Jonathan P. Caulkins and Peter Reuter

How Drug Enforcement Affects Drug Prices

ABSTRACT

Enforcement against drug selling remains the principal tool of drug control in the United States and many other countries. Though the risk of incarceration for a drug dealer has risen fivefold or more over the last 25 years in the United States, the prices of cocaine and heroin have fallen substantially. Different models of how enforcement affects drug supply may help explain the paradox. There are substantial periods in which drug markets are not in the stable equilibrium that has informed much of the empirical research. Enforcement is likely to be more effective in preventing the formation of a mass market than in suppressing such a market once it has formed. Once a mass market is established, there may be little return to intense enforcement. A modest level of enforcement may generate most of the benefits from prohibition.

The U.S. drug problem has lost some of its political salience since the late 1980s when it was briefly the principal domestic policy concern of the public, but every year there are more than 30,000 deaths from drug-induced causes and roughly 1 million drug-related emergency room visits. More than one-third of AIDS deaths are drug related. Seven and a half million Americans are believed to need drug treatment. The majority of arrestees test positive for at least one illegal drug, and illegal drugs generate an estimated $180 billion per year in social costs (Office of National Drug Control Policy [ONDCP] 2004, 2009).

The nation invests ever-increasing amounts in an effort to control illegal drug markets. In 2007 the federal government probably spent
$20 billion per year on drug programs, and it devoted almost two-thirds of federal prison space to locking up 100,000 drug offenders. State and local governments made more than 1.5 million arrests for drug offenses and incarcerated 400,000 individuals for drug offenses on any given day. A plausible estimate of the total annual government expenditure for drug control in the United States in the latter part of the first decade of the twenty-first century is about $40 billion, of which approximately three-quarters goes to enforcement, both here and overseas.

Drug enforcement, rather than the demand-side programs of prevention and treatment, is thus the cornerstone of drug policy in the United States. Surprisingly, enforcement also dominates drug control budgets in other countries for which budget estimates are available. This holds even in countries that explicitly embrace less punitive approaches or "harm reduction" regimes that give more attention to the consequences of drug use than to its prevalence in the general population. For example, Moore (2008) finds that enforcement accounts for 55 percent of the $1.3 billion Australia spent on proactive drug control programs, and the majority of the $1.9 billion spent dealing with the consequences of drug use pertained to crime-related consequences. Likewise, Rigter (2006) finds that enforcement accounts for 75 percent of the Netherlands' €2.185 billion spending on drug control in 2003. Both Australia and the Netherlands emphasize harms rather than prevalence as the most important outcomes.

Most enforcement effort and the great bulk of imprisonment for drug-related offenses are focused on people involved in drug distribution rather than on drug users themselves (Sevigny and Caulkins 2004). Though concerns about justice and retribution do play some role, the basic justification for this considerable investment of taxpayer funds and great sacrifice of liberty is the hope that constraining supply will limit drug use and dependence.

Enforcement is directed at all levels of the distribution chain, begin-

---

1 In 2003 the Office of National Drug Control Policy (ONDCP) quietly redefined what constituted federal drug control expenditures. Under the new definition, which emphasized those budget items that ONDCP claimed could be controlled ("proactive"), expenditures on prosecution and incarceration were excluded as "reactive." This led to a rebalancing of federal expenditures with a higher share classified as demand reduction, as well as a reduction in estimated total expenditures. The traditional definitions are more informative for present purposes; our $20 billion figure is an educated guess as to what the current budget would be under the traditional reporting rules. For comments on this, see Walsh (2004).
ning with supporting crop eradication and destruction of cocaine processing in Peru, Bolivia, and Colombia, and also arrest of high-level traffickers in Colombia. Even more is spent interdicting drugs at the border and in the “transit zone” between source countries and the United States. There are other types of enforcement, such as programs designed to control the chemical precursors of synthetic drugs such as methamphetamine and diversion of pharmaceuticals that can be abused as recreational drugs. Yet the largest investment is in arrest, prosecution, and incarceration of people involved in drug distribution within the United States. Higher-level domestic wholesale dealers get longer sentences, but there are so many more retail dealers and low-level employees such as couriers that “kingpins” account for a minority of drug-related incarceration.

This essay examines how all of this drug enforcement affects the price of drugs. We focus on that relationship because it is through price and availability that tough enforcement should reduce drug use in the population. We also focus on enforcement aimed at sellers both because that is where most of the resources go and because enforcement against users presents a different set of analytic issues: it constrains demand rather than supply.

The essay draws on a much expanded literature and data that have appeared over the 25 years since *Crime and Justice* published “Risks and Prices” (Reuter and Kleiman 1986, hereafter R&P), the first systematic effort at relating drug enforcement and prices. The essential logic of R&P is that drug dealers are in business to make money, so when enforcement imposes costs on drug dealers, those costