

original single ticket-writer will be able to enforce zero tolerance on the entire lot.

That raises the obvious question: If warnings alone might do the trick, then why not rely on warnings alone? That is to say, why not bluff and hope that your bluff becomes a self-fulfilling prophecy?

The answer is: Bluffing sometimes works, but it carries a risk that shouldn't be ignored. A threat not carried out reduces the credibility of future threats. If the interaction between the law enforcement system and potential violators were zero-sum, then deception by law enforcement would be harmless, though also somewhat difficult. But if the interaction is variable-sum – if some outcomes are better for both sides – then both sides have a stake in the possibility of arriving at those outcomes. In a variable-sum game, such as a negotiation, there exist opportunities to make promises, threats, and factual claims with the intention of having them believed by the other party (or parties), and there is often short-term gain from successful swindling. But a negotiator who engages in a series of negotiations – as the law enforcement system engages in a series of interactions with various groups of actual and potential offenders – needs to worry about the impact on the next negotiation of engaging in deception in the current negotiation. There is a large advantage to being known as someone whose word is good: who doesn't say what isn't so, and doesn't promise or threaten and then fail to deliver, even when doing so is beneficial in

the short run. Most threats, including legal threats of punishment, are costly to carry out, so someone whose word is not relied on may not be able to threaten successfully.

The term usually used in negotiation for one's reputation for following through on promises and threats is "credibility." In the current analysis, the central problem of law enforcement is the production and maintenance of *sanctions credibility*: an agency that can issue threats that are known to be valid can change the behavior of actual and potential offenders at low cost, while an agency known for bluffing will need to do a lot of actual punishing to acquire in a particular case the credibility it lacks in general. (This is analogous to the "policy credibility" discussed by students of macroeconomic management.) Bluffing is the enemy of sanctions credibility, and bluffing by one enforcement agency imposes costs on other agencies through what is sometimes called a "reputational externality."

Of course, whether one is bluffing or not is a matter of degree. Declaring the entire parking area zero tolerance when in fact you only have enough capacity to ticket 10% of the potential misparked cars is a bluff. But doing so when you have capacity to ticket only 70% may well not be a bluff, if you have reason to think that your existing credibility, supporting whatever sort of communications strategy you plan to pursue (you could, for example, raise the ante from posters and flyers to warning notices placed on the windshields of misparked cars) is strong enough to

scare away 30% of the commuters. Nonetheless, promising more than you are (morally) certain you can deliver is risky business.

So ramping up the enforcement level temporarily can, in principle, move a situation from low-compliance to high-compliance, and the cost of doing so and the time required can be reduced by making the temporary intervention intensive, by concentrating it, and by communicating the threat, especially starting from a position of credibility.

What about the more typical situation, in which the enforcement decision-maker lacks control not only over the level of punishment per offense but also over the level of enforcement activity? Imagine the parking-control problem where the 100-ticket-per-day limit is an inflexible constraint. Starting from the high-violation equilibrium, it might seem impossible to move a group of rational but not conscientious potential violators to the low-violation equilibrium, since universal violation is a Nash equilibrium: as long as everyone else is doing it, the cost of compliance is above the expected cost of violation.

Indeed, no strategy that gives each offender the same chance of detection can do the job. But relaxing equal opportunity by concentrating enforcement resources on a subset of offenses or offenders allows you to make offending irrational for the target individuals or

offense types, and doing so sequentially allows you – again, in this simple model -- to bootstrap your way to the high-compliance world.

For example, you could tell the ticket-writer, as above, to work from the front of the lot. Doing so dissolves the high-violation Nash equilibrium; the 100 commuters who park closest to the front are now making a mistake. Once commuters figure that out (which you can help them do more quickly by direct communication) they will start to leave some of the first 100 spaces free for shoppers (which means that the late arrivals will have to pay to park legally to avoid the ticket they know they would otherwise surely get). But every commuter who parks legally puts someone else into the effective zone of zero tolerance: if the first twenty spaces are left vacant, then the zero tolerance zone stretches back to Space #120. Slowly the zone will expand to cover the entire lot, and you have moved from the Boston equilibrium to the Minneapolis equilibrium.

In principle, this would work just as well with any system for prioritizing enforcement – for example, starting with the lowest license plate numbers and working up – as long as those commuters singled out for certainty of enforcement have some way of determining who they are:

a principle of concentration too mysterious to figure out is like a threat never delivered.

Thus in this example concentration alone, even without adding to the total level of enforcement, is sufficient to "tip" the system from one equilibrium to the other. That is, potentially, very good news indeed.

Insofar as enforcement swamping is the major problem, concentration and communication look like workable solutions; it will often be possible even with limited enforcement capacity, to raise benefit to the offender, thus discouraging any offender sufficiently rational enough to figure out the odds and sufficiently self-controlled enough to act on the result of his calculations.

The bad news is that making crime unattractive to a rational actor may not be enough, because offenders may not be entirely rational actors.

References for Chapter 1

Ainslie, George. 1991. “Derivation of 'Rational' Economic Behavior from Hyperbolic Discount Curves.” American Economic Review 81(2):334-340.

Beccaria, Cesare. 1986. On Crimes and Punishments; translated from the Italian in the author's original order, with notes and introduction by David Young. Indianapolis: Hackett Pub. Co.

Bentham, Jeremy. 1830. The Rationale of Punishment. London: R. Heward.

Becker, Gary and William Landes. 1974. “Essays in the Economics of Crime and Punishment.” New York: Columbia University Press

Gladwell, Malcolm. 2002. The Tipping Point: How Little Things Can Make a Big Difference. Boston : Little, Brown.

Howard, Ronald A. 1980. “On Making Life and Death Decisions.” In Societal Risk Assessment: How Safe is Safe Enough? eds. Richard C. Schwing and Walter A. Albers, Jr. New York: Plenum Press.

Kahneman, Daniel and Amos Tversky. 1979. “Prospect Theory and Analysis of Decision Under Risk.” Econometrica. 47(2):263-91

Kleiman, Mark. 1993. “Enforcement Swamping: A Positive-Feedback Mechanism in Rates of Illicit Activity.” Mathematical & Computer Modeling 17 (2):65-75.

Kelling, George and James Wilson. March 1982. “Broken Windows: The Police and Neighborhood Safety.” The Atlantic Monthly. Volume 249, [3]: 29-38.

Raiffa, Howard. 1982 The Art and Science of Negotiation, Harvard University Press, Cambridge, Mass.

Schelling, Thomas C. 1978. Micromotives and Microbehavior. W.W. Norton: New York.

Schelling, Thomas. 2003. “Foreword” in Collective Choice: Essays in Honor of Mancur Olson, edited by Jac Heckelman and Dessin Coates Springer, Berlin.

Viscusi, W. Kip. December 1993. "The Value of Risks to Life and Health," Journal of Economic Literature, Vol. XXXI, (4):1912-1946.

CHAPTER 2

CRIME DESPITE PUNISHMENT

A simple economic model of crime would hold that crime is only committed when the threat of punishment is insufficient to outweigh the benefits of offending. That would follow immediately from the assumption that offenders behave perfectly rationally.

If that were so, crime could be eliminated, or nearly so, by increasing the threatened punishment, in severity or probability or both. ("Nearly so" if there is some upper bound – institutional or technical – on the severity of punishment, and if some offenders prize the commission of the offense more than they fear that punishment.)

Thus, in situations where criminal activity is open enough to be easily observable and in which offenders are perfectly rational, it should be possible in principle, by concentrating enforcement resources, to "tip" from high-violation equilibria to low-violation equilibria, pushing past the hump of the inverted U of total punishment. In such cases a low violation rate, which brings as a secondary benefit a small total amount of punishment inflicted, will also bring low budgetary costs. (By contrast, for crimes that are harder to observe, the residual surveillance costs might be high, perhaps too high to be worth bearing for the

reduction in the offense rate. The expense of surveillance is part of what makes enforcement of the drunk driving laws so difficult.)

That leaves open the question of *how much* punishment would be enough to greatly reduce offending. Treating the rewards of successful offending and the penalties for being caught as if they were monetary values helps make one element of the structure of the problem easy to see, but at the cost of creating a big gap between the world of the model and the world of actual crime and punishment. What happens when the typical punishment is measured, not in dollars, but in time behind bars?

Perhaps we should think of time behind bars as a potential cost of offending, and then ask how much of that cost a rational offender would be willing to accept in return for the gains from a successful crime. For income-producing offenses, we can compare the risk (measured in days, months, or years) to the offender's potential financial gain (measured in dollars).

Doing that analysis for some of the more common income-producing crimes yields a surprising result: current punishment levels in the United States already appear to be high enough to make a career in crime thoroughly unattractive, even for someone whose licit alternatives are not very attractive. To that extent, brute force – the attempt to control crime by increasing punishment – has, if not failed, at least

underperformed what might have been expected of it in a world of rational offenders.

The continued frequency of such crimes as residential burglary and street corner crack dealing in the face of what seem to be more-adequate levels of punishment therefore constitutes a puzzle, and we can ask what features of the offenders, and of the actual criminal justice system, might account for the anomaly, and whether any of them could be changed to achieve in practice crime rates closer to the low ones that the theory indicates "ought to be" the result of current punishment rates.

The outlines of the risk-reward calculation are fairly straightforward. Aside from whatever intrinsic gains there are from committing a successful residential burglary (what Jack Katz calls the "existential pleasures of crime," [Katz, 1998]. including the thrill of a risky activity, the satisfaction from the exercise of skill, the feeling of superiority over the victim, and bragging rights among one's acquaintances, including potential sexual partners), the burglar can expect to earn some amount of money. On the other side of the equation, in addition to whatever other risks the activity might have – being wounded or killed by a homeowner, for example, or the short-run embarrassment and long-run career consequences of a brush with the law – the burglar risks spending some amount of time locked up.

Evaluating the different risk-and-reward prospects confronting different potential burglars would be hard. But evaluating the average

risk and the average reward is relatively uncomplicated. Within the limits of measurement error, the victimization surveys and crime reports provide an estimate of how many burglaries occur each year, and also of the losses experienced by homeowners from those burglaries. The prison and jail surveys indicate how many people are behind bars at any given moment for those crimes. Combining those numbers allows a computation of the risk-reward profile of the average residential burglary, measured in dollars gained versus days spent confined.

In 2001, there were just under 2,110,000 burglaries reported to the police [FBI, 2001]. At year-end 2001, there were approximately 208,000 persons behind bars for burglary (7.6% of the 631,000 jail inmates and 12% of the 1,334,000 prison inmates) [Bureau of Justice Statistics, 2001, 1996, 1991]. That implies just under 76 million prisoner-days per year served for burglary (365 x 208,000). Dividing 76 million prisoner-days by 2.11 million burglaries gives an average of 36 days served per burglary committed. Of course, that figure is a probability-weighted average, rather than a typical result; most burglaries lead to no punishment at all, while a relative few lead to substantial prison terms.

On the earnings side, the FBI estimates the average loss in a residential burglary at $1381 [FBI, 2001]. Since stolen property is worth substantially less than the same property without a cloud on its title, the gain to the burglar is substantially less than the loss to the victim; in addition, the gain is often shared with a receiver of stolen property, or

"fence;" those individuals notoriously pay only a small fraction of the value of the goods they accept. If the gain to the burglar is 20% of the loss to the victim – the actual figure will be higher for cash and weapons, lower for, e.g., television sets – the average burglar gains something between $250 and $300 from the average residential burglary.

Thus burglary yields less than $10 in "earnings" to burglars for each day they spend behind bars. When the terms are stated that bluntly, burglary does not appear to be an attractive proposition; it seems implausible that many of those who commit residential burglaries would accept a straight offer to spend a day in jail in return for $10. Even throwing in the "existential pleasures," it is hard to believe that an offer of the thrill of a burglary, plus something less than $300, in return for 36 days behind bars (plus the other costs and risks attendant on housebreaking) would find many takers.

(Since burglary is sometimes a group crime, the average risk of punishment for an individual committing a burglary is less than 36 days, but by the same token the reward is proportionally smaller. The ratio between average risk and average reward is unchanged. The calculation omits punishment within the juvenile system while including the takings of juvenile burglars, and implicitly treats probation and parole as not involving punishment at all; to that extent it understates just how bad a deal burglary actually is.)

Thus, in this case at least, the goal set out by the proponents of a deterrence-based approach to crime control – making crime unattractive by making the punishment greater than the reward – would seem to have been accomplished. And indeed, since 1975, [Wilson, 1975] as the expected punishment per burglary has grown from 6 days to 36 days, the burglary rate has fallen about 50%, a trend broadly consistent with the idea that crime obeys the Law of Demand. (As outlined in the previous chapter, the downtrend in burglary rates is, in purely computational terms, part of the explanation in the rising punishment per burglary; any given amount of punishment is being stretched over a smaller number of offenses in computing the expected value.)

The statistics for retail drug dealing are, or ought to be, comparably daunting. The retail cocaine dealers in Washington, D.C. studied by Reuter and MacCoun took in about $30 per hour of dealing (plus the cocaine they used but did not pay for). In a year of full-time dealing, those dealers could expect to spend about three months, on average, behind bars. Given the risks of being shot and of becoming addicted to crack, those dealers were not making an especially wise set of choices, judged from outside, but the monetary rewards of the trade were real enough.

More recent studies, however, show that wages in the cocaine trade have fallen to below the Federal minimum wage (Levitt and Venkatesh 2000); while the risks of violence have diminished, the risk of

imprisonment has risen substantially. Yet wages (and retail cocaine prices) have fallen sharply, even in the face of an increased threat of punishment (Walsh 2004).

So burglary and retail crack-dealing seem to be very poor career choices, even for those with unattractive alternative uses for their time and energy. Yet burglary and crack dealing persist. Why?

No doubt some burglars, more skilled than others, face smaller-than-average risks for greater-than-average rewards. But that means that the rest of the burglars face odds even worse than those computed above. Even if burglary were rational for some high-skilled burglars, the persistence of the balance would remain puzzling.

Misperception of risks and rewards

One possible explanation for continued crime in the face of punishment is that offenders misperceive the risks they face and the rewards they can expect. Burglars, for example, might not know the average risk and reward of a burglary, or they might not know how the risks and rewards they would face as burglars compare with that average.

Such misperceptions, if they turned out to exist, would hardly be surprising. There might, for example, be perceptual lag. The risks of burglary have been increasing steadily for a generation, but folk-wisdom among potential burglars might not have fully adjusted to the fact that

the expected punishment for burglary is six times what it was a generation ago.

Like other gamblers, burglars may tend to talk more about their successes and their failures, leading those who listen to them to overestimate the gains and underestimate the risks. Moreover, a burglar who is arrested and imprisoned is no longer around his neighborhood to talk about it, which may make instances of successful burglary more psychologically "available" to potential burglars than instances of unsuccessful burglary.

Even if the risk-reward mix of the average burglary were well known among burglars – and there is no particular reason why it should be – there are several well-established psychological phenomena that might lead someone contemplating burglary to think that his individual odds were better than average, even if they are not.

The "optimistic biases" include the belief in luck, the illusion of control, and self-confidence bias. In the laboratory and in surveys, people on average tend to believe that luck favors them (i.e., in a situation they know to be purely chance, that their chances of winning are higher than average). Moreover, they tend to think that they can exert some control over what are in fact chance events. The illusion of control matters in part because there is a widespread conviction that one's own skills are unusually sharp: 90% of drivers, asked to estimate

the risks facing "the average driver" and the risks they personally face, will attribute to themselves less-than-average risk. (This last effect, formally referred to as "self-confidence bias," is sometimes called the "Lake Wobegon paradox," after Garrison Keillor's village where "all the children are above average.")

Even having been caught and punished before may be insufficient to disabuse someone of such optimistic biases. Since more than 85% of burglaries never lead to arrest, any actual arrest must necessarily be the product of unusual circumstances: some combination of random bad luck and poor planning or execution. Given the human predilection towards unjustified optimism, it may be easy for burglars to dismiss previous failures as unlikely to happen again, either because they resulted from random circumstances unlikely to recur or because they resulted from mistakes not to be repeated. The reflection that, given enough burglaries, either bad luck or a new mistake is a virtual certainty requires more statistical insight than most burglars may be able to summon.

The relative rarity of arrest could also contribute to perceptual lag; when the base probability is relatively low, it can be difficult to detect even a significant increase if all one has to go one are personal experience and word of mouth.

On the other hand, misperceptions can work in both directions; offenders may overestimate some threats even as they underestimate others. In particular, very-low-probability threats tend to be overestimated, while moderate-probability threats are underestimated; in some experimental circumstances, decision-makers will treat both a risk of one in a thousand and a risk of one in ten as if they were close to one in a hundred. [Allais1979; Ellsberg 1961; Hey 1995; Starmer 2000)

The valuation of uncertainty

If the expected value of punishment for a burglary is greater than the expected value of the gain, and yet burglaries continue, that might be because of the variability concealed behind the expected-value calculation. It is not, after all, the case that each burglary yields $280 to the burglar and results in 36 days' confinement. Both the yield and the punishment vary from incident to incident.

To start with, the typical burglary, as opposed to the average burglary, yields no punishment whatever, because the offender is never caught; in recent years, the FBI data show about one-seventh as many burglary arrests as burglaries known to the police in any given year. Nor does every arrest lead to a conviction, or every conviction to a sentence of confinement. For his average gain of less than $300, a burglar faces, not 36 predictable days behind bars, but something like one chance in twenty of a prison term lasting an average of about two years. That is, for each two-year prison term served, the average burglar will enjoy

about $6000 in criminal "earnings." While it seems implausible that anyone would be willing to spend a day in jail for something less than $10, the offer of $6000 (now) for spending two years in prison (sometime later) might well find takers.

That uncertainty can reduce the effectiveness of deterrent threats is far from a new idea; it goes back at least to Beccaria. It is a maxim that certainty of punishment is more important than severity in determining deterrent effectiveness. So perhaps the randomness of the punishment for burglary that keeps the activity attractive in the face of what would seem to be a net negative expected value.

But from the viewpoint of a rational actor as economists understand that concept, that phenomenon is puzzling. Uncertainty should in general decrease the attractiveness of an option rather than increasing it. (That is, uncertainty of punishment ought to make it more frightening rather than less so.) At least, that should be the case for monetary penalties; the diminishing marginal utility of income means that bigger losses are felt more sharply per dollar lost than smaller ones. The diminishing marginal utility of income leads, rationally, to what economists call risk aversion: the willingness to sacrifice some expected value in return for a reduction in uncertainty, is the principle that explains the existence of insurance markets. In order for rational burglars to accept a gamble whose expected value is negative, they would have to be, not risk-averse, but rather the opposite: risk-seeking.

That might be the case. The prevalence of gambling, which is the opposite of insurance, seems to show that under some circumstances people voluntarily seek out risk, and it is not hard to see how "thrill-seeking" in the psychological sense – known to be a characteristic more common among offenders than among non-offenders [Grasmick et. al, 1993; Longshore, 1998; Eysenck, 1989; Herrnstein, 1983; White et al., 1994; Wilson & Herrnstein, 1985). – might lead to risk-seeking in the economic sense, despite the fact that risk-seeking fails to maximize expected utility.

Another possible explanation does not require assuming any subjective taste for risk. Perhaps some offenders' choices are influenced by the phenomena described by prospect theory: [Kahneman & Tversky, 1979, 1992] perhaps, that is, they focus on whether each individual transaction produces a gain or a loss, thus overweighting both small gains and small losses relative to large ones. That behavior proves, in some laboratory conditions, to be much more prevalent than subjective thrill-seeking, and seems to be one explanation, for example, for the fact that most individuals' stock-market behavior does not in fact minimize risk for any given level of expected reward.

Or perhaps it is a mistake, descriptively, to treat prison time as a quantity that can be averaged out, as monetary losses can be averaged out: to assume that is, that (before risk-preference is taken into account) a two-year sentence starts out being twice as bad as a one-year sentence.

Losing $100 twice is certainly the same as losing $200 once, and (again, ignoring risk-preference) twice as bad as losing a hundred dollars only once; but are two one-year sentences the same as one two-year sentence, and twice as bad as one one-year sentence? It is not obvious that they are, considering both to the fixed costs of arrest, incarceration, and the subsequent criminal record – costs that increase only slightly with the length of the term – and the psychological phenomena of acclimation and duration invariance. If so – if two years behind bars is much less than twice as bad as a year behind bars, and much less than twenty times as bad as 36 days behind bars – then a burglary that a given offender would find not worth committing at the cost of 36 days locked up might still seem worthwhile to him at the cost of one chance in twenty of two years locked up, which is more or less the deal on offer to the average burglar.

The fixed costs of being incarcerated would give someone without a prior criminal record very strong reasons to prefer an even-money chance of two years locked up (with the alternative being to go off scot-free) to the certainty of a year locked up. The embarrassment and stigma of arrest and trial, the disruption of family relationships, housing and employment arrangements, and other parts of the life routine, the subsequent difficulty of finding employment as an ex-convict, the stress of entering prison and the stress of leaving it to return to an uncertain future: all of these result from going to prison at all and vary relatively

little with the length of the term. The same is true, albeit to a lesser extent, for someone who does have a prior criminal history, or even prior experience of prison. Some of the fixed costs are already "sunk," but not all of them. Thus there are reasons to think that the deterrent impact of incarceration is subject to diminishing marginal returns at the intensive margin: that, for example, the second year of a two-year sentence packs less deterrent punch than the first year, and that it is therefore rational for offenders to act in a manner that seems to embody risk-seeking.

But while fixed-cost effects support the idea that certainty is more important than severity in determining the deterrent impact of punishment – the idea that, for any fixed amount of imprisonment for a given crime, the total deterrent effect would be increased by increasing the probability of incarceration while decreasing the average sentence length – there is a conceptual difficulty in using fixed effects to explain away the apparent irrationality of household burglary and similar crimes. Even ignoring those fixed costs and simply comparing a twenty-four hour period spent behind bars to a twenty-four hour period spent on the street, it is hard to imagine many people accepting an offer to spend that period confined for a payment of $10. (Or, to think about it a slightly different way, it is hard to imagine someone about to leave prison being willing to stay an extra day in return for a $10 payment.) If the gain from a crime cannot cover the *variable* costs of incarceration, then the fixed

costs simply stand as additional reasons not to commit the underlying crime.

But even on a day-to-day basis, all days in prison may not be equally aversive. Studies of self-reported happiness show that a wide range of people have astonishing capacities for acclimating themselves to what seem in prospect, or from the outside, like virtually intolerable conditions. Quadriplegics, for example, report themselves as only moderately less happy than people with full use of their limbs. Insofar as prisoners succeed in psychologically acclimating themselves to the prison environment, much of the aversive power of the prison experience may inhere in the first few days, weeks, or months.

Acclimation may be part of the explanation for another phenomenon, called by its discoverers "duration neglect" or "duration invariance." (Fredrickson and Kahneman 1993). It might seem logical that someone's overall evaluation of an experience, good or bad, would be something approximating the integral over time of the moment-to-moment evaluations: that is, to the average value of a moment during that experience multiplied by the duration of the experience. But that turns out not to be the case.

Whether people are asked to give a numerical rating of how good or bad an experience was, or asked instead, having experienced it once, to make a choice between experiencing it again and some fixed alternative,

the answers they give turn out to be sensitive to the most intense moment of the experience and to its end, but virtually not at all to its duration. In fact, the overall aversiveness of an unpleasant experience can actually be reduced by adding to its end an additional period of somewhat reduced, but still noticeable, unpleasantness.

The studies to date deal with durations in minutes or at most hours, not months or years; perhaps duration neglect is less prominent over longer time periods. But if the phenomenon does extend to periods of months or years, then incomplete success of the attempt to increase deterrence by increasing sentence length is not at all surprising.

Duration neglect, even if it were present and virtually complete, would not utterly defeat that project; the plea-bargaining process demonstrates that offenders do care about the length of their sentences in prospect. But duration neglect might mean that a former prisoner's memory of how bad a spell behind bars was, and consequently his willingness to repeat the behavior that produced that spell – and the fear his recollected accounts of imprisonment conveyed to his acquaintances – might not vary much with sentence length. That provides one more reason to think that the total deterrent effect of a smaller number of longer sentences would be less than the effect of a larger number of smaller sentences amounting to the same aggregate of prisoner-days. If so, the expected-value calculation will tend to overestimate the deterrent effectiveness of our actual system of randomized severity.

Timing

The punishment for burglary occurs later in time than the financial gain from burglary; that results not only from the delay between crime and punishment occasioned by the legal process, but also from the randomness of the arrest and prosecution process. Except for the unfortunates who get caught the very first time, burglars get to enjoy some of the gains from their activities before ever getting caught. Even when a burglary leads to arrest, the arrest may not lead to any time behind bars; that increases the average time-lag between the stream of gains from burglary and the stream of losses. In addition, unlike a fine or corporal punishment, a term of confinement is by its nature extended in time; the average day of a two-year sentence is a year after its beginning.

The time lag between gain and loss requires an adjustment in the calculation of the balance between the gains from crimes such as burglary and the risks the criminal justice system imposes on burglars.

Even a strictly rational actor prefers present to future gains, and future to present losses. When the gains and losses come in monetary terms, the reason is obvious: money now can be lent out or invested, earning interest, while money that arrives later must, to be used now, be borrowed, paying interest. Thus the actor's capacity to borrow and lend – the interest rates he can collect or must pay – create lower and upper

bounds on the "discount rate" a rational actor uses in comparing present and future gains and losses.

Given a choice between $100 today and $120 (with certainty) a year from now, someone who can borrow at 15% on a credit card would be foolish to take $100 now; better, if the money is needed now, to wait a year for $120 and borrow $100 at 15% in the meantime; when the year is up, the loan can be repaid with $115, leaving the borrower $5 ahead of where he would have been taking the $100 immediately.

By the same token, someone who can earn 3% on invested funds would be foolish to take $102 a year from now in preference to $100 today.

Since all of us face continual opportunities to trade money for pleasure (or the avoidance of unpleasantness) or to save or earn money by denying ourselves pleasures or accepting unpleasantness, consistency dictates that even non-monetary gains and losses be discounted at the same rate; someone who can trade an hour of leisure for $20 by working overtime, and could do so either now or a year from now, ought to discount leisure at the same rate he discounts monetary gains and losses.

If offenders used financial-market rates of interest to discount the value of punishment, the time-lag could not account for much of the anomaly. Even if the average punishment-day occurs two years after the

average financial gain from burglary, at a discount rate of 20% compounded the value of punishment would be reduced, and thus the present-dollar gain per hour spent in prison increased, by 44%. As large as that adjustment is, it cannot really make sense of the activity; instead of spending time behind bars for illicit gain amounting to 34 cents per hour confined, a 44% discounting adjustment makes the figure 47 cents per hour confined. That still does not add up to a sensible decision.

However, the typical burglar may not be able to borrow on credit-card terms; he may be unable to borrow at all, or face the very high interest rates charged by loan sharks. At 20% per month – not an especially high interest rate by loan-shark standards – a two-year delay would require an adjustment by a factor of almost 80, making the earnings per discounted confinement day hundreds of dollars: not obviously an unreasonable proposition. Thus very high discount rates could fully explain the puzzle of continued crime in the face of apparently ample punishment levels.

That explanation, however, merely pushes the puzzle up one level. The interest rate at which one can borrow sets an upper bound on one's personal discount rate, not a lower bound. Even those unable to borrow at reasonable rates do not uniformly discount the future so heavily compared to the present, whether by borrowing at usurious rates or by committing residential burglary or other crimes with high risk-to-reward ratios. Facing a high interest rate as a borrower is a social fact; acting

on a very high personal discount rate is, generally speaking, simply another label for improvidence. (It is possible to construct examples where a very high discount rate would be perfectly rational, because delaying an expenditure would have disastrous consequences; it is not irrational to borrow from a loan-shark if the alternative is starvation. But that is not the usual case.)

In laboratory experiments, it is possible to measure discounting behavior, and there is evidence that very high discount rates are correlated with criminality. (Whether this is the same phenomenon as the psychological trait of impulsiveness remains an open question.) There is evidence from other studies that personal discount rates go up (and measured intelligence goes down) under various forms of social stress, and especially social exclusion. This may reflect an evolutionarily supported mechanism that suspends concern for the future when the immediate present is sufficiently threatening. In social environments sufficiently unpredictable to discourage any sort of long-range planning, a radical improvidence and present-orientation may seem to be the only sensible stance to take.

Another line of laboratory studies has shown that, under appropriate conditions, most people (and other organisms) will demonstrate behavior inconsistent with any fixed discount rate, giving great weight to the present compared to even the near (and predictable) future. [Ainslie, 1991] While even a very high discount rate is consistent with rationality

under some circumstances, such "hyperbolic discounting" is strictly irrational, since decisions made under its influence are not even consistent over time; decisions are taken that, in Schelling's phrase, are "deprecated in advance and predictably regretted in retrospect." (Schelling 1984) Procrastination, breaking a diet, and scratching a rash reflect hyperbolic discounting over different time-intervals, and the ubiquity of such behavior suggests that irrationality about the future is built fairly deeply into the normal human psyche. Sometimes these seem subjectively like choices; in other cases, temptation or fear seems to override the mechanism of rational choice entirely, in the phenomenon Aristotle called "akrasia," or "weakness of will."

That time-inconsistent discounting is a characteristic of the situation as well as the person is shown by the observation that, even in pencil-and-paper questionnaires, frequent heroin users make more present-oriented choices when asked questions about heroin (e.g., "Would you prefer a bag of heroin now or two bags tomorrow?" than about money ("Would you prefer $10 now or $20 tomorrow?"). [Bickel et al., 1999] Why some circumstances and decisions bring it out more than others remains an open question. George Loewenstein points convincingly to pain, fear, hunger, addiction, and sexual desire among the "visceral influences" that defeat the attempt to make rational choices between present and future, though those factors seem to play relatively little role in procrastination, for example. [Lowenstein 1996]

Whether irrationally strong present-orientation is best described as discounting at a very high rate, as discounting hyperbolically, as giving in to visceral influences, or as weakness of will, its potential to disable the mechanism of deterrence makes it a threat to any crime-control effort organized around the threat of punishment, especially delayed punishment.

The same conclusion follows if instead of imagining potential offenders as weighing risk against reward we think about adjusting behavior to the threat of punishment as a kind of learning process. In that context, the observation that delay reduces the efficacy of punishment is completely unsurprising. Cause-and-effect relationships are much more salient psychologically when the effect follows almost immediately on the cause. (Think about the difference between learning to type with a computer program that rings a bell for every error and learning to type when the feedback is delayed even a few minutes.)

However, recent studies by Daniel Nagin and his colleagues have shown that for one – perhaps atypical – group of offenders, college students facing jail time for driving under the influence, punishment is actually discounted *negatively*: the subjects of that study apparently wanted to "get it over with" and be able to resume normal life. [Nagin & Pogarsky, 2001]

Compared to burglars, college-student drunken drivers probably have more normal lives to resume; an overhanging prison term may be less of

a disruption to an already disordered life. Moreover, the college students were dealing with sentences that they knew were going to be carried out, with timing the only unknown factor. Waiting for a punishment that is certain to occur may generate much more anxiety than arises when a punishment is not only delayed but also uncertain. In the situation facing burglars, delay and uncertainty are so intertwined that it may be hard to distinguish empirically between uncertainty effects and discounting effects.

Implications for crime control

All of this has implications for our imagined designer of crime control regimes. Confronted with perception, uncertainty, and time as threats to the effectiveness of deterrent threats, what are you going to do? The obvious answers are to make the risks of crime to criminals easier for them to perceive, reduce the uncertainty, and reduce the time-gap between offense and punishment. All of that, of course, is easier said than done.

There is good evidence that many offenders do not accurately grasp the risks they face, especially when the rules have change relatively recently. The Draconian punishments available under the federal Armed Career Criminal (ACC) statute, for example, were for the first several years of the statute's existence almost completely unknown to the people subject to it. The same seems to have been true in jurisdictions where

federal prosecutors selected some retail drug-dealing cases, sometimes virtually at random (in Manhattan, for example, the U.S. Attorney's office for several years prosecuted arrests where the arraignment fell on a Tuesday). Defense lawyers report that their clients were utterly bewildered to be facing five-year mandatory sentences for offenses that would otherwise have led to at most jail time. Obviously, a completely unknown deterrent threat does not deter. How to make the threats known is a different question.

But that question presupposes that making those threats known is part of the task of law enforcement, an idea without much currency. It might be instructive to survey groups of people who might engage in burglary (selected either by criminal history or by other personal characteristics predictive of a high probability of burglary arrest in the future) to see what they think the odds are. If in fact they generally underestimate the risks and overestimate the rewards, you could then consider what sort of marketing approach might bring their beliefs into closer alignment with the realities.

In other cases, the goal might to be encourage misperception, or at least to leave it undisturbed. If, for example, California's offenders overestimate the risk they face of being sentenced under California's "Three Strikes" law, that overestimation will tend to increase the law's deterrent impact. The law's memorable name, and the press accounts of the (fortunately, relatively rare) absurd injustices done under it – as in

the infamous incident where someone with an unspectacular criminal history was sentenced to 25 years to life for stealing a slice of pizza – contribute to giving "Three Strikes" a saliency out of all proportion to the actual frequency of "third strike" sentences, especially in California's more populous counties.

An alternative to mass-market communications – such as the advertising campaign surrounding Virginia's "Project Exile" crusade against gun possession by persons with violent criminal histories – is direct communication to a list of high-risk potential offenders.

Insofar as the uncertainty of punishment reduces its deterrent efficacy, the obvious solution would be to reduce that uncertainty. Reducing uncertainty implies spreading any given stock of punishments more evenly over the group of eligible offenders. That could be done at either or both of two stages: by increasing the probability of arrest (and perhaps the probability of conviction given arrest, for those arrestees who are in fact guilty of the underlying crime) and by increasing the number of sentences involving some non-trivial punishment. If the only non-trivial punishment available is confinement, that means increasing the number of confinement sentences and thus necessarily (if we hold the total amount of confinement constant) reducing the average sentence lengths for those sentences not of length zero.

Alternatively, more use could be made of punishments other than confinement, including fines, restitution, and forced unpaid punitive work (now usually called, euphemistically but largely inaccurately, "community service"). In practice, fines and restitution payments have proven difficult to collect; moreover, they risk inducing those who must pay them to return to crime to secure the requisite funds. "Community service" has also proven harder in practice than it seems in concept, partly due to the difficulty of finding appropriate assignments and partly due to the difficulty of getting unpaid workers to show up and to work productively.

Probation supervision, with its rules, restrictions, and reporting requirements, is also a form of punishment, though probation offices actually administer a mix of rules and services that partly obscures the punitive nature of the relationship; many offenders reportedly regard a probation term as virtually equivalent to no sentence at all. Another approach to increasing the probability of non-trivial punishment for any given offense would be to beef up probation supervision to make it seem non-trivial to those subject to it. Currently, probation supervision costs, on average per person, less than 5% as much as imprisonment. Even if making it more "real" meant quadrupling the per-capita cost of probation, it would be possible within a fixed corrections budget to administer community-based punishment five offenders for each one released from prison. Further economies could be achieved by

shortening probation terms; a month of truly intensive probation supervision might easily prove more aversive, and therefore a better deterrent, than two years of the merely nominal supervision currently characteristic of probation.

Any punishment that involves the voluntary (even if unwilling) compliance of the person punished needs to have behind it a back-up punishment for non-compliance, and a mechanism for detecting non-compliance and adjudicating and administering those back-up sanctions when needed. Creating that back-up sanctions system is among the neglected tasks in the field of "community" (i.e., non-incarcerative) corrections.

Increasing the probability of arrest for offenses overall requires either hiring more police or somehow changing police practices to increase the number of (valid) arrests per officer-day. Increasing the probability of arrest for any given crime type, or for any given crime type in any given neighborhood, can be accomplished by concentrating resources. Whether that change in probability will be noticed by offenders in a way that leads them to reduce offense rates is a different question. One advantage of announcing and enforcing focused zero-tolerance programs is that doing so is a means of ensuring that a policy change that increases the objective probability of arrest will change the subjective probability as well. At the other extreme, there may be value in making sure that every offense type occasionally leads to arrest, on the principle

that the subjective difference between zero risk and a tiny risk is sometimes behaviorally important.

Reducing the variance in punishment by leveling out the sentencing regime might or might not be worth its costs, which would include an increase in the number of people carrying the social disabilities that go with being ex-prisoners, a reduction in the capacity of the criminal justice system to make appropriate distinctions among different levels of culpability, and perhaps a reduction in the crime control achieved through incapacitation, insofar as the current system succeeds in locking up for relatively long periods a subset of convicted offenders whose severity-adjusted personal crime rates are, on average, higher than the severity-adjusted personal crime rates of those convicted but not confined, or confined for only short periods. Whether the possible increase in deterrence would be enough to justify those costs would depend in part on whether a feasible level of change in the probability that an arrest leads to a jail or prison term would be noticed by offenders sufficiently to change their behavior. A change to more certain sentencing might be expected to be more effective in combination with a change to targeted zero tolerance on the policing side than it would be alone.

Insofar as the efficacy of punishment in shaping behavior is diminished by delay, the prescription would seem to be reducing the time-gap between crime and punishment. As noted above, increasing the

probability of punishment has such a reduction as a side-effect, so the measures discussed under that heading will do double duty. But there are also ways to increase the average celerity of punishment, even holding the probability fixed.

In principle, arrest, booking, and the time spent locked up before bail is set or release on recognizance granted are not supposed to constitute punishment, since the arrestee, not having been convicted of anything, is still legally presumed to be innocent. Practice is otherwise. Arrest is naturally unpleasant, and police and jailkeepers have developed practices to make it even more unpleasant than it has to be. (Handcuffing, for example – nominally a security measure – is deliberately made both humiliating and uncomfortable, and it is applied to arrestees who obviously pose no real risk of flight or resistance. Pre-arraignment holding facilities, which hold the presumptively innocent, are frequently dirtier, more uncomfortable, and more dangerous than the prisons to which the convicted are sent. The fact that courts are closed for the weekend converts a Friday arrest into a three-day jail sentence without trial, and some police officers make strategic use of that situation.)

Officially, these facts are deprecated when they are not merely denied. Yet the analysis above suggests that these swift and unofficial punishments may carry with them a large portion of the total deterrent effectiveness of the entire criminal justice process. Thinking of arrest as

merely the prelude to trial (or, in the typical case, guilty plea) and formal punishment may be a mistake: arrest alone may be an important deterrent.

Insofar as that is the case, the high attrition rate from arrest to incarceration may be less of a problem than it seems at first blush. Increasing the number of police or their arrest activity may have little impact on the distribution of prison and jail terms, because in most jurisdictions there are already more arrests than the courts can process and more convicted offenders than the prisons and jails have room for. That being the case, increasing the arrest rate, either by hiring more police officers or by boosting the number of arrests per officer, might seem to be largely wasted effort. But if arrest and its immediate consequences are doing much of the work of deterrence, perhaps that is not the case.

Another, grimmer implication is that reforms designed to bring enforcement practice more closely into alignment with legal theory, however desirable ethically, might have unwanted side effects in the form of higher crime rates. That would not be a conclusive argument against making such reforms, but it would be an argument worth pondering.

The overall time-gap between offense and punishment could also be reduced by speeding the process of adjudication. (Thus, "speedy trial" rules, imposed to protect defendants' rights, may actually be important

crime-control initiatives.) Some prosecutors' offices offer large sentencing discounts for plea-bargains arrived at before arraignment. That policy is usually described as an attempt to economize on prosecutorial and court resources, but again it may by accident actually increase the total deterrent value of any fixed stock of jail and prison time. In the constant struggle between the defense bar and prosecutors over procedural issues – a struggle waged both in the courtroom and in legislative chambers – perhaps prosecutors should focus less on increasing the probability of conviction and more on making the process move more quickly. In this context, the logic of present-orientation suggests that speeding up processes that already move fairly quickly would have a higher payoff than speeding up those that move slowly: moving from arrest to sentencing in a month rather than two months might make a big difference, while doing so in six months rather than a year might barely matter at all.

In dealing with long prison terms, a good part of the time-lag between the offense and the average punishment-day is unavoidable. That, and the possibility that duration neglect applies to prison terms, constitutes an argument for shortening very long sentences. (In some cases, though hardly all, retribution or selective incapacitation or both with furnish strong counterarguments.) That suggests the desirability of finding ways to shorten prison terms while maintaining, or perhaps even increasing, their deterrent effectiveness. Obviously, there are some very bad ways of

doing so, but there might also be some good ones. As John Stuart Mill pointed out more than a century ago, there may be no very close connection between how aversive a punishment is in behavioral terms – how much, that is, it deters – and how much actual suffering it entails (or, he might have added, how much actual damage it does to the long-run capacities of the person punished). (Mill 1986)

Insofar as what Kahnemann and his colleagues discovered to be true of colonoscopies is true of prison terms – the remembered unpleasantness of the process depends almost entirely on its peak intensity and its intensity near its end, and virtually not at all on its duration – then it ought to be possible to shorten sentence lengths without reducing the deterrent effects on those who actually go to prison and on those who hear about the prison experience from them, though presumably there would still be a loss of deterrence among the broader population. In particular, the practice of moving prisoners to progressively less restrictive (and, presumably, progressively less unpleasant) prison conditions as they approach their release dates may have the unintended result of greatly reducing the remembered disutility of the prison experience. If it were to turn out – as it might, or might now – that a one-year prison term that started and ended with two weeks in solitary confinement was remembered as more unpleasant than a five-year prison term ending with a year on a pre-release "honor farm" and six months in halfway house, that might mean an opportunity to reduce

the expense and the overall cruelty of the current system while increasing is crime-control efficacy. That idea is, on our current knowledge base, entirely speculative, but it embodies the sort of speculation that correctional authorities might well be asked to do more of.

At an even more speculative level, if it turns out that crime in the face of punishment is due in significant part to imperfect rationality – in more old-fashioned terms, to impulsiveness, lack of self-command, improvidence, and imprudence – that suggests that increasing impulse control, self-command, providence, and prudence among actual and potential offenders deserves a place among the techniques of crime control. Doing so positively might – or, again, might not – be beyond the range of our current ingenuity. But it might well be possible to reduce the ways in which the criminal justice system makes the offenders it deals with *less* capable of acting in their own long-run self-interest.

References for Chapter 2

Ainslie, George. 1991. "Derivation of 'Rational' Economic Behavior from Hyperbolic Discount Curves." American Economic Review 81(2):334-340.

Allais, Maurice and Ole Hagen (eds). 1979. Expected Utility Hypotheses And The Allais Paradox. Dordrecht. Holland: Reidel.

Beccaria, Cesare. 1986. On Crimes and Punishments; translated from the Italian in the author's original order, with notes and introduction by David Young. Indianapolis: Hackett Pub. Co.

Bickel, Warren K. (and Nancy M. Petry), "A Behavioral Economic Analysis of Polydrug Abuse in Heroin Addicts," In Frank J. Chaloupka, Michael Grossman, and Henry Saffer (eds), The Economic Analysis of Substance Use and Abuse: An Integration of Econometric and Behavioral Economic Research, National Bureau of Economic Research Conference Report, 1999.

Bureau of Justice Statistics. 1991. Profile of Jail Inmates.

Bureau of Justice Statistics. 1996. Survey of State Prison Inmates.

Bureau of Justice Statistics. 2001. Prison and Jail Inmates at Mid-Year.

Ellsberg, D. 1961. "Risk, Ambiguity and the Savage Axioms." *Quarterly Journal of Economics* 75 (Spring): 643-669.

Eysenck, Hans J. 1989. Personality and Criminality: A Dispositional Analysis. In William S. Laufer and Freda Adler (Eds.) Advances in Criminological Theory , [1]:89-110.

Federal Bureau of Investigation. 2001. "Crime in the United States 2001." U.S. Department of Justice. pp. 44

Fredrickson, B.L., and Daniel Kahneman. 1993. "Duration Neglect in Retrospective Evaluations of Affective Episodes." *Journal of Personality and Social Psychology*, 65: 45-55.

Grasmick, H. G.; C. R. Tittle; R. J. Bursik; and B.J. Arneklev. 1993. "Testing the Core Empirical Implications of Gottfredson and Hirschi's General Theory of Crime." Journal of Research in Crime and Delinquency 30(1):5-29.

Hey, J.D. 1995. "Experimental Investigations of Errors in Decision Making Under Risk." *European Economic Review* 39: 633-640.

Herrnstein, Richard. 1983. Some Criminogenic Traits of Offenders. In James Q. Wilson (Ed.) Crime and Public Policy. San Francisco: Institute for Contemporary Studies.

Kahneman, Daniel and Amos Tversky
1973 Availability: A heuristic for judging frequency and probability. Cognitive Psychology 5:207-232.
1979 Prospect Theory and Analysis of Decision Under Risk. Econometrica. 47(2):263-91.
1992 Advances in Prospect Theory: Cumulative Representation of Uncertainty. Journal of Risk and Uncertainty [5]:297-323.

Katz, Jack. 1988. Seductions of Crime: Moral and Sensual Attractions in Doing Evil. New York: Basic Books.

Levitt, Steven D. and Sudhir Alladi Venkatesh. 2000. "An Economic Analysis of a Drug-Selling Gang's Finances." *Quarterly Journal of Economics* 115 (3):755-789.

Loewenstein, George. March 1996. "Out of Control: Visceral Influences on Behavior." Organizational Behavior and Human Decision Processes 65[3]:272-92.

Longshore, Douglas. 1998. Self-Control and Criminal Opportunity: A Prospective Test of the General Theory of Crime. Social Problems 45(1):102-113.

Mill, John S. 1986. "Speech in Favour of Capital Punishment." *Hansards Parliamentary Debates*, 3rd. Series, 21 April 1968 (London, 1868). Reprinted in *Applied Ethics*, Peter Singer (ed). Oxford: Oxford University Press.

Nagin, Daniel and Greg Pogarsky. 2001. "Integrating Celerity, Impulsivity, and Extralegal Sanction Threats into a Model of General Deterrence: Theory and Evidence." Criminology [39]:865-892.

Reuter, P.; R. J. MacCoun; and P. Murphy. 1990. "Money from crime: A study of the economics of drug dealing in Washington D.C. Santa Monica, CA". RAND.

Schelling, Thomas C. 1984. Choice and Consequence. Cambridge: Harvard University Press.

Starmer, C. 2000. "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice Under Risk." *Journal of Economic Literature* 38: 332-382.

White, J.L.; T. E.Moffitt; A. Caspi; D. Bartusch; D. Needles; and M. Stouthamer-Loeber. 1994. "Measuring Impulsivity and Examining its Relationships to Delinquency." Journal of Abnormal Psychology [103]:192-205.

Walsh, John M. 2004. "Are We There Yet? Measuring Progress in the US War on Drugs in Latin American." *Drug War Monitor* from the Washington Office on Latin America (December).

Wilson, James Q. 1975. Thinking About Crime. New York: Basic Books.

Wilson, James Q. and Richard J. Herrnstein. 1985. Crime and Human Nature. New York: Simon and Schuster.

CHAPTER 3

POLICING AND PROSECUTION

Focusing on crime control

It is now conventional wisdom that the mission of the police is to control crime, and that police chiefs should direct their attention to that end and that the quality of their leadership should be judged largely according to their success or failure in that regard. To some extent, that set of beliefs was always widespread in the general public, among elected officials, and in the mass media; what, after all, is the point of police work if not crime control? Yet it is only recently, since the development of the ideas associated with "problem-solving" or "community" policing in the 1980s, and especially since the spectacular crime decrease that followed the appointment of William Bratton as Police Commissioner in New York, that those seemingly obvious doctrines have received the assent of the scholarly community, or indeed of the police themselves.

The doctrines held in those seemingly more sophisticated circles as recently as ten years ago held that crime rises and falls for largely extrinsic reasons rooted in large social processes (the famous "root causes of crime"). Therefore, on this view, the job of the police is to respond swiftly to citizen reports of crimes in progress and crimes completed and to enforce the law even-handedly, with as little intrusion on civil liberty as possible and staying as closely as possible within the

procedural strictures laid down in the Constitution and by statute. Police - still in this view - should invest no effort in what was seen as the largely vain attempt to actually reduce the victimization risks faced by ordinary citizens.

Despite arguments that prosecutors' offices should direct themselves to the same end (Kleiman, 1987) the currently received view of the prosecution function resembles the previous view of policing. Prosecution is widely regarded as a quasi-judicial activity, in which judgments about which cases to take and how hard to push them ought to be made on grounds of justice rather than expediency. Politically, there is a widespread demand (which prosecutors' offices, mostly led by elected officials or by appointees with electoral ambitions, have been more than willing to meet) that prosecutors be "tough," but next to no demand that such "toughness" have demonstrable impacts on the crime rate.

And, indeed, no one has proposed prosecution strategies likely to have the sort of impact on crime that some policing strategies appear to have had; if prosecution strategies can influence the crime rate, their influence would be more subtle and indirect than that of new approaches to policing.

On the other hand, it is undeniable that the allocation of prison cells among offenders and offense types depends largely on decisions made by

prosecutors, and that prosecutors' power has grown with the introduction of guideline or mandatory sentencing (which moves discretion from judges to prosecutors) and the elimination of discretionary probation. Unless it is true for some reason not yet articulated that the crime rate is utterly insensitive to how those prison cells are allocated, then it must be the case that prosecutors' decisions can influence the level and distribution of crimes, by (jointly with the police, and within the limits created by the laws and by the remaining discretion of judges) setting the punishment-prices paid for various kinds of offending. Ninety-five out of a hundred criminal cases that result in a finding of guilt are never taken to trial, but are settled through a negotiated plea. Thus the amount of punishment most offenders face depends, in the first instance, on the amount the prosecutor insists on, though how hard the prosecutor can bargain depends in part on what sort of sentence a judge will be inclined to impose if the case is one of the rare ones that goes to trial. Surely some possible solutions to that set of decision problems will do more to control crime that others. At a minimum, prosecutorial follow-through is essential to the success of some police initiatives.

Taking the police and the prosecutors together, then, how should they craft policies and operations to reduce the criminal threats faced by the people they are supposed to protect and serve?

Brute-force incapacitation

One answer is that the police should try to arrest as many serious criminals as possible, and that prosecutors should try to get as many of them as possible imprisoned for as long as possible, concentrating on those who have committed the most serious crimes. That is, more or less, the story of the past quarter-century, in which the population behind bars has more than trebled. Whatever its deficiencies in terms of deterrence, that approach has the capacity to reduce crime through the purely mechanical effect of incapacitation: while they are locked up offenders cannot commit new crimes against ordinary residents, no matter how much they may threaten one another or their guards. Sheer incapacitation effects no doubt account for some portion – perhaps as much as a quarter – of the crime decline of the past fifteen years or so, perhaps as much as a quarter of it: a decline that has been substantial, though uneven by region, by demographic group, and by crime type (Spelman, 2002).

However, that approach may be near, or at, the limit of its effectiveness. Under fiscal pressure, the states seem to be losing the appetite for prison-building that proved so voracious in the 1980s and 1990s and that has brought the country to so astonishing an incarceration rate both in historical terms and by comparison with other advanced societies. The U.S. incarceration rates in 2003 was 714 per

100,000 residents, five times as high as any other advanced democratic country. (Sentencing Project, 2004).

In addition, incapacitation *via* incarceration is subject to the law of diminishing returns. Insofar as police, prosecutors, and judges succeed in catching, convicting, and sentencing more serious offenders rather than less serious ones for any given level incarceration, then as the number of people sent to prison grows the prisons fill with less and less serious offenders. (If you take the cream first, what is left is skim milk.) If imprisonment is expanded instead by giving longer sentences, then the sentences are more and more likely to extend beyond what would otherwise have been the serious criminal career of the prisoner, thus providing no additional incapacitation effect. (Philip Heymann calls this the policy of "providing retirement homes for former armed robbers.")

Selective Incapacitation and its Limits

Some offenders are far more worth incapacitating than others, because they commit some combination of more crimes and more serious crimes than average each year they are not in confinement. (The technical term is "severity-weighted personal crime rate.") (Moore et al, 1985). The differences among offenders are dramatic; even among the select group sent to prison, the 10% most active offenders account for more than half the total offenses (Chaiken and Chaiken, 1982). The obvious policy implication of that finding was that the incapacitative crime control benefit from any given level of incarceration could be

multiplied if it were possible to identify the most dangerous offenders and single them out for long prison terms (Greenwood and Abrahamse, 1982).

Unfortunately, that turned out to be a very big "if;" an offender's criminal history as reflected in official records turns out to be only modestly useful in predicting his future criminality, largely due to the relatively brief course of a typical criminal career; by the time an offender has racked up a serious enough record to justify a long prison sentence, he is often in the declining phase of his career. The longer he stays in prison, the older he gets, and therefore (on average) the fewer and less serious the crimes prevented by each addition year of incarceration. Indeed, "repeat offender" statutes can actually reduce the incapacitative effect of imprisonment by shifting prison cells from younger, more active offenders to older, less active ones. (There has always been a mismatch between the age of prisoners and the age of active offenders; fifteen years ago, it was estimated that the median age of someone committing a serious crime was 23, while the median age of the prison population was 30 (BJS, 2000). In the meantime, serious offenders have been getting younger, with the upsurge in youth homicide, while various "tough-on-crime" sentencing laws and practices have been making the prison population older.) Even a system fully optimized for selective incapacitation would only improve current performance on that dimension by a few percent, (Spelman, 1994) even ignoring the potential

loss in deterrence value from giving out fewer, longer sentences rather than giving shorter sentences and more of them.

But that analysis considers only the use of selective incapacitation as a guide to sentencing; it still makes good sense for the police, as one element of a crime control strategy, to identify and try to make cases against the most dangerously active offenders in their jurisdictions. Some police departments have formalized that approach by keeping a list of especially dangerous offenders (sometimes called the "Varsity" or the "A Team") who are singled out for special investigative attention. Some prosecutors' offices have formal policies of prosecuting "A Team" offenders to the maximum extent possible rather than allowing them to plea-bargain or sentence-bargain. Those considerations are likely to play a role in the decision-making of police and prosecutors even in the absence of formal policies; one advantage of having a formal program is the possibility of having written criteria and a formalized decision process to prevent rumor, prejudice, or personal animus from having too much weight in determining which offenders are singled out for such treatment.

Direct Communication: The "B Team" Approach

Logically, any jurisdiction with a varsity must also have a junior varsity; an "A Team" implies the existence of a "B Team," whether or not anyone has made a list. That is, there must be a group of offenders who

fall just below whatever cutoff, formal or informal, would qualify them for selective investigation and full-bore prosecution. Especially where the existing "A Team" policies are effective in arresting, convicting, and incarcerating their targets, direct communication with the "B Team" may offer a low-cost, high-value opportunity to increase deterrence (low-cost because the "B Team" is likely to be fairly small, high-value because it consists of people who, if not deterred, will commit serious crimes at a high rate). Simply warning "B Team" members that one more offense will land them on the "A Team," and showing them the statistics about the rate at which "A Team" members wind up behind bars, might cause them to reconsider their current course of action. Minneapolis reportedly reduced its overall rate of homicide by 45% by warning a "B Team" of repeat gang member that one more crime would put them on the target list (Kennedy, 1998).

The principle of having an "A Team" could extend beyond having a single list for an entire jurisdiction. Every police unit focusing on a particular crime type (e.g., domestic violence or auto theft) and every patrol area could make a list of its most troublesome offenders and single them out for special attention; again, that is likely to happen informally if it is not formalized. If the police are doing such targeting properly and accurately, prosecutors ought to be willing to do their part as well, treating a burglary by someone who, while not a sufficiently serious or high-rate offender to make a city-wide "A Team," is nonetheless a well-

known neighborhood menace differently from a burglary by someone else, and therefore not bargaining it out for whatever the usual price of a burglary is in that courthouse. As long as police and prosecution resources are scarce, it makes sense to focus them on the most troublesome offenders in each category.

By the same token, taking the time to assemble the list of the "B Team" for each crime type and each patrol area, and communicating directly with the people on that list, might well provide some bonus deterrence. This is the application of the same principle of concentration as singling out some vehicles for special attention in our hypothetical parking lot: a level of enforcement effort inadequate to deter everyone may well be adequate to deter a subgroup of offenders, especially if that subgroup is explicitly warned that it is being targeted for special attention.

Concentration on Offense Types

The same principle of concentration can be applied to offense types rather than offenders. Perhaps the most famous success story involving that principle was the crackdown on "squeegee men" in New York, which is often thought of as signaling the beginning of the Bratton era even though the operation actually started under his predecessor, Raymond Kelly (Bratton and Knobler, 1998).

The "squeegee men" were, in effect, something between aggressive beggars and low-grade extortionists. They would approach cars stopped at red lights, clean their windshields, and then ask to be paid. Drivers feared, accurately or not, that a refusal would lead to their cars' being damaged, and enough of them paid to make the practice worthwhile for those who engaged in it. Perhaps some drivers even regarded the activity as providing a service, but the majority, or at least the ones who made themselves heard, regarded the experience as unpleasant and even frightening. The squeegee operators concentrated on the entry points to Manhattan, especially the exits from the tunnels, annoying visitors and giving them the impression of a city out of control.

The activity was illegal, but its very triviality made it seemingly impossible to control. The offense was seen as too minor to warrant a full custodial arrest, with the time and expense involved in taking the offender to a lockup, booking him, and holding him for arraignment; instead, police issued Desk Appearance Tickets, which were no more than orders to show up for a hearing later. The squeegee men frequently ignored the Desk Appearance Tickets (known to New York police officers as "disappearance tickets" for the high no-show rate of those ticketed); even a squeegee artist who decided to show up in court was unlikely to face any substantial punishment. In a city averaging between five and six homicides per day, the notion of devoting substantial police resources to curbing what was it worst a nuisance seemed absurd, and Kelly and

Bratton were the targets of mockery when they announced that they were going to put an end to the squeegee problem.

As it happened, the announcement that squeegeeing was no longer to be tolerated, plus a few months of aggressive enforcement (including the use of actual custodial arrest in lieu of DATs), put an end to the problem, seemingly permanently. It turned out that what had been perceived as a huge squeegee problem consisted of the actions of fewer than 100 people, and that their earnings from the activity were not high enough to keep them active once it got to be clear that they would be arrested and charges would be pressed. The activity went away, and so far has not returned, and the New York Police Department promptly turned its attention to other problems.

Concentration plus communication had tipped a high-violation equilibrium to a low-violation equilibrium, which was then self-sustaining, without any need for a substantial residual enforcement effort. Compared to dabbling, concentration actually saved police resources over time; devoting some focused attention to squeegeeing for long enough to accomplish that transition meant that the department could then devote roughly zero attention to it from then on.

In a process with two equilibria and a tipping point between them, an equilibrium that is robust in the face of small interventions may vanish quickly in the presence of a sufficiently large intervention: that is,

enforcement enjoys increasing returns to scale. In a sense, the problem of enforcement management is the inverse of the problem of fisheries or forest management; instead of seeking out the maximum sustainable yield to get as large a flow of fish or timber as can be maintained over time, the enforcement agency looks for a level of arrests that "overfishes" the pool of offenses, leading to a lasting diminution in the stock.

Several characteristics of squeegeeing made that cheap and easy victory possible: it was easy to observe, it was extended in time but was geographically concentrated, it involved a small number of offenders, the rewards of squeegeeing were small, and there was no mechanism by which removing one squeegee man induced someone else to replace him (by contrast with, for example, illicit drug sales).

If NYPD had tried to wipe out horn-blowing instead of squeegeeing, the outcome would have been different, because horn-blowing is diffuse geographically and transient temporally, with a very large number of potential offenders. A long history of crackdowns on drunken driving, using checkpoints at which each vehicle is stopped and the driver scanned for signs of intoxication, has shown that the crackdowns succeed in reducing the prevalence of drunken driving, and the accident rate, while they last, but they are too expensive to sustain and the benefit is about as transient as the effort itself. The problem is that drunken driving is hard to observe, and enforcement against it therefore remains resource-intensive even when it manages to drive the actual

offense rate down. Thus some offenses will, and others will not, be subject to the "tipping" strategy.

"Broken Windows" and "Zero Tolerance"

A mythology has developed around New York's success at reducing both serious crime and disorder. In the myth, the success was achieved using a "broken windows" strategy featuring "zero tolerance." Like any good myth, that account mixes fact and fancy in complicated ways.

Taken literally, "zero tolerance" for violations of the law is an impossibility in an urban setting. Making an arrest takes time, and following up on that arrest takes more time and uses up scarce prosecutor, courtroom, and corrections capacity. Police in any big city already arrest more people than the courts can process without aggressive plea-bargaining, and the courts already generate more convicted offenders than the prisons and jails can hold. (Notoriously, the "market price" in the plea-bargain bazaar for any given offense is lower in high-crime jurisdictions than in low-crime jurisdictions; that's enforcement swamping at work.)

The Bratton strategy for dealing with disorder, as typified by the squeegee crackdown, was *targeted* zero tolerance. Some offenses in some areas were marked for extinction; others got no more than routine attention. The selection was largely left to the precinct captains. There was a city-wide mandate for zero tolerance of open retail drug dealing;

enforcing that mandate was, and remains, expensive, because the economic motivation is stronger, because the number of participants is larger, and because arresting one dealer, as long as there are still buyers, creates a market niche for a new dealer.

In the myth, cracking down on minor disorder offenses contributed to reducing serious offenses by discouraging potential offenders, making the city look like a less hospitable venue for offending. That may be true, but it is not known to be true, nor is it the only good reason to try to control disorder.

The original "broken windows" metaphor (Wilson and Kelling, 1982) refers to a bit of folklore that does not seem ever to have been subjected to a scientific test, but at a sufficiently general level scarcely needs one. The claim is that a vacant house in which a window is broken and not promptly repaired will soon have all of its windows broken by neighborhood kids who read the unfixed broken window as a virtual announcement that the house is un-owned and undefended, thus making it a virtual free-fire zone. Similarly, a street or park that is litter-free will tend to remain somewhat litter-free, while a litter-strewn street or park will quickly be buried in litter, and the behavior of a few rowdy teenagers on a street corner wills stimulate rowdiness in others.

More than one mechanism might explain "broken windows" effects. Simple cue-taking no doubt plays a part. Given a general desire to

conform to the expectations of a situation, it is natural to imitate the behavior one observes; a quiet library does not really need a sign saying "Please be quiet." So may offenders' calculation of how much harm their offenses actually impose on others; the twentieth broken beer bottle in a park does not detract nearly as much from the park experience as the first broken beer bottle.

But insofar as enforcement swamping is a real phenomenon, committing almost any offense where that offense is common will be safer, in terms of the risk of punishment, than committing the same offense where it is rare. Thus a somewhat rational offender who sees some sort of offense being committed and not seeing any police response will make reasonable inference that he is in a relatively safe place to commit that offense.

For all of those reasons, we might expect a generalized "broken windows" effect, such that letting some offense type go unpunished in an area has a strong and rapid effect on its frequency in that area, while a focused crackdown on that behavior will have a strong and rapid effecting suppressing it. That explains cases such as squeegeeing, where it is possible to maintain a very low level of violations more cheaply than it is to tolerate a large number of violations while making more or less token arrest efforts in response to complaints.

There is strong evidence that minor disorder offenses play a large role in shaping residents' and visitors' experience of their neighborhoods, and in influencing their perception of the risk of serious crime. That being so, the police can serve the public by "fixing broken windows," especially when that can be done with only a temporary burst of police attention. That alone is an adequate justification for putting some amount of effort into "broken windows" policing, rather than ignoring minor offenses entirely and concentrating on major ones. Moreover, the gratitude the police earn by protecting the neighborhood from such nuisances will sometimes pay off in the form of cooperation with the police in the solution of major crimes. We might call that the "weak" version of the "broken windows" theory.

It might also be true, or might also be true in some circumstances and not in others, that observing unpunished low-level offending serves as a signal that this is a relatively safe place for serious offending. When that is true, then focused enforcement directed at minor disorder will not only have the benefits of making residents feel safer and perhaps saving on enforcement resources in the long run, it will also reduce serious crime by signaling to offenders that they are not in a place where "anything goes." Call that the "strong" version of "broken windows." It is an entirely plausible story, for which there is to date exactly zero evidence (either way).

Even where the strong "broken windows" theory does not hold, enforcement of the laws against some minor offenses may contribute directly to reducing major crime when the minor-crime offenders are also major-crime offenders. Here the example from New York involved an offense even more trivial on a per-incident basis, though far more widespread, than squeegeeing: turnstile-jumping in the subways. Here again, the offense was so common that enforcing the law seemed impossible, but it turned out that a policy of aggressive arrest and prosecution was able in relatively short order to re-establish the norm of paying to ride the subway, and that maintaining that norm made only modest ongoing demands on police and prosecution resources. That alone was an adequate justification for the crackdown.

But the big payoff was elsewhere, and largely unanticipated. For turnstile-jumping as for squeegeeing, part of the policy change involved making full custodial arrests in lieu of issuing Desk Appearance Tickets. That meant running a warrant check, and it turned out that the turnstile-jumping population had a high density of outstanding warrants, some of them for serious offenses. A full arrest also meant a pat-down for weapons, which is where the program hit the jackpot. Almost one percent of arrested turnstile-jumpers were carrying guns, almost always illegally. Over time, that proportion fell dramatically suggesting that the new policy had succeeded in deterring illegal gun-carrying. (That inference is not water-tight; perhaps instead the policy

merely encouraged fare-paying among illegal gun carriers.) In the meantime, it led to the confiscation of thousands of weapons and to the arrest and incarceration of an unknown but substantial number of truly dangerous offenders. In addition to those direct effects of capturing dangerous offenders or deterring them from carrying guns, the turnstile policy and other similar policies led to an increasing number of direct interactions between police and the people most likely to commit serious crimes. It is possible, though not demonstrated, that such an increase exerts its own deterrent effect, giving those offenders the impression that the police are watching them. (The extent to which police in New York and elsewhere have improperly stopped people who seemed to them likely to be high-rate offenders on no lawful grounds, or on merely pre-textual grounds, remains a troubling one; even operationally successful programs need to be operated within Constitutional norms. The use of traffic stops as an adjunct to general law enforcement creates the same issues.)

Warrant Service

One way to increase the interaction between police and dangerous offenders is to serve bench warrants. A bench warrant issues when an offender fails to appear for a court date, or absconds from probation. (Unfortunately, a bench warrant can also issue for something as trivial as a forgotten jaywalking ticket, so selectivity is essential.)

Empirically, it turns out – unsurprisingly – that people who fail to show up for court or ignore a probation officer's order to come in for a meeting are disproportionately likely to be high-rate offenders. The sort of person who defies not just specific laws but the institutions of the criminal justice system is likely to defy many specific laws, and do so often. So serving bench warrants, in addition to helping maintain the sanctions credibility of the entire system, is also a good way to target those most likely to develop into dangerous offenders.

Serving warrants is also relatively straightforward. While anyone subject to an arrest warrant is technically a "fugitive from justice," few have in fact left their neighborhoods. When the U.S. Marshal Service started to send out its fearsomely-named Fugitive Investigation Strike Team (FIST) units in the early 1980s, most of the apprehensions were made either at the subject's last known residence address or his last known place of employment.

Many police departments, however, treat warrant service is a low-priority task, and arresting a probation violator is not considered the same level of professional accomplishment for a police officer as making a new arrest on a prosecutable felony charge. From the viewpoint of the department's statistical record, it does not result in "clearing" an additional case; whatever crime the wanted person has committed were

already “cleared” when he was arrested, and the fact that the criminal becomes a fugitive does nothing to change that.

But what that suggests is that neither the clearance rate nor the number of prosecutable felony arrests is really a good measure of police performance, either at the individual level, the precinct or patrol-district level, or the department level. When the New York Police Department started to measure itself by its success at crime control rather than its clearance rate, warrant service moved from low priority to high priority.

Multi-Offending Groups

The leading cause of death for African-American and Latino males 15-30 in big cities is homicide. Most of those murders are committed in connection with the activities of persistent multi-offending groups. It is conventional to call them “gangs,” though that locution calls up a Jets-and-Sharks imagery not really consistent with current realities, and also suggests that all persistent multi-offending groups of young men are more or less the same, which turns out to be false. They vary enormously in size, structure, and activity; a gang that wears blue and calls itself "Crips" in Indianapolis may have almost nothing in common with a gang that wears blue and calls itself "Crips" in Los Angeles. Even in LA, "Crips" refers to a set of culturally similar groups, not to a unified organization with a central command structure.

Most of the victims of gang homicide are themselves gang-involved. But the damage spreads much more widely. Gang violence is a neighborhood-killer. The combination of witness/juror intimidation and the "code of the streets" that prevents testimony even against gang rivals means that even first-class homicide investigation does not work as a solution to this particular homicide problem. Like the Mafia families before the advent of the Organized Crime Strike Forces, some of these groups have made themselves substantially enforcement-proof when it comes to homicide, though they remain vulnerable on charges such as drug-selling that do not require civilian witnesses. Successful witness intimidation is self-reinforcing; the more invulnerable the group looks, the more reluctant potential witness are to testify. Some groups have become more ambitious, attempting to intimidate jurors as well. Since they usually face state-level charges, and since the states have little of the machinery of witness protection that the Federal government developed in its successful effort to crack “traditional” organized crime (the “families” of the Mafia, also called La Cosa Nostra). In this context, the recently-announced decision of the FBI to begin to work on witness intimidation cases makes good sense (Butterfield 2005). Putting individual perpetrators in prison, though a worthy achievement, is only half the point: the key is converting the use of witness and juror intimidation from a source of organizational safety to a source of organizational risk.

Not all persistent multi-offending groups of young minority men use deadly violence, and most members of even violent groups aren't themselves killers. The actual number of shooters, even in a big city, is likely to be in scores or hundreds, not the thousands.

The situation, though out of control in many places, is not uncontrollable. Effective law enforcement against gangs rests on two principles: direct communication of credible deterrent threats, and group-level accountability for violence. Those are the principles that David Kennedy and his colleagues employed to such good effect in Boston's Operation Cease-Fire, and subsequently elsewhere (Kennedy, 2003; Kennedy 1998).

The logic of the approach is to enlist the gang's internal structure and multi-offending nature in restraining its most violent members. In a world in which the reputation for skill at deadly violence and for the willingness to employ it are survival advantages both for individuals and for the group, a member who kills a member of a rival group increases the prestige, power, and safety of his own group, though he also incurs a risk of retaliation for himself and his colleagues. He is therefore likely to be encouraged in advance and rewarded afterwards by the group of which he is a part.

The Cease-Fire approach is to gather each group together, confront the members with a solid phalanx of enforcement and criminal justice

agencies, and threaten that any deadly violence by any member of the group will result in concentrated enforcement attention on the entire group, ranging from the enforcement of probation conditions to bringing in federal agents to develop drug-conspiracy or Armed Career Criminal cases leading to long mandatory prison terms. The multiple offenses committed by gang members -- from selling drugs to driving unregistered cars to playing hooky from school -- represent multiple vulnerabilities to enforcement action if a given gang is chosen as a target.

A variant of this approach threatens to put pressure, not on the current members, but still-influential gang alumni (sometimes called O.G.'s, for "original gangstas") with profitable illegal ventures of their own, such as gambling venues, that the police can shut down.

The "Cease-fire" approach amounts, then, to targeted zero tolerance for deadly violence, operating at the group level rather than the individual level. It is not a simple "crack-down" on gang activity in general; a group that refrains from deadly violence will not be specifically targeted, though a "Cease-fire" operation is not a treaty guaranteeing impunity for other crimes, and group members remain as vulnerable as they ever were to arrest and prosecution if apprehended in the ordinary course of enforcement business.

If it seems intolerable that multi-offending groups should be more or less tolerated as long as they don't kill anyone - that simply testifies to

the mismatch between the crime problem and enforcement resources in high-crime jurisdictions. When, for example, officials of the Los Angeles Police Department's gang unit, a small part of a Department whose *total* sworn strength is only 9,000, estimate that there are 50,000-80,000 "gang members" in Los Angeles and say at the same time that the goal of the unit is to shut down gang activity entirely, that goes beyond bravado into the realm of pure fantasy. Keeping the violence down may not seem a very high aspiration, but it is at least a potentially achievable one. Achieving it, though, requires concentration of enforcement attention, not its dispersion over all groups and all offense types. In Los Angeles and other cities where the situation is especially desperate, it might not even be possible, at first, to deliver on a threat to crack down on any group responsible for a killing, if there are too many killings and too few officers to make that threat credible. An alternative would be to restrict the threat at first to cases of violence against witnesses and jurors, or to start in only one part of the jurisdiction rather than trying to impose a cease-fire everywhere.

Police and the rest of the criminal justice system have one great adventure in trying to impose a cease-fire: their actions are very much in the interests of the members of the groups whose violence they are trying to constrain, and in many cases the members know it. They are one another's targets for homicide, and their mortality risks are comparable to those faced by U.S. infantrymen in Vietnam. Any one group that

refrains from violence risks making itself a target for others. In this truly Hobbesian situation, "a common power to keep them all in awe" is not only necessary; it may even, at some level, be welcomed by the young offenders, especially when combined with offers of social-service help to escape from the gang life. It will certainly – by contrast with heavy-handed "sweep" operations -- be welcomed by the neighbors.

References for Chapter 3

Bratton, William and Peter Knobler. *The Turnaround: How America's Top Cop Reversed the Crime Epidemic*. Random House. 1998.

Bureau of Justice Statistics. "Correctional Populations in the United States, 1997" Office of Justice Programs U.S. Department of Justice. November 2000, NCJ 177614

Butterfield, Fox. "Guns and Jeers Used by Gangs to Buy Silence." *The New York Times*. January 16, 2005.

Chaiken, J. M. and Chaiken, M. R. (1982). *Varieties of Criminal Behavior*. Santa Monica, CA: Rand Corp

Greenwood, Peter and Allan Abrahamse. 1982. Selective Incapacitation. Santa Monica: The Rand Corporation

Kennedy, David, "Pulling Levers: Getting Deterrence Right," *National Institute of Justice Journal*, No.236. July 1998,

Kennedy, David M. "Reconsidering Deterrence". A Report to the Office of Justice Programs. U.S. Department of Justice. December 2003

Kleiman, Mark. "The Punishment Deficit and the Prosecutor's Job," Kennedy School Working Paper 87-02-01, 1987

Moore, Mark H. Susan Estrich, Daniel Mcgillis, William Spelman *Dangerous Offenders: The Elusive Target Of Justice*. Harvard University Press. 1985.

Spelman, William. "The Limited Importance of Prison Expansion" in Alfred Blumstein and Joel Wallman, eds. *The Crime Drop in America*, Cambridge University Press, 2002.

Spelman, William.*Criminal Incapacitation*. Plenum Press. New York. 1994.

The Sentencing Project. "New Incarceration Figures: Rising Population Despite Falling Crime Rates". Report. www.sentencingproject.org/pdfs/1044.pdf. 2004.

Wilson, James Q. and George L. Kelling. "Broken Windows". *Atlantic Monthly*. March 1982.

CHAPTER 4

PROBATION

The need for community corrections

Incarceration is cruel, both to those incarcerated and those who care about them. Providing incarceration at even the minimally decent standards demanded by the courts is also expensive, averaging about $25,000 per prisoner per year. Partly because it is expensive, incarceration is also scarce; at any given time, there are more recently convicted offenders than the prisons and jails can hold, though the population behind bars has quadrupled in the past thirty years. Thus many convictions, especially first adult convictions, never lead to incarceration at all.

Community supervision would seem to be a good way to increase the probability that an offense leads to punishment. Freed of the obligation to feed and house their clients, community corrections agencies have the advantage of being able to handle more offenders for less money than the prisons. And indeed they do so; nationally, there are almost three times as many probationers as prison inmates (4.1 million probationers compared to 1.4 million prisoners), yet total probation expenditures are only $4.65 billion (which includes the preparation of pre-sentencing reports as well as community supervision) while prison expenditures come to $31 billion (Glaze and Palla 2004, 2003). Thus holding

someone in a prison costs an average of $22,000 per year, while supervising a probationer costs about $1,000 per year.

With more than two-thirds of those released from prison returning within three years, it is hard to argue that prisons do much in general to rehabilitate inmates, and even the literature on special prison-based programs designed to rehabilitate, ranging from drug treatment to "faith-based" prisons is largely discouraging, at least if read with a critical eye. (Any evaluation that creates a high standard for "completion" of the program under study, and then compares "completers" with the general population, is almost certain to be able to achieve apparent success through selection bias. Such results should not be taken seriously unless they hold up on an "intent-to-treat" analysis comparing the entire program group with a group of randomized, or at least matched, controls (Kleiman, 2003; Manski et al 2001)

The great obstacle to realizing the supposed rehabilitative effects of imprisonment is the enormous gulf between prison life and life in the community. Whatever patterns of behavior offenders learn in prison will translate to the outside with difficulty, if at all (Petersilia, 1997). It is a commonplace of the literature evaluating prison-based rehabilitation programs that "community follow-through" – much easier to arrange for as an experiment than as an ongoing programmatic reality – is all-important, raising the question of whether the follow-through alone, without the prison aspect of the programs, would work nearly as well.

(Inciardi, 1996) (Literacy is the exception that proves the rule. Literacy is a skill rather than a habit. The reading skill acquired in prison will naturally carry over into post-release life. The logic linking literacy programs to reduced recidivism works through changing the qualities of the non-criminal opportunities available to the offender, not on changing attitudes or behavior patterns. That makes it unsurprising that literacy programs are not subject to the general rule that prison-based rehabilitation efforts show at best modest results (Piehl, 1994; Bazos and Hausman, 2003.)

Community-based rehabilitation efforts start, then, with a great advantage over prison-based efforts: they take place in the same setting in which the behavior change being sought is to be manifested. Teaching someone to manage time, control anger, or abstain from intoxicants in a prison may have only a slight impact on the prisoner's behavior in the community where he lives after release.

It is not hard to imagine an ideal community corrections process as a much superior alternative to prison, except for the relatively small number of offenders whose severity-weighted personal crime rates while not incarcerated are high enough to repay the cost of incarceration. (The average prisoner is well worth locking up from the perspective of those who might otherwise become his victims, but in this case the average is a somewhat misleading statistic; the median prisoner committed only a handful of crimes in the year before incarceration, but the average is

pulled up by the extreme criminality of the 10% of the prison population that constitutes the right-hand tail of the distribution. (Chaiken and Chaiken, 1982; DiIulio and Piehl, 1991)

Such an ideal community corrections process would provide a mix of services and rules; each rule would have a direct link to the goal of reducing recidivism and encouraging re-integration (or in some cases integration for the first time) to the "straight" world, including stable housing and employment situations and family relationships; the design of the rules and the capacities and management of the probation department would create a high probability of detection whenever a rule was violated (even if that meant eschewing the creation of otherwise desirable rules for which adequate monitoring is infeasible); and every rule violation would lead to predictable and immediate, though generally mild, unpleasant consequences.

The reality, alas, is quite otherwise.

How probation became "getting a walk"

About 70% of the people under criminal justice supervision are being managed by the community corrections system: they are on probation or parole rather than behind bars (Glaze and Palla 2004). (Probation and parole differ in important ways; what follows will concentrate on the probation system, which is by far the larger of the

two. A third system exerts rather minimal, and nominally non-punitive, supervision over those on bail or other pretrial release.)

And yet, in practice, and especially in high-crime jurisdictions, offenders experience probation as more an annoyance than an actual punishment. An offender convicted and sentenced to a term of probation is said to "get a walk," a result perceived as more a victory than a defeat for the offender and his lawyer. While the goal of probation is said to be maintaining offenders more or less crime-free in the community, and helping them establish, or re-establish, ordered and law-abiding lives, few really expect the system to accomplish those noble goals. Perhaps the strongest, if unintended, testimony to the low expectations of the probation system is that when a probationer commits a new crime, as many do, neither the mass media nor elected officials are likely to hold the probation department accountable.

Unlike incarceration, community supervision relies on the voluntary compliance of the offenders subject to it. Someone put on probation by a judge who never reports for a meeting with a probation officer can hardly be said to be undergoing punishment at all, unless that default causes a police officer to go out and arrest him. Some offenders may find the wide variety of requirements and prohibitions to which probationers are subject, and the obligation they are under to account for their behavior, highly aversive, but the extent to which they obey those requirements and observe those prohibitions is almost entirely up to them; some are

more compliant than others, and some abscond entirely, either never appearing at the probation office at all or ceasing to show up before the term of probation has expired. As much as a third of the nominal caseload of some probation offices consists of absconders.

Thus the first task of a probation office would seem to be inducing compliance by probationers. To accomplish that task, probation officers have, in addition to their personal powers of persuasion, both inducements and threats. As inducements, they have a variety of social services that they can offer directly or help their clients receive from other agencies, both public and private, and to reduce the level of supervision and control they exert over those clients, for example by reducing the number of required visits or the frequency of drug tests. As threats, they have the capacity to increase the supervision level, to impose curfews, community-service obligations, and "day reporting" (a requirement that, for some period of time, the probationer comes to the probation office each morning and remain during the working day) and, if they detect a violation of probation conditions (including any new crime), to refer a client back to court for possible incarceration. (Parole departments, whose clients were sentenced to prison and released early under parole supervision, can revoke probation and return the offender to prison administratively, and do so frequently; that is the major administrative difference between the two systems.)

Thus probation offices face a version of the general problem of designing crime-control regimes; they have a limited capacity to monitor and punish misconduct, and they need to manage that capacity to induce what compliance they can. Therefore we might expect to encounter the same sort of dual-equilibrium situation we found in the hypothetical parking problem. If probationers expect that the rules will be strictly enforced – that violations are likely to be detected, and that detected violations are likely to lead to sufficiently unpleasant consequences (compared to the cost of compliance or the benefit of successful violation) – they are likely to be highly compliant. If most probationers are highly compliant so that the overall violation rate is low, then whatever capacity to punish the office commands will not have to be spread over too many violations, which in turn means that probationers who try to get away with breaking the rules will find that an unwise course of action. If, on the other hand, the rates of detection given a violation of the rules and of punishment given detection are low, then the violation rates will be high, which in turn will keep the sanctions rates low. The inverted U will still be at work: if the violation rate is high, even a relatively low sanctions rate may involve a large volume of sanctions, while if the violation rate is low it may be possible to maintain high sanctions rates per offense without having to inflict much actual punishment.

Probation departments, especially the large ones with many clients in high-crime jurisdictions, are largely trapped in high-violation equilibria. Violations are frequent, and sanctions relatively rare, and the departments lack both the resources and the authority they would need to turn that situation around.

The statement that it costs about $1000 per year to supervise a probationer is only half-true. It might be more accurate to say that *failing* to supervise a probationer costs about $1000 per year. In high-crime jurisdictions, caseloads of 150 offenders per probation officer are typical, giving the average probationer about 15 minutes of a probation officer's attention per week. Probation officers are typically paid substantially less than police officers. Probation offices have been slow to adopt information technology, with much of the paperwork still being done by hand. Drug tests, which cost less than $10 each, are regarded as scarce resources to be rationed out; even probationers known to be drug-involved face testing frequencies of once per month or less. (In Los Angeles County, the overall rate of drug tests averages out to 1.3 tests per probationer per year (Kleiman et al, 2003)). Under those circumstances, it is hardly surprising that many probationers largely disregard the requirements nominally imposed on them by their probation status.

The resource deficit is only one half of the problem, and perhaps the smaller half. Probation also suffers from a deficit of authority.

In most jurisdictions, probation officers typically lack the power of arrest, and must therefore rely on the police to chase down probationers who fail to appear when required to do so, or even "abscond" entirely from probation. A judicial warrant for an arrest ("bench warrant") is not hard to obtain, but in most police departments warrant service is a low-priority task, and arresting a probation violator does not bring the same professional rewards for a police officer as making a new arrest on a prosecutable felony charge. The fact that a bench warrant can also issue for something as trivial as failing to pay a traffic fine devalues the warrant service process in the eyes of the police. In addition, police officers and their managers correctly perceive that a probation absconder is very likely to be arrested for something else eventually, at which time the warrant will show up in the records and the probation violation can be dealt with then, along with whatever led to the new arrest; why, then, make a special effort?

From the viewpoint of the probationer, however, the threat that the bench warrant will be an additional complicating factor at his next arrest is not a very serious or imminent threat. If the new arrest is for a serious crime, the old probation violation, and having absconded, may add little if anything to the period of incarceration for the new offense. In any case, all of that is in the indefinite future; lax handling of probation bench warrants by police eliminates both certainty and celerity from the probation sanctions process, since a probationer facing the threat of

incarceration for a probation violation has – except in the fairly unusual circumstance that he is arrested and held pending the hearing -- the option of simply not showing up.

That reflects the fact that probation departments do not have the authority to impose confinement as a sanction for violation of their rules, including the rules against committing new crimes. Under the 1972 Supreme Court decision in *Morrissey v. Brewer* (408 US 471), a probationer has a "liberty interest" in not being locked up that cannot be taken away without due process, which the Court held meant a full adversarial hearing before a judge. That puts the system most likely to induce high compliance – one in which a detected violation has a high probability of resulting in a swift sanction – largely out of reach.

Judicial hearings take time: an average of about two months in one study of 32 cities (Burke 1997). Therefore, if a probation officer wants to impose a swift sanction it must be a non-incarcerative sanction, unless the department is prepared to have the probationer arrested and held pending a hearing, which if pursued as a regular policy would be highly unpopular with the managers of overcrowded jails.

In addition to making swiftness impossible, having judicial hearings reduces the probability of a significant sanction both directly – the judge may not find a violation, or may find a violation but simply return the

offender to probation, as happens about 40% of the time – and indirectly, by making the sanctions process expensive for the probation department.

Even on a routine matter where the facts are not in serious dispute – most typically, a missed or "dirty" drug test – preparing for and going through such a hearing, which involves drafting the complaint, having it served on the probationer, getting ready to testify, testifying and being cross-examined, represents several hours' work for a harried probation officer and for his or her supervisors, especially in unautomated poorly automated office environments. That creates a significant disincentive for the officer and the department to ask for such a hearing, in addition to the disincentive inherent in the possibility that the court will undercut the probation officer by not ordering a significant sanction, thus weakening the probation officer's credibility in making threats of sanctions against that probationer, and other probationers, in the future.

Judges, who might be expected to regard a probation violation – which is, after all, a violation of a court order – as a professional affront, in fact tend to regard "technical" violations of probation conditions (those that would not constitute crimes for someone not on probation) as minor matters, not worthy of jail or prison time.

For jail or prison time measured in months – the typical length of confinement on a probation violation – the judges may well be right. A much milder sanction – as little as a couple of days' confinement – might

be ample to change behavior, if it were applied quickly and certainly. But judges rarely think in terms of forty-eight-hour sentences, and such a sentence would be a somewhat anticlimactic conclusion to a two-month process. The jails, which house offenders doing short time, are crowded, and face high costs of mustering offenders in and out.

Therefore, if a judge takes any action at all against an offender who has been detected of a violation (by no means assured given the prison crowding problem) it is likely to be severe: a few months behind bars is typical, and offenders have been sent back to finish multiyear sentences for a single positive marijuana test.

As a result, there are strong incentives for probation officers not to take every positive test back to the judge. Probationers may be counseled, warned, or referred to treatment providers several times before being (in the perhaps unintentionally graphic jargon term) "violated." It is hard to fault probation officers for attempting to "jawbone" their charges out of drug use, or into attending their anger-management classes or paying restitution in a timely fashion, rather than proceeding immediately to drastic measures.

Thus the process of adjudicating violations of probation reproduces the process of detecting and punishing crime in general: even once a violation is detected, the sanctions are highly uncertain and significantly delayed. (When Multnomah County (Portland, OR) decided to set up a

"structured sanctions" program to deal with technical violations of probation conditions, the first three "sanctions" were various levels of warnings, giving each probationer a virtual license to break all rules until he had been caught three times (Cavanagh and Harrell, 1995; Harell et al, 1994). Given the large caseloads and the plethora of rules without an associated effective monitoring mechanisms, the probability that a violation will ever be detected is small.

Thus it should be no surprise that probation is regarded by many offenders as not constituting real punishment at all, and that they regard compliance with probation conditions as largely optional. A study of probation supervision in three counties in California (Los Angeles, San Diego, and Santa Cruz) found that in all three, despite large differences in the composition of the probation population and in the practices of the probation offices, a probationer called for a drug test had more than three chances in ten of either not showing up at all or testing "dirty" for one or more drugs on any given occasion. As might be expected, a high violation rate and a low rate of punishment per violation go hand in hand. None of those offices regarded that high violation rate as constituting a problem in need of urgent attention; indeed, none of them had even calculated that statistic, either on an office-wide basis or officer-by-officer. That probationers should disregard even simple, bright-line instruction with a clear link to criminal recidivism did not appear to the managers of those three probation offices – two of them

regarded by their peers as unusually well-run, dedicated, and innovative – as a crisis; it was merely a fact of life (Kleiman et al, 2003).

The current system of unpredictable and uncertain sanctions for probation or parole violations creates at least four unfortunate results. First, handing out merely nominal punishments to those just starting their (adult) offending careers may reduce their fear of arrest and conviction, thus encouraging continued offending. Second, crime victims whose victimization results in community-corrections sentences are frustrated. Third, judges who want to impose real punishment are driven to over-use incarceration. Fourth, legislatures are driven to tie the hands of judges through mandatory sentencing, and to limit or eliminate parole through "truth-in-sentencing" laws, further increasing the number of persons behind bars.

The probation system, designed as a punishment short of prison that would monitor offenders in the community, keep them from committing fresh crimes, and provide both a mandate and assistance to establish, or re-establish, law-abiding lives, seems to be caught in a low-compliance trap that makes it impossible to accomplish any of those tasks.

Escaping the low-compliance trap

Probation could come to represent a serious alternative to incarceration, rather than being a mere placeholder for absent punishments, if community-corrections agencies had the right resources,

the right authority, and the right mind-set. They need the capacity to set and enforce specific rules, including the availability of swift and certain low-intensity punishment for each detected violation of those rules. That in turn requires both technology to observe probationers' behavior and authority and capacity to administer sanctions swiftly and predictably. Existing probation departments lack the staffing levels and the technology to do that job. Changing that would require more money, but an exclusive focus on higher budgets would be a mistake. Money by itself is not a magic bullet. New funds are needed to acquire and employ that technology, authority, and capacity; just adding more money to a broken system will merely make it fail more expensively.

The cost of a community-corrections system employing the full range of supervision technologies would be several times that of existing probation supervision, but it would remain a small fraction – perhaps a quarter – of the costs of imprisonment.

Some of the elements that might make up an effective system are already in place. By law, probationers are far more subject to search than other persons. Probation departments have wide discretion over conditions of probation and have the power to vary those conditions from client to client and from time to time without going back to court. The due-process requirements for incarcerating a probationer for violations of the terms of community supervision, while burdensome, are as nothing

compared to the process requirements surrounding a fresh criminal charge.

The challenge for probation departments is to turn those legal facts into facts on the ground, so that probation becomes a true alternative to incarceration in terms both of punishment and social control, while retaining its advantages in terms of budgetary cost, the suffering inflicted, and its effects on offenders' chances of establishing non-offending patterns of behavior.

Existing probation and parole regimes impose many rules, arguably too many, although they tend not to require, even on paper, that their clients improve their job-related basic cognitive skills in reading and mathematics. Where they fall down is in their capacity to detect violations of those rules and to impose and administer appropriate sanctions. The keys to making community corrections real are better technology to allow community corrections officers to monitor various aspects of their clients' behavior other than by interview and increased authority and capacity to impose immediate and predictable sanctions in response to violations of terms of conditional liberty. The result would be a corrections program that truly "corrected," with benefits for taxpayers, offenders, and potential future victims alike.

Such a program can be thought of as having four elements specific to the powers of community corrections agencies, in addition to whatever

social-service elements might be added by those agencies or others: the establishment of rules, the detection of violations, the imposition of sanctions for violations, and the execution of those sanctions, which includes seeking out those offenders who fail to appear, or fail to comply with the sanctions imposed.

(This puts to one side the role of probation departments in providing or brokering social services for their clients. Providing services is not a task that can be ignored; whatever gives a probationer a better set of non-criminal options, and better capacity and willingness to take advantage of those options, will tend to reduce his criminal activity. But the role of probation departments in providing or brokering social services for their clients seems inconsistent with their role in making and enforcing rules, and there is little evidence that they do an especially good job as service providers. Arguably, that function ought to be voucherized, and the probationers allowed to choose the services, or at least the service providers, they need or want (Dilulio, 1998).

Monitoring

Logically, it might seem that rule-setting ought to precede consideration of how to verify compliance. But if violations of a rule cannot be detected, the rule cannot be enforced, and an un-enforced rule, by weakening sanctions credibility, is likely to decrease compliance with other rules. Face-to-face interviews do not constitute an effective

monitoring technology. So in redesigning probation on paper, perhaps it is best to start by asking what capacity exists, or might exist, to monitor compliance with the rules we might want to impose. Three examples of promising technologies are drug testing, electronic position monitoring, and automated data transfer to community corrections departments from other institutions that serve and supervise the same clients.

Of these, drug testing, while it covers the narrowest slice of behavior, is the best developed and least problematic. Urine tests reflecting the past three day's use of up to five drugs at a time can be administered, in mass production, for a total cost of less than $5 per test, even including confirmation tests when a client challenges the validity of a "dirty" (positive) finding (Kleiman, 1997).

A variety of techniques, from the "electronic handcuff" that monitors an offender's proximity to a fixed base unit, such as in the offender's residence, through currently undeveloped applications of cell-telephone or geo-positioning satellite technology or even the use of mobile video scanners equipped with face-recognition capacity, can be applied to the problem of detecting whether an offender is in a required (or forbidden) location. A probationer might be required to be at home from midnight to 6 a.m., at work during working hours, and at a drug treatment program when scheduled to be there, but never to be within six blocks of the street corner where he used to deal drugs. This would have value both in enforcing detailed time-and-place requirements that might be

imposed and also in accusing (or exonerating) individual probationers of offenses that can easily be pinned down in time and space, such as robberies and assaults.

Less ambitious technically, though more demanding organizationally, would be the development of data links between community corrections offices and other public and private entities - drug, anger management, and other treatment programs; agencies supervising community service; agencies collecting child support or restitution payments or fines; family services agencies; schools; and workplaces – which have in their operating records evidence about whether individual probationers are complying with the terms of their conditional liberty.

In the absence of computerization, position monitoring and data links would simply flood probation officers' in-boxes with more information than they could process. But a well-designed system might allow a probation officer to look at a computer screen each morning to find out which of his or her clients had been out of compliance the previous day. That probation officer could then exercise a degree of control scarcely dreamed of today. Such a system would produce much of the benefit of "day reporting" programs at a small fraction of the cost.

Rule-setting

Since rules must be monitored and enforced, and since monitoring and enforcement are expensive, rules ought to be chosen (from among

those with which compliance can be monitored) for their connection with the process of establishing, or re-establishing, a law-abiding lifestyle. For example, probationers might to be expected to abstain from the use of illicit drugs, obey curfews and requirements to stay away from specific places, such as drug markets and the residences and workplaces of their former victims, and comply with orders to receive treatment, perform community service, meet financial obligations and family responsibilities, and show up for school or work.

While to some extent the rules are set by statute or by specific order of the court, probation departments need to have some amount of constrained administrative discretion in making or changing rules, both for administrative parsimony and for the maintenance of offenders' incentive to comply.

Not all rules will have the same relevance to the problems of all probationers. Requiring someone on probation for hurting someone in barroom brawl to stay out of bars makes sense; imposing the same requirement on someone who stole a car while completely sober makes less sense. Moreover, since being subject to rules – even rules beneficial to oneself – is always somewhat onerous, the number and onerousness of the rules can be varied as a means of rewarding and punishing probationers for compliance or minor noncompliance. Using such small rewards and punishments to induce compliance can help economize on the need for larger sanctions, which are both more painful to the

probationer and more effort for the probation department. But that approach ultimately depends for its effectiveness on the belief of probationers that the rules matter, which in turn depends on the existence of real sanctions if they are disregarded entirely.

To the extent that the clients of community corrections systems are more likely than average to be reckless, then it becomes especially important that the coupling between announced rules and enforced rules be kept very tight, to avoid compromising the credibility of the system by ambiguity. Therefore, the rules, rather than their enforcement, should be relaxed when experience shows that they are not, or no longer, necessary in a particular case.

Sanctions

The capability to detect violations is of little use without the authority and capacity to punish violations. The current system of "seeing the judge" is too slow, too time-consuming for probation officers, too random in its outcomes, and, when it does result in any sanction at all, almost invariably too severe. A properly-functioning corrections system will detect many technical violations, and it needs to be able to handle them on a routine basis, within probation status, rather than as exceptional events calling for revoking that status and moving the offender from the community to the institutional part of the corrections system.

Typical punishments should be hours of unpaid labor, tightening of curfews and other restrictions, and short periods of confinement. Ideally, such sanctions would be assigned administratively, according to a set of published protocols, and with a hearing to determine only the fact of the violation, not the sanction. That would both create the sanction predictability essential to effective deterrence and also help shift the client's attention from the likely behavior of others to the links between his own behavior and its results, creating the “internal locus of control” (Rotter, 1966) known to be a central element in creating lasting behavioral change.

“Community service” hours are already in widespread use both by the courts and by probation offices, but the capacity to use them is generally limited to a level well below their potential by the lack of assignments and the expense of supervision. Ideally, the capacity of that system would be unlimited, because there would be organizations with tasks to be accomplished willing to provide the requisite supervision in return for access to unpaid labor. (In some ways the perfect client organizations would be community-based non-profits with specific tasks that do not compete with the market economy, such as cleaning up vacant lots. Perhaps a small grant program could be established to pay such organizations for providing supervision.) But the reality is not so simple. Naturally, people who work for a living resent working beside those who are working to expiate violations, or competing against them. In

addition, many probationers lack the skills and attitudes that would make them valuable workers. Moreover, imposing community service and collecting it are two different things. A probationer who fails to appear for "community service," or who shirks or even sabotages on the job, faces only the relatively weak threat that the default will be reported, that the probation officer will decide to do something about it, and that the judge will, some time in the future, apply a sanction for what after all is a fairly trivial violation.

In the last analysis, any sanction except confinement is essentially voluntary and therefore depends in the last analysis on the existence of a credible confinement threat in the background. But long confinement is an over-reaction to minor breaches of probation rules. So the question is how to create the right sort of capacity for short-term confinement, and how to make it possible to impose a confinement sanction quickly and predictably rather than slowly and randomly.

Conventional prisons or jails are ill-suited to the task; instead, there is a need to create specially designed and operated community corrections sanctions facilities designed to take advantage of the short-term nature of their task. That would avoid the problems created for jailers by having large numbers of very-short-stay inmates to muster in and out. Such special facilities could be much cheaper than jails, since fewer services would be required; less brutal, since short stays prisoners could be kept apart from one another; and, by the same token, more

aversive due to the imposition of solitude and silence (no radio or television, just books).

That leaves the problem of how to create a confinement option that did not require a two-month wait to see the judge. With sufficient legal ingenuity, it ought to be possible to conform such a program to the due-process requirements of Morrissey v. Brewer (408 US 471).

For example, a probation officer already has the authority to assign a client to "day reporting," in which the offender is in effect confined for several hours per day. But that confinement is legal, rather than physical; the probationer is not behind a locked door and can – at some risk of arrest or subsequent sanction – walk out at any time. By extending "day reporting" around the clock, with the person being sanctioned ordered not to leave a specific (unlocked) room, it might be possible to reproduce the experience of short-term incarceration without depriving the offender of a Constitutional "liberty interest." (Such quasi-confinement would, again, require special facilities rather than using normal jails; an abandoned motel might be ideal.)

Alternatively, it might be possible to induce probation violators to consent to administrative discipline in lieu of judicial processing that imposes tougher sanctions, or to have an "on-call" judge, perhaps physically co-located with the probation department, available to hold swift hearings and hand out formulaic sanctions.

One likely side-effect of tightening supervision in this way would be increasing the rate at which probationers abscond from supervision, thus accentuating the deficiencies of the current system of warrant service. Persuading police departments to take probation bench warrants seriously would be in some ways the most attractive option; one of the initiatives undertaken by William Bratton as Police Commissioner in New York was to require precinct captains to deliver to his office each Monday a list of bench warrants for residents of their precincts that had been outstanding for more than a week. Arresting probation absconders presents no special challenge; most are not fugitives in any genuine sense, and can be found at their previous addresses, workplaces, or street hangouts. Moreover, as a statistical matter absconding is an excellent predictor of re-offending, so the current police practice of waiting for them to be re-arrested is far from optimal as policing strategy, assuming that the goal of the police is to reduce crime. However, that approach requires an adequately-staffed police department; Bratton himself has not instituted in under-policed Los Angeles the policy he deployed so successfully in New York. The deployment of face-recognition technology in police cruisers could greatly facilitate the task; if each cruiser has a database of probation absconders, and cameras to scan the street as it drives along, all that would remain for the officers would be to respond when the camera recorded a possible “hit.”

If the police department will not cooperate, two alternatives suggest themselves. Probation departments could be given independent authority to make arrests, either by making all of their sworn employees "peace officers" or by creating a special absconder-apprehension unit with peace-officer powers. That would allow a useful sort of specialization, at the cost of the specialized neighborhood knowledge that precinct officers often have. Alternatively, and more simply in organizational terms, the probation department could be given a budget with which to hire police officers on paid overtime details.

Implementing a New System

While the full-dress version of this approach would cost a multiple of the current cost of probation, it would still be a just a fraction of the cost of incarceration. Nor would every probationer need all, or even most, of the supervision capacity of the new community corrections system. Even those starting out with the entire menu of restrictions would be expected to earn reduced restriction and scrutiny by periods of compliance.

Implementing the program described above will require that probation officers be more numerous, more broadly trained, and better equipped with information technology and with authority than they now are. Pay, working conditions, and recruiting standards should reflect the increased level of responsibility implied by the idea of community corrections as a front-line crime-control agency. With the new resources

and powers proposed here, probation offices, and even individual officers, could reasonably be held accountable for improving measurable aspects of their clients' behavior and condition, as measured by criminal behavior, drug use, and acceptance of family and workplace responsibilities.

References for Chapter 4

Bazos, Audrey and Jessica Hausman, "Correctional Education as a Crime Control Program". Applied Policy Project, UCLA School of Public Affairs, March 2003.

Burke, Peggy. "Policy-Driven Responses to Probation and Parole Violations", Center for Effective Public Policy. Silver Springs, MD. March 1997.

Cavanagh, D.P. and Harrell, A. "Evaluation of the Multnomah County drug testing and evaluation program." Final report to the National Institute of Justice. Cambridge: BOTEC Analysis Corporation, 1995.

Chaiken, J. M. and Chaiken, M. R. (1982). Varieties of criminal behavior. Santa Monica, CA: Rand Corp

DiIulio, John, "Reinventing Probation and Parole," *The Brookings Review* 1998 (4):40-42.

Dilulio, John Jr. and Ann Morrison Piehl. 1991. "Does Prison Pay? The Stormy National Debate Over the Cost-Effectiveness of Imprisonment." Brookings Review 28-35.

Glaze, Lauren and Seri Palla. "Probation and Parole in the United States, 2003". Bureau of Justice Statitstics Bulletin NCJ 205336. Washington D.C. July 2004.

Harrell, Adele, William Adams, and Caterian Gouvis Roman, "Evaluation of the Impact of Systemwide Drug Testing in Multnomah County, Oregon", Research Report. April 1994.

Inciardi, James A. "Corrections-Based Continuum of Effective Drug Abuse." Research Brief. Washington, DC: U. S. Department of Justice, National Institute of Justice. June 1996.

Kleiman, Mark. "Faith-Based Fudging: How a Bush-promoted Christian prison program fakes success by massaging data." Slate.com. August 2003.

Kleiman, Mark,, Thomas H. Tran, Paul Fishbein, Maria-Teresa Magula, Warren Allen, and Gareth Lacy. "Opportunities and Barriers in Probation Reform: A Case Study in Drug Testing and Sanction." California Policy Research Center, Los Angeles, 2003.

Kleiman, Mark A. R., "Coerced Abstinence: A Neopaternalist Drug Policy Initiative," in Lawrence M. Mead (ed.) *The New Paternalism*, D.C: Brookings Institution, (1997).

Petersilia, Joan, "Probation in the United Sates," *Crime and Justice*, *A Review of Research* , Vol 22, 1997, University of Chicago Press.

Piehl, Anne Morrison, "Learning While Doing Time," Kennedy School working paper #R94-25, Harvard University, April 1994.

Rotter, Julien. **"**Generalized expectancies for internal versus external control of reinforcements" , *Psychological Monographs*, *80*, Whole No. 609 1966.

U.S. Department of Justice. "Truth in Sentencing in State Prisons," U.S. Department of Justice, Office of Justice Programs, January 1999, NCJ 170032.

US Probation and Pretrial Services System. "Year in Review Report, Fiscal Year 2003." August 2004.

CHAPTER 5
DRUGS

The Strange Economics of the Illicit Drug Markets

If one sort of criminal activity above all others might be expected to operate on economic principles, illicit drug dealing is the one. The logic seems simple enough: the price of illicit drugs, like the prices of other goods traded in competitive markets, should move toward the cost of production and distribution. Since the physical drugs are cheap to produce, most of the costs in the illicit drug business stem from enforcement: the risk of arrest, incarceration, and the loss of drugs and other assets, and the cost of taking precautions against those risks (Reuter and Kleiman 1986). (The other big risk faced by drug market participants is the risk of violence from other dealers, and from those who attempt to steal money and drugs from those who cannot complain to the police.)

In theory, the wages of drug dealers' employees, and the dealers' own earnings from the trade, should tend toward being equivalent to their valuation of the risks they face. If prices for drugs are "too high" compared to that standard, new entrants will crowd into the market (or existing dealers will expand their businesses) forcing prices and earnings down. If at the current price the risks outweigh the rewards, new entrants will not be attracted to the business to replace those who quit, die, or go to prison, and the remaining participants, facing less

competition, will be able to demand higher wages and charge higher prices.

Therefore, according to this analysis (an approach economists call "comparative statics,") and applied to the drug markets in a model called "risks and prices," raising enforcement pressure on any given drug should tend to increase the price of that drug, and falling enforcement pressure should tend to allow prices to decline [Reuter and Kleiman 1986). For these purposes, the relevant measure of enforcement pressure is not the total amount of enforcement (measured, for example, by total person- years of incarceration imposed per year) but the ratio of that amount to the size of the market, measured in the physical volume of drugs bought and sold. More years in prison per kilo should translate into higher prices, fewer years in prison per kilo into lower prices.

That analysis has a clear, though surprising, implication: for any given level of enforcement activity, rising volume in the market due to increased demand should cause prices to fall, rather than to rise. (Economists call this phenomenon a "downward-sloping supply curve," and attribute it to "industry-wide economies of scale." In this case, the process works through the enforcement mechanism: the attention of the enforcement system is spread across dealers, and expanding the number of dealers diminishes each dealer's share of that (unwanted) attention. [Kleiman 1992). That makes sense of the otherwise puzzling observation that cocaine prices fell sharply during the late 1970s and early 1980s,

even as the number of active cocaine users grew rapidly: the rising size of the market swamped the enforcement effort, leaving each dealer or each gram of cocaine at smaller enforcement risk than previously.

In trying to move drug prices, enforcement faces an uphill battle. Because drug dealing occurs in a market context, with dealers competing for customers, removing a drug dealer from the trade, either by locking him up or scaring him off, cannot be expected to reduce by one the number of dealers, as locking up or scaring off a burglar reduces the number of active burglars. The natural result of removing one dealer from an active market is his replacement by another dealer, either a new dealer entering the business or another existing dealer increasing his volume of sales.

As long as there are drug buyers looking for sellers, removing one dealer simply creates a market niche for another dealer to fill. Since retail dealing demands no special skill, the supply of potential dealers, especially in poor urban neighborhoods, seems to be effectively unlimited. Thus, through the mechanism of replacement, the economic logic of the drug markets tends to defeat both deterrence and incapacitation.

By contrast, burglars do not compete for houses to burglarize nearly as directly as crack dealers do for users to sell to, because houses to break into are not scarce compared to the number of potential burglars.

Thus incarcerating or deterring one burglar does not create new opportunities that will lure other potential burglars into action.

Thus the expected effect of drug enforcement levels on the quantity of drugs sold is indirect, through the price mechanism. Higher enforcement risks to dealers would be expected to cause higher prices, and higher prices to reduce the quantity purchased, to an extent depending on the price-elasticity of demand for the drug.

But, if that simple economic theory were an accurate representation of the actual situation, the enormous expansion in drug enforcement activity since 1985 should have led to a substantial increase in price, even in the face of the replacement problem. The number of cocaine dealers in prison, for example, is about 3 times as high today as it was in 1995, while the quantity of pure cocaine available is about three times as high, implying an expected price increase on the order of fivefold. Yet in fact prices have moved in the opposite direction; adjusted for inflation, cocaine costs about a fifth as much today as it did in 1982. (Heroin has behaved similarly.) [Walsh 2004, Sourcebook of Criminal Justice Statistics 2004, United States Sentencing Commission 2002.]

What happened? And what should we do now?

The "what happened" part is complicated. If the "risks and prices" model is wrong, as the evidence suggests it is, there is no alternative model of the determination of prices and volumes in illicit markets, and

how enforcement does or does not influence prices and volumes, to replace it, though some suggestions have been made. (Kleiman 1989, Boyum 1992, Caulkins 1990).

One explanation that does not require a complete abandonment of the comparative-statics framework is that the market for illegal labor functions rather oddly. Despite the great increase in the risk of going to prison, retail crack dealers work today at much lower wages than they did fifteen years ago. (Reuter 1990) That might in part reflect a perceptual lag, as young people enter the trade with unrealistic beliefs about its risks and rewards based on earlier conditions. If so, that suggests that there should be more effort than there is to warn young people in drug-impacted neighborhoods about the risks of selling drugs, to parallel the national effort to make young people aware of the risks of using drugs [Kleiman 1997a]. An alternative, or perhaps complementary, explanation would focus, not on the information available to retail dealers, but on their alternatives to dealing. One consequence of the expansion in imprisonment rates during the 1980s and 1990s is that the flow of ex-prisoners back into the community has expanded, now running at a rate of about 2,000 persons per day. Ex-prisoners notoriously face very poor prospects in the legitimate labor markets; those with experience of cocaine dealing may be “employable” as crack hustlers when no other employer will have them, and thus be willing to work (as retail crack dealers reportedly now do) for less than

the minimum wage. Moreover, developing a personal crack habit is an occupational hazard of crack dealing; someone with such a habit may be willing to work for very low wages in a job that also secures him access to wholesale-priced cocaine.

Whatever the explanation, the fact remains: despite an unprecedented level of law enforcement directed at the cocaine trade, prices have not only not risen but actually fallen. With approximately a third of a million people now behind bars for cocaine dealing, and police departments under budget pressure and forced to divert resources to the homeland-security effort, it is hard to imagine much in the way of further expansion of the cocaine enforcement effort.

If the phenomenon of falling prices in the face of rising enforcement were restricted to cocaine, it would be tempting to look for some explanation in the idiosyncrasies of that market; but the pattern has been very similar in the heroin market. Perhaps it is time to confess that, under current U.S. conditions, drug law enforcement has a very limited capacity to raise the prices and reduce the availability of mass-market drugs, and thereby to reduce the extent of drug abuse.

Demand-side options

And yet drug abuse, the crime committed by drug addicts, and the violence and disorder surrounding the illicit markets, are all big problems. So what is to be done?

In principle, these illicit market side effects could all be eliminated by repealing the drug laws, but even if that were politically feasible – as it certainly is not – no one has proposed a set of regulations short of prohibition that would prevent a major expansion in the prevalence of substance abuse disorders without simply recreating the enforcement problems created by the current prohibitions: regulations tight enough to bind behavior will always provide profitable opportunities for evasion (Kleiman 1992).

Another approach is to work from demand rather than supply, either preventing drug abuse through information and persuasion or treating those who become the victims of addictive disorders. Clearly, reducing drug abuse other than by enforcement is pure gain from a crime-control perspective; it will reduce both crime by addicts and the violence and disorder incident to the illicit markets, and any volume shrinkage due to demand reduction will make the enforcement task against the smaller remaining market that much easier.

But there is little evidence that even the best prevention programs have an impact on hard drug use (as opposed to their demonstrated ability to retard initiation to nicotine, alcohol, and cannabis). And most actual prevention programs are very far from the best (Caulkins et. al 1999).

Some drug treatment programs are demonstrably effective, and cost-effective, as crime-control measures, even putting aside their benefits to the people treated and to the intimates of those people.

In particular, the opiate maintenance, or substitution therapies, (both the old stand-by methadone and the newer LAAM and buprenorphine) are measurably successful at reducing heroin use and crime among heroin users, relatively inexpensive, and always in demand. Increasing the public budgets that support methadone clinics, loosening the unnecessarily burdensome regulations that limit their operations, and continuing the process of making buprenorphine available outside the clinic system all have real promise as crime control measures, though they are usually thought of as social services or health care. Of the estimated 1 million Americans with heroin problems, only about 100,000 are now in substitution treatment; that number could easily be expanded greatly, and the benefits would be substantial. (Experiments in Switzerland and The Netherlands have shown that some very refractory and socially expensive heroin addicts can be managed successfully by allowing them to have as much heroin as they like in clinic settings, on the condition that all the heroin be consumed on the spot (Van den Brink et al. 1999, Rehm, et al.2001, Farrel and Hall 1998, Ali et al. 1999, Perneger et al. 1998). That is a costly approach – probably too costly ever to be a significant part of the heroin treatment system – but probably worth it for the most problematic heroin users.

Political and administrative feasibility is a different, and in this country perhaps insurmountable, issue.)

However, it seems no more likely that we can treat our way out of our current drug problems than that we can arrest and imprison our way out of them. Problems with stimulants, including cocaine and methamphetamine, do not respond to maintenance therapies, and the efficacy of non-maintenance drug therapies is, in general, lamentably small. Almost any treatment works well enough while the sufferer is attending treatment to justify its cost (Rydell and Everingham 1994), but few have the capacity to keep their clients engaged for the long haul, and the evidence of reduced drug use after treatment ceases is generally not very impressive (Informing America's Policy on Illegal Drugs 2001). Increasing treatment supply is feasible, but the paucity of treatment demand is another, and in many cases the more important, limiting factor.

But the very overlap between the population in need of substance abuse treatment and the population of offenders that gives the illicit drug problem some of its urgency also provides a potential solution to the problem of inadequate treatment demand and treatment persistence. About three-quarters of the frequent, high-dose cocaine-using population gets arrested in the course of any given year. [Kleiman 1997b]. Among a population so criminally active, a large proportion is likely to be on probation if not currently incarcerated, and attending drug treatment

and complying with treatment instructions can be, and often is, made a condition of probation, either legislatively, judicially, or administratively.

Alternatively, judges sometimes make acceptance of drug treatment a condition of imposing a probationary sentence in lieu of a sentence of confinement. Such arrangements have been formalized as "drug diversion" programs, of which California's Proposition 36 is an example, and as "drug courts." The institutional arrangements vary, but the basic structures are the same: an offender is required to attend treatment in order to stay out of prison or jail.

But the coercion in such "coerced treatment" is no stronger than the capacity of the supervising authority to enforce it, and that capacity has proven to be limited, especially when the supervising authority is a probation agency. Both in the classical diversion programs organized under the TASC rubric and in Proposition 36, a third or more of those subject to "coerced treatment" never show up even for a first treatment visit, and most of the rest drop out of treatment before completing its prescribed term. In most cases, such defaults either never come to the attention of probation authorities at all – because the treatment providers, which are after all not law enforcement agencies, do not usually regard reporting their clients to the authorities as part of their professional role, and the probation offices lack direct access to the data – or are never acted on by them. (One advantage claimed for the "drug

court" approach is that judicial supervision will increase compliance; the truth of that claim remains unresolved.)

In addition, diversion programs, including Proposition 36, and drug courts provide coercion only for those drug-involved offenders who volunteer for it. Those who choose to take their chances with ordinary criminal-justice processing, and those ineligible for treatment-in-lieu-of-incarceration programs, face only routine probation-department scrutiny, which leaves them pretty much at liberty to continue their drug use.

With the arrestee population accounting for approximately 40 percent of the total consumption of cocaine (Kleiman 1997b), and probably a comparable percentage of the consumption of heroin and methamphetamine, leaving most of that population untreated for its substance abuse disorders looks like an expensive mistake to make.

One alternative would be to tighten up on the existing forms of coerced treatment, by expanding the size of the population mandated to treatment, or by increasing the pressure to attend, comply with, and persist in treatment, or both. But monitoring treatment attendance and compliance, across the gulf between the criminal justice system and the world of the treatment provider, is always going to be difficult. Moreover, at some point the (induced) demand for treatment will begin to outstrip the supply; there are substantially more heavily drug-involved offenders

under criminal justice supervision than there are total publicly-paid treatment slots, let alone vacant publicly-paid treatment slots, and adding treatment capacity of decent quality levels costs real money, on the order of one to several thousands of dollars per client per year. And though treatment is virtually always useful in reducing drug consumption by those attending it, treatment attendance is not the real endpoint to be aimed at: treatment is a means, and not the only means, to the end of reduced drug use and criminal behavior. Most people with substance abuse disorders recover, and most of those who recover never receive formal treatment. And some people who are arrested for drug-related crimes – in particular, for simple possession – do not in fact suffer from any substance abuse disorder at all, according to the diagnostic standards in that field.

That suggests an alternative to coerced treatment: coerced abstinence from the use of expensive illicit drugs for offenders known to be drug-involved, through the use of drug testing and sanctions for probationers. Such a system might work in the following manner:

Probationers and parolees would be screened for cocaine, heroin, or methamphetamine use, using a combination of records review and chemical tests. Those identified as users would be subject to twice-weekly drug tests, leaving effectively no "safe window" for undetected use. Every positive test would result in a brief period of incarceration, perhaps forty-eight hours: a sanction to be applied immediately and

automatically, with no discretion except perhaps for a delay to the following weekend to avoid interference with work. The offender would be entitled to a hearing only on the question of whether the test result was accurate; the penalty itself would be fixed. Missed tests would count as "dirty." The underlying theory that swiftness and certainty can substitute for severity, increasing compliance rates while decreasing the need for actual sanctoins.

After a substantial period – measured in months – of no missed or positive tests, or alternatively achievement of some score on a point system, offenders would be eligible for less frequent testing. Continued good conduct would lead to removal to inactive status, with only random testing.

Operating such a system would require the capacity to do tests at locations reasonably accessible to those being tested (since they have to appear twice a week); on-the-spot test results, both to shrink the time gap between misconduct and sanctions and to reduce the administrative burden of notifying violators and bringing them back for hearings and punishment; the capacity for quick-turnaround (within hours) verification tests on demand; authority to apply sanctions after an administrative hearing or the availability of an on-call judge who can hear a case immediately; confinement spaces for short-term detainees available on demand; and the capacity to quickly apprehend those who fail to show up for testing. Ideally, such a system would embody rewards

as well as sanctions; that rewards shape behavior more powerfully than punishments is a well-established result. Of course, the political problems of rewarding law-breakers for obeying the law are substantial ones, and the best feasible approach may be to use praise and reduced supervision as the primary forms of reward. But collecting an up-front "participation fee" or "fine" that is then returned in small increments for each "clean" test might greatly reduce the failure rate.

Some participants would prove unable or unwilling to reform under punitive pressure alone. For that group, treatment would be essential, if only to reduce the burden they put on sanctions capacity, and the cost of such treatment counts as part of the cost of the system. But by the same token, it is quite possible that many of those now sent to coerced treatment could succeed without it if subjected to testing and sanctions, and the treatment services they did not consume would represent an offsetting savings. From the perspective of treatment providers, having clients motivated to succeed (because failure will lead to sanctions from the criminal justice system) rather than clients who appear only grudgingly, should be substantial.

A rough calculation suggests that such a system could be mounted for about $3600 per participant per year, including a budget for treatment and a budget for sanctions. To subject all of the nation's known hard-drug-using probationers to such a system would cost about $7 billion a year, which is a large number compared to the probation

budget but a modest one compared to the $50 billion cost of current drug enforcement, incarceration, and publicly-paid treatment operations.

If successful nationwide implementation of such a program reduced drug use among its participants by two-thirds – a reasonable number, based on the efficacy of existing treatment programs and the fact that testing-and-sanctions outperformed treatment in their one experimental head-to-head match— and if in fact sixty percent of the hard drugs now go to the criminally active population, then the result would be a forty percent reduction in the volume in the illicit markets. Insofar as the continued criminality of that group is directly linked to its hard-drug use, the result would be a significant drop in property crime. Moreover, shrinking the illicit drug markets would be expected to shrink the associated violent crime and disorder, with great benefit to currently drug-market-impacted neighborhoods.

In addition, there would be benefits to the user-offenders and to their families from the reductions in drug use and associated problems. Even a modest reduction in crime, and consequently in arrest and incarceration, in that population would more than make up for the cost of the occasional two-day confinement terms served as a result of detected drug use. (In the absence of experimental studies, it is hard to guess whether being subject to testing and sanctions for a period of months would lead to a reduction in drug use after the testing period ended.)

By making probation a more effective alternative to incarceration, a testing-and-sanctions system might also be expected to increase the willingness of judges to use it, and of legislatures to allow its use, but the existence and extent of such a benefit remains a matter of speculation. Less speculative is the prospect that shrinking the drug markets would shrink the number of drug dealers sent to prison, thus allowing either a reduced prison population or more punishment for other crimes.

On top of all that, there would be a round of secondary benefits: as the enforcement system confronted smaller amounts of drug dealing and property crime, enforcement pressure on the remaining dealers and property criminals would increase.

Given the limitations of the current probation system, the plan sketched above seems more like a dream than does like a practicable proposal. The addiction management problems are all reasonably tractable, but the public management problems may be virtually insoluble in the absence of a great policy entrepreneur.

Getting drug enforcement right

None of the available demand-side options, then, seems likely to come riding to the rescue of the drug law enforcement effort. But whether they do or do not, we are left with the question of how best to deploy enforcement resources against the illicit drug markets.

If, as argued above, drug law enforcement as applied to drugs with established mass markets has only limited capacity at the margin to

reduce drug abuse by making those drugs more expensive and harder to find, what other good purposes could it serve? ("At the margin" is a significant restriction on this proposition; if drug enforcement were eliminated or cut back to a fraction of its current levels, the result might well be a significant increase in drug availability and a significant increase in drug prices, leading to an upsurge in drug abuse and its consequences. The spread of opiate addiction into areas where it had previously been largely unknown as a result of easier, though still illegal, access to prescription opioids such as Oxycontin demonstrates that availability matters. But it still might be true that increasing or decreasing cocaine law enforcement by 20% would have no noticeable impact on cocaine consumption.)

Targeting new drugs and minding the boundaries

One answer is to concentrate on drugs that do not yet have established mass markets, or on places yet unreached by drugs with established markets elsewhere. The arrest of one crack dealer out of thousands in a big city is largely futile, but the heavily publicized arrest of the first crack dealer to appear in a rural town, without an existing base of heavy users, might be extremely useful. Similarly, a drug whose marketing machinery and user base are still in the process of being established may be much more vulnerable to enforcement pressure than a drug whose market is mature. That the current enforcement effort against cocaine does not seem to be gaining on the problem does not

imply that an effort of that magnitude mounted in 1979 or even 1982 might not have greatly reduced the ultimate size of the cocaine problem, both because its impact would have been greater on a smaller market and because the epidemic pattern of the spread of drug abuse can be quite sensitive to forces that slow it down during its exponential-growth phase (Naik et. al 1996). Of course, drug epidemics are easier to spot in retrospect than in prospect, which makes the approach Churchill called "strangling the baby in its cradle" harder to apply in practice than it is to discuss in theory.

Targeting market side-effects

But just as drug abuse is not the only bad result of the illicit drug markets, reducing the extent of drug abuse should not be the only goal of drug law enforcement. Not every drug transaction, every dealer, every organization, every transaction process, or every market location makes the same contribution to violence and disorder. Enforcement has the capacity to reduce the side effects of drug market activities by singling out the most noxious individuals, organizations, and activities for special enforcement attention, thus exerting both Darwinian and economic pressure to push drug-market activities in less harmful directions.

In choosing targets for enforcement action, and in setting the sentences for convicted drug offenders, the current system focuses on the kind of drug involved and the quantity of the transaction (or the

volume handled by the target organization). Sentences can indeed be enhanced for using violence, or employing or selling to juveniles, but the basis is always drug and quantity, and enforcement agents and agencies frequently use sentence length as a measure of the importance of the case and the quality of their handling of it. Under those circumstances, acquiring a reputation for violence, especially against informants, may actually reduce the vulnerability of a dealer or dealing organization to enforcement action, which is about as perverse an incentive effect as could be imagined.

There is a strong case to be made for turning the system around and focusing instead on violence (especially the intimidation of witnesses) and on the use of juveniles, rather than on drug volumes. That would have the dual effect of getting the most dangerous dealers off the street and encouraging dealers considering alternative styles of dealing to choose the less violent styles. Since much drug-market violence is against other market participants, identifying the most violent dealers in a given city should require no more than interviews with informants and with currently imprisoned dealers, asking the simple question, "Who in this town scares you the most?"

Another targeting rule should be to focus on the most flagrant dealing processes. Drug transactions vary on the dimension of flagrancy vs. (relative) discretion. Some proclaim themselves, as when a dealer does business openly on a busy streetcorner or in a public park, with

runners to approach the drivers of passing vehicles to ask what they might be looking for, or when a house or storefront is converted into a dedicate drug-selling location. Others hide themselves, as when a customer pages a dealer and the dealer arranges for a delivery to the customer's home.

Flagrant dealing is a disaster for the neighborhood in which it occurs. An open street market, or a proliferation of drug houses, is about the strongest evidence possible that the forces of order are not in control, something sure to frighten the law-abiding and likely to embolden the criminally-minded with the thought that where drug dealing goes largely unpunished it might be possible to get away with other crimes as well. Equally worrisome is the effect of visible open markets on public perceptions of the police: residents who see open drug transactions and cannot understand why the police do not stop such flagrant lawbreaking may perceive the police as incompetent, indifferent, or even corrupt.

But the public order threat from open drug markets only starts there. Property crime and prostitution are two major sources of money for hard-drug purchases, so drug-market neighborhoods are likely to face more than their share of robbery, burglary, and streetwalking.

Worse yet, drug buyers and sellers carrying cash and valuable drugs, and reluctant to call the police to complain if they are victimized, provide highly tempting robbery targets. Consequently, the dealers in particular have strong reasons to go armed. In addition to violence directly related

to drug-market activities – disputes over territory or debt, retribution against dishonest employees, and intimidation of potential witnesses – the ubiquity of firearms will tend to convert some ordinary interpersonal disputes into incidents of deadly violence. A quarrel initiated by a slight, an insult, or courtship competition, of the sort that in a different neighborhood might lead to a fistfight, may lead to gunplay instead.

This gives the police both operational and community-relations reasons to "do something" about street drug markets. As a result, police departments across the country continue to make large numbers of low-level drug-dealing arrests, in most cases with little hope that anything substantial will change as a result.

Low-arrest crackdowns on focal-point markets

Drug dealers and drug buyers cluster, for two reasons: to find transactions partners, and to avoid police. Buyers will be able to "score" more quickly and reliably where there are many sellers, and sellers will be able to dispose of their inventory more quickly where there are many buyers, who in turn are likely to be attracted by the concentration of sellers. (Thus drug dealers who appear to be rivals may actually be beneficial to one another; only when a single drug dealing organization is big enough to maintain enough of its own sellers in a given location to make it attractive to buyers will it attempt to keep competitors out.)

In addition, dealing drugs where drug-dealing is prevalent takes advantage of enforcement swamping: the presence of many people engaged in illicit dealings reduces the risk that any given one of them will be arrested, by providing competing targets for police attention. A single buyer on a streetcorner is conspicuous; one of twenty may easily be lost in the crowd. As long as dealing activity is more concentrated than the police resources to deal with it, the risk of arrest will, other things equal, be lower where the density of dealers is higher.

In the absence of advertising, buyers and sellers have no easy way to agree to gather in one location rather than another, or to change a location that has become established. Thus it may take considerable time for a street market to develop; once developed, it will tend to stay put unless something forces it to move. It acts as what Schelling calls a "focal point."

In principle, the solution to enforcement swamping is the concentration of enforcement resources in focused and sustained market crackdowns such as New York's Operation Pressure Point. (Kleiman 1988) Clustering tends to provide safety in numbers for lawbreakers only to the extent that illicit activity is more concentrated than is the enforcement response. If, instead of spreading effort across many drug markets in a city, the police focused on one or a few markets, they could make the risk of arrest in one of the target areas higher than the risk a dealer or user would face elsewhere. If that is true, and perceived to be

true by market participants, one of the two reasons that led them to cluster there will no longer be operative, giving them an incentive to look for less hazardous locations. Once enough of them do so, the second reason for clustering – the easy availability of transaction partners – will also diminish or even vanish as applied to the target area.

The same positive feedback mechanism that led to the growth of a concentrated market in the first place will start to work in reverse: as enforcement drives away some participants, their departure will increase the enforcement pressure on the remainder, and the diminished numbers of potential transaction partners will make buying and selling there harder, making the target area less attractive for new buyers and sellers alike. Eventually, the target area will reach the tipping point at which the density of illicit activity is no longer large enough, relative to the level of enforcement pressure, to be an attractive venue for buyers and sellers. If that situation persists for long enough to be noticed by market participants, the "focal point" convention will have come undone; again, in the absence of advertising, it may take time for a new consensus location to emerge.

Still, crackdowns tend to be highly expensive. Once established, a concentrated focal-point market is likely to be quite robust to enforcement action, due to the replacement effect; arresting one dealer creates a market niche for another.

Arrest is an ever-present risk of buying and selling in open drug markets, and how a given individual, gets arrested is largely a random process, punctuated by occasional organized but short-lived "sweeps." As a result, dealers may be slow to perceive that the risk of arrest in a given dealing area has undergone a persistent increase, and therefore slow to change their place of business. Since "sweeps" are common, and police decisions to make the long-term commitment required to shut down an open market are rare, market participants have good reason to expect that the changes in enforcement pressure they observe are transient rather than permanent. That perception, and difficulty of creating a new "focal point," will tend to make street markets slow to respond to focused enforcement.

So the police would actually benefit if drug buyers and sellers in the target area had a more accurate picture of what the police are doing. This is counterintuitive, but it follows from the idea of enforcement swamping plus the simple observation that arrests are costly. If the cases are prosecuted (without which the effort will lose most of its impact) putting pressure on the dealers in a big open market will also put great pressure on prosecutors and the courts. Operation Pressure Point, the crackdown that eventually broke the famous "Alphabet City" drug market on New York's Lower East Side, also paralyzed Manhattan's criminal courts for a year (Press 1987).

That suggests a way to reduce the cost of breaking up a drug market: the police can "telegraph their punch" by announcing the focused enforcement effort, ideally doing so in advance. Every dealer and every buyer who heeds the warning will make continuing to deal at the target location more dangerous, and less rewarding, for those who remain. The more obviously credible the initial threat is, the fewer actual arrests will be required to "tip" the target market area out of existence.

That approach has now had its first practical test, with apparently excellent results. (Kennedy 2003) In High Point, North Carolina, the unwilling home to a major drug market that was, as usual, a source of violence and disorder, and which had been resistant to sustained routine enforcement efforts. The market, though substantial compared to the size of the city, turned out to involve only about two dozen dealers. Instead of arresting them one by one, giving replacement a chance to work, the police in High Point patiently identified all of the dealers and made the "buys" required to prepare airtight cases against them. The dealers (and, in a brilliantly seriocomic touch, their mothers) were then invited to a meeting, at which they found a solid phalanx of enforcement and prosecution officials and a group of social-service providers. They also found a set of chairs with their names on them (plus three empty chairs), and a set of loose-leaf notebooks, also labeled with their names.

The head of the High Point enforcement effort then explained that the three empty chairs were for the three most violent dealers in town, all of

whom had been arrested that day, and that the notebooks contained the evidence on which any of the twenty-five dealers in the room could be arrested and convicted with no further investigation necessary. The shocked dealers were also told that, as of that day, the open drug market was closed, and that any one of them so much as suspected of dealing from then on would be prosecuted on the evidence already gathered and in the notebook. The social-service providers were available for those who needed various kinds of help (literacy, drug treatment, job training, housing, tattoo removal) in turning their lives around.

The result, as reported by one of the designers of the initiative, was virtually magical: the drug market dried up overnight. Two new dealers who thought they could take advantage of the sudden supply shortage in the open market were promptly arrested. The effect on the volume of drug transactions and the extent of drug abuse in the High Point area is unknown, and may not have been substantial or lasting. But the contribution to crime reduction was dramatic.

Even where the number of dealers makes the approach of preparing cases against all of them infeasible, as will usually be the case in big cities, the concept of a pre-announced crackdown still has potential. Pick a particularly bad drug market, pick a shut-down date, and then start a month in advance to spread the word that the market will close down as of that date. The means of communication are limited only by your ingenuity: posters, postcards to the drivers of vehicles that seem to

be cruising for drugs, face-to-face warnings to known dealers (and perhaps to their mothers as well). Have enough enforcement muscle available on the announced day to be able to carry through on the threat to arrest anyone who tries to deal in the teeth of the crackdown, and budget enough officer time for the subsequent few weeks to keep ahead of the curve of any returning activity. If it works, the effect on the neighborhood will be virtually the same as the result of an old-fashioned crackdown, at a fraction of the cost in arrests and prosecutions. By keeping your powder dry in that way, you will be able to move on fairly quickly to the next-most-threatening market, or to any nascent market that springs up to serve the displaced buyers and sellers from the site of the initial crackdown.

That might not reduce drug abuse very much; markets can adapt in the direction of discretion without greatly inconveniencing customers, and sufficiently determined customers – likely to be present in any location that once housed a major open market – will almost always find a source of supply somehow. The point of the exercise, however, was not to reduce drug abuse, but to reduce the damage done to the neighborhood by the side-effects of drug dealing, and to do so at a low enough cost to make the effort feasible and sustainable.

References for Chapter 5

Ali, R., M. Auriacombe, M. Cases, L. Cottler, M. Farrel, D. Kleiber, et al. 1999. "Report of the External Panel on the Evaluation of the Swiss Scientific Studies of Medically Prescribed Narcotics to Drug Addicts." *Sucht* 45: 160-170.

Boyum, D.A. 1992. "Reflections on Economic Theory and Drug Enforcement." Ph.D. Thesis, Harvard University.

Caulkins, Jonathan P., Peter C. Rydell, Susan S. Everingham, James R. Chiesa and Shawn Bushway. 1999. *An Ounce of Prevention, A Pound of Uncertainty: the Cost Effectiveness of School Based Drug Prevention Programs.* Santa Monica, CA: RAND.

Caulkins, Jonathan P. 1990. "The Distribution and Consumption of Illicit Drugs : Some Mathematical Models and Their Policy Implications." Ph.D. Thesis, MIT.

Farrel, M. and W. Hall. 1998. "The Swiss Heroin Trials: Testing Alternative Approaches." *BMJ* 316:639.

Kennedy, David M. 2003. "Reconsidering Deterrence." Report to the Office of Justice Programs, US Department of Justice.

Kleiman, Mark. 1992. *Against Excess: Drug Policy for Results*; New York: Basic Books.

Kleiman, Mark. 1989. *Marijuana: Costs of Abuse, Costs of Control*; Greenwich, Conn.: Greenwood Press.

Kleiman, Mark. 1997a. "Reducing the Prevalence of Cocaine and Heroin Dealing Among Adolescents." *Valparaiso University Law Review* 31(2): 551-564.

Kleiman, Mark. 1992. "Neither Prohibition Nor Legalization: Grudging Toleration in Drug Control Policy." *Daedalus* (Summer).

Kleiman, Mark. 1997b. "Coerced Abstinence: A Neo-Paternalistic Drug Policy Initiative," in Lawrence A. Mead, ed., *The New Paternalism*, Brookings Institution Press.

Kleiman, Mark. 1988. "Crackdowns: The Effects of Intensive Enforcement on Retail Heroin Dealing." In *Street Level Drug Enforcement: Examining the Issues,* ed. Marcia R. Chaiken. Washington DC: National Institute of Justice.

Manski, Charles R., John V. Pepper, and Carol V. Petrie, eds. 2001. *Informing America's Policy on Illegal Drugs: What We Don't Know Keeps Hurting Us*. Washington DC: National Academy Press.

Naik, Ashish V., Alok Baveja, Rajan Batta, and Jonathan P. Caulkins. 1996. "Scheduling Crackdowns on Illicit Drug Markets." European Journal of Operational Research, Vol. 88, pp.231-250.

Perneger, T.V., F. Giner, M. Del Rio, and A. Mino. 1998. "Randomized Trial of Heroin Maintenance Program for Addicts who Fail in Conventional Drug Treatments." *BMJ* 317: 13-18

Press, Aric. 1987. *Piecing Together the System: The Response to Crack*. New York: Bar Association.

Rehm, J., P Gschwend, T Steffen, F Gutzwiller, A. Dobler-Mikola, A. Uchtenhagen. 2001. "Feasibility, Safety, and Efficacy of Injectable Heroin Prescription for Refractory Opioid Addicts: A Follow-Up Study." *Lancet* 358: 1417-1420.

Reuter, Peter and Mark Kleiman. 1986. "Risks and Prices: An Economic Analysis of Drug Enforcement." In *Crime and Justice: An Annual Review of Research* Michael Tonry and Norval Morris (eds.), Vol. 7.

Reuter, Peter, Robert J. MacCoun, P.J. Murphy, Allan F. Abrahamse and Barbara Simon. 1990. *Money from Crime: A Study of the Economics of Drug Dealing in Washington D.C.* Washington DC, RAND publications.

Rydell, Peter C. and Susan S. Everingham. 1994. *Controlling Cocaine: Supply Versus Demand*. Santa Monica, CA: RAND.

Sourcebook of Criminal Justice Statistics 2004. "Table 6.54: Federal Prison Population, and Number and Precent Sentenced for Drug Offenses."

US Sentencing Commission. 2002. "Report to the Congress: Cocaine and Federal Sentencing Policy May 2002." http://www.ussc.gov/r_congress/02crack/ch6.pdf January 28, 2005.

Van den Brink, W., VM Hendriks, and J Van Ree. 1999. "Medical Co-prescription of Heroin to Chronic, Treatment-Resistant Methadone Patients in the Netherlands." *J Drug Issues* 29: 587-608.

Walsh, John M. 2004. "Are We There Yet? Measuring Progress in the US War on Drugs in Latin American." *Drug War Monitor* from the Washington Office on Latin America (December).

Bibliography

Ainslie, George. 1991. "Derivation of 'Rational' Economic Behavior from Hyperbolic Discount Curves." American Economic Review 81(2):334-340.

Ali, R., M. Auriacombe, M. Cases, L. Cottler, M. Farrel, D. Kleiber, et al. 1999. "Report of the External Panel on the Evaluation of the Swiss Scientific Studies of Medically Prescribed Narcotics to Drug Addicts." *Sucht* 45: 160-170.

Allais, Maurice and Ole Hagen (eds). 1979. Expected Utility Hypotheses And The Allais Paradox. Dordrecht. Holland: Reidel.

Bazos, Audrey and Jessica Hausman, "Correctional Education as a Crime Control Program". Applied Policy Project, UCLA School of Public Affairs, March 2003.

Beccaria, Cesare. 1986. On Crimes and Punishments; translated from the Italian in the author's original order, with notes and introduction by David Young. Indianapolis: Hackett Pub. Co.

Bentham, Jeremy. 1830. The Rationale of Punishment. London: R. Heward.

Becker, Gary and William Landes. 1974. "Essays in the Economics of Crime and Punishment." New York: Columbia University Press

Bickel, Warren K., Frank J. Chaloupka, Michael Grossman, and Henry Saffer (eds).1999. The Economic Analysis of Substance Use and Abuse: An Integration of Econometric and Behavioral Economic Research: (NBER-C) National Bureau of Economic Research Conference Report

Boyum, D.A. 1992. "Reflections on Economic Theory and Drug Enforcement." Ph.D. Thesis, Harvard University.

Bratton, William and Peter Knobler. *The Turnaround: How America's Top Cop Reversed the Crime Epidemic*. Random House. 1998.

Bureau of Justice Statistics. "Correctional Populations in the United States, 1997" Office of Justice Programs U.S. Department of Justice. November 2000, NCJ 177614

Bureau of Justice Statistics. 1991. Profile of Jail Inmates.

Bureau of Justice Statistics. 1996. Survey of State Prison Inmates.

Bureau of Justice Statistics. 2001. Prison and Jail Inmates at Mid-Year.

Burke, Peggy. "Policy-Driven Responses to Probation and Parole Violations", Center for Effective Public Policy. Silver Springs, MD. March 1997.

Butterfield, Fox. "Guns and Jeers Used by Gangs to Buy Silence." *The New York Times*. January 16, 2005.

Caulkins, Jonathan P., Peter C. Rydell, Susan S. Everingham, James R Chiesa and Shawn Bushway. 1999. *An Ounce of Prevention, A Pound of Uncertainty: the Cost Effectiveness of School Based Drug Prevention Programs.* Santa Monica, CA: RAND.

Caulkins, Jonathan P. 1990. "The Distribution and Consumption of

Illicit Drugs : Some Mathematical Models and Their Policy Implications." Ph.D. Thesis, MIT.

Cavanagh, D.P. and Harrell, A. "Evaluation of the Multnomah County drug testing and evaluation program." Final report to the National Institute of Justice. Cambridge: BOTEC Analysis Corporation, 1995.

Chaiken, J. M. and Chaiken, M. R. (1982). *Varieties of Criminal Behavior.* Santa Monica, CA: Rand Corp

DiIulio, John, "Reinventing Probation and Parole," *The Brookings Review* 1998 (4):40-42.

Dilulio, John Jr. and Ann Morrison Piehl. 1991. "Does Prison Pay? The Stormy National Debate Over the Cost-Effectiveness of Imprisonment." Brookings Review 28-35.

Ellsberg, D. 1961. "Risk, Ambiguity and the Savage Axioms." *Quarterly Journal of Economics* 75 (Spring): 643-669.

Eysenck, Hans J. 1989. Personality and Criminality: A Dispositional Analysis. In William S. Laufer and Freda Adler (Eds.) Advances in Criminological Theory , [1]:89-110.

Farrel, M. and W. Hall. 1998. "The Swiss Heroin Trials: Testing Alternative Approaches." *BMJ* 316:639.

Federal Bureau of Investigation. 2001. Uniform Crime Reports.
Gladwell, Malcolm. 2002. The Tipping Point: How Little Things Can Make

a Big Difference. Boston : Little, Brown.

Fredrickson, B.L., and Daniel Kahneman. 1993. "Duration Neglect in Retrospective Evaluations of Affective Episodes." *Journal of Personality and Social Psychology*, 65: 45-55.

Glaze, Lauren and Seri Palla. "Probation and Parole in the United States, 2003". Bureau of Justice Statitstics Bulletin NCJ 205336. Washington D.C. July 2004.

Grasmick, H. G.; C. R. Tittle; R. J. Bursik; and B.J. Arneklev. 1993. "Testing the Core Empirical Implications of Gottfredson and Hirschi's General Theory of Crime." Journal of Research in Crime and Delinquency 30(1):5-29.

Greenwood, Peter and Allan Abrahamse. 1982. Selective Incapacitation. Santa Monica: The Rand Corporation

Harrell, Adele, William Adams, and Caterian Gouvis Roman, "Evaluation of the Impact of Systemwide Drug Testing in Multnomah County, Oregon", Research Report. April 1994.

Herrnstein, Richard. 1983. Some Criminogenic Traits of Offenders. In James Q. Wilson (Ed.) Crime and Public Policy. San Francisco: Institute for Contemporary Studies.

Hey, J.D. 1995. "Experimental Investigations of Errors in Decision Making Under Risk." *European Economic Review* 39: 633-640.

Holzer, Harry J., Steven Raphael and Michael A. Stoll. 2003.

"Employment Dimensions of Reentry: Understanding the Nexus between Prisoner Reentry and Work." Paper presented at Urban Institute Reentry Roundtable, May, New York City, NY.

Howard, Ronald A. 1980. "On Making Life and Death Decisions." In Societal Risk Assessment: How Safe is Safe Enough? eds. Richard C. Schwing and Walter A. Albers, Jr. New York: Plenum Press.

Inciardi, James A. "Corrections-Based Continuum of Effective Drug Abuse." Research Brief. Washington, DC: U. S. Department of Justice, National Institute of Justice. June 1996.

Kahneman, Daniel and Amos Tversky
1973 Availability: A heuristic for judging frequency and probability. Cognitive Psychology 5:207-232.
1979 "Prospect Theory and Analysis of Decision Under Risk." Econometrica. 47(2):263-91.
1992 Advances in Prospect Theory: Cumulative Representation of Uncertainty. Journal of Risk and Uncertainty [5]:297-323.

Katz, Jack. 1988. Seductions of Crime: Moral and Sensual Attractions in Doing Evil. New York: Basic Books.

Kelling, George and James Wilson. March 1982. "Broken Windows: The Police and Neighborhood Safety." The Atlantic Monthly. Volume 249, [3]: 29-38.

Kennedy, David, "Pulling Levers: Getting Deterrence Right," *National Institute of Justice Journal*, No.236. July 1998,

Kennedy, David M. "Reconsidering Deterrence". A Report to the Office of Justice Programs. U.S. Department of Justice. December 2003

Kleiman, Mark. "Faith-Based Fudging: How a Bush-promoted Christian prison program fakes success by massaging data." Slate.com. August 2003.

Kleiman, Mark, Thomas H. Tran, Paul Fishbein, Maria-Teresa Magula, Warren Allen, and Gareth Lacy. "Opportunities and Barriers in Probation Reform: A Case Study in Drug Testing and Sanction." California Policy Research Center, Los Angeles, 2003.

Kleiman, Mark A. R., "Coerced Abstinence: A Neopaternalist Drug Policy Initiative," in Lawrence M. Mead (ed.) *The New Paternalism*, D.C: Brookings Institution, (1997).

Kleiman, Mark. "The Punishment Deficit and the Prosecutor's Job," Kennedy School Working Paper 87-02-01, 1987

Kleiman, Mark. 1993. "Enforcement Swamping: A Positive-Feedback Mechanism in Rates of Illicit Activity." Mathematical & Computer Modeling 17 (2):65-75.

Kleiman, Mark. 1992. *Against Excess: Drug Policy for Results*; New York: Basic Books.

Kleiman, Mark. 1989. *Marijuana: Costs of Abuse, Costs of Control*; Greenwich, Conn.: Greenwood Press.

Kleiman, Mark. 1997a. "Reducing the Prevalence of Cocaine and Heroin Dealing Among Adolescents." *Valparaiso University Law Review* 31(2): 551-564.

Kleiman, Mark. 1992. "Neither Prohibition Nor Legalization: Grudging Toleration in Drug Control Policy." *Daedalus* (Summer).

Kleiman, Mark. 1988. "Crackdowns: The Effects of Intensive Enforcement on Retail Heroin Dealing." In *Street Level Drug Enforcement: Examining the Issues,* ed. Marcia R. Chaiken. Washington DC: National Institute of Justice.

Levitt, Steven D. and Sudhir Alladi Venkatesh. 2000. "An Economic Analysis of a Drug-Selling Gang's Finances." *Quarterly Journal of Economics* 115 (3):755-789.

Loewenstein, George. March 1996. "Out of Control: Visceral Influences on Behavior." Organizational Behavior and Human Decision Processes 65[3]:272-92.

Longshore, Douglas. 1998. Self-Control and Criminal Opportunity: A Prospective Test of the General Theory of Crime. Social Problems 45(1):102-113.

Manski, Charles R., John V. Pepper, and Carol V. Petrie, eds. 2001. *Informing America's Policy on Illegal Drugs: What We Don't Know Keeps Hurting Us*. Washington DC: National Academy Press.

Moore, Mark H. Susan Estrich, Daniel Mcgillis, and William Spelman *Dangerous Offenders: The Elusive Target Of Justice*. Harvard University Press. 1985.

Nagin, Daniel and Greg Pogarsky. 2001. "Integrating Celerity, Impulsivity, and Extralegal Sanction Threats into a Model of General Deterrence: Theory and Evidence." Criminology [39]:865-892.

Naik, Ashish V., Alok Baveja, Rajan Batta, and Jonathan P. Caulkins.

1996. "Scheduling Crackdowns on Illicit Drug Markets." European Journal of Operational Research, Vol. 88, pp.231-250.

Perneger, T.V., F. Giner, M. Del Rio, and A. Mino. 1998. "Randomized Trial of Heroin Maintenance Program for Addicts who Fail in Conventional Drug Treatments." *BMJ* 317: 13-18

Petersilia, Joan, "Probation in the United Sates," *Crime and Justice, A Review of Research* , Vol 22, 1997, University of Chicago Press.

Piehl, Anne Morrison, "Learning While Doing Time," Kennedy School working paper #R94-25, Harvard University, April 1994.

Press, Aric. 1987. *Piecing Together the System: The Response to Crack*. New York: Bar Association

Raiffa, Howard. 1982 The Art and Science of Negotiation, Harvard University Press, Cambridge, Mass.

Raphael, Steven. 2003. "The Socioeconomic Status of Black Males: the Increasing Importance of Incarceration." Working Paper. Goldman School of Public Policy, Berkeley, CA.

Rehm, J., P Gschwend, T Steffen, F Gutzwiller, A. Dobler-Mikola, A. Uchtenhagen. 2001. "Feasibility, Safety, and Efficacy of Injectable Heroin Prescription for Refractory Opioid Addicts: A Follow-Up Study." *Lancet* 358: 1417-1420.

Reuter, P.; R. J. MacCoun; and P. Murphy. 1990. "Money from crime: A study of the economics of drug dealing in Washington D.C. Santa Monica, CA". RAND.

Reuter, Peter, Robert J. MacCoun, P.J. Murphy, Allan F. Abrahamse and Barbara Simon. 1990. *Money from Crime: A Study of the Economics of Drug Dealing in Washington D.C.* Washington DC, RAND publications

Reuter, Peter and Mark Kleiman. 1986. "Risks and Prices: An Economic Analysis of Drug Enforcement." In *Crime and Justice: An Annual Review of Research* Michael Tonry and Norval Morris (eds.), Vol. 7.

Rotter, Julien. **"**Generalized expectancies for internal versus external control of reinforcements" , *Psychological Monographs*, *80*, Whole No. 609 1966.

Rydell, Peter C. and Susan S. Everingham. 1994. *Controlling Cocaine: Supply Versus Demand*. Santa Monica, CA: RAND.

Schelling, Thomas C. 1978. Micromotives and Microbehavior. W.W. Norton: New York.

Schelling, Thomas. 2003. "Foreword" in Collective Choice: Essays in Honor of Mancur Olson, edited by Jac Heckelman and Dessin Coates Springer, Berlin

The Sentencing Project. "New Incarceration Figures: Rising Population Despite Falling Crime Rates". Report. www.sentencingproject.org/pdfs/1044.pdf. 2004.

Sourcebook of Criminal Justice Statistics 2004. "Table 6.54: Federal Prison Population, and Number and Precent Sentenced for Drug Offenses."

Spelman, William. "The Liminted Importance of Prison Expansion" in Alfred Blumstein and Joel Wallman, eds. *The Crime Drop in America*, Cambridge University Press, 2002.

Spelman, William. *Criminal Incapacitation*. Plenum Press. New York. 1994

Starmer, C. 2000. "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice Under Risk." *Journal of Economic Literature* 38: 332-382.

U.S. Department of Justice. "Truth in Sentencing in State Prisons," U.S. Department of Justice, Office of Justice Programs, January 1999, NCJ 170032.

US Probation and Pretrial Services System. "Year in Review Report, Fiscal Year 2003." August 2004.
US Sentencing Commission. 2002. "Report to the Congress: Cocaine and Federal Sentencing Policy May 2002."
http://www.ussc.gov/r_congress/02crack/ch6.pdf January 28, 2005.

Van den Brink, W., VM Hendriks, and J Van Ree. 1999. "Medical Co-prescription of Heroin to Chronic, Treatment-Resistant Methadone Patients in the Netherlands." *J Drug Issues* 29: 587-608.

Viscusi, W. Kip. December 1993. "The Value of Risks to Life and Health,"

Journal of Economic Literature, Vol. XXXI, (4):1912-1946.

Walsh, John M. 2004. "Are We There Yet? Measuring Progress in the US War on Drugs in Latin American." *Drug War Monitor* from the Washington Office on Latin America (December).

White, J.L.; T. E.Moffitt; A. Caspi; D. Bartusch; D. Needles; and M. Stouthamer-Loeber. 1994. "Measuring Impulsivity and Examining its Relationships to Delinquency." Journal of Abnormal Psychology [103]:192-205.

Wilson, James Q. and George L. Kelling. "Broken Windows". *Atlantic Monthly*. March 1982.

Wilson, James Q. 1975. Thinking About Crime. New York: Basic Books.

Wilson, James Q. and Richard J. Herrnstein. 1985. Crime and Human Nature. New York: Simon and Schuster.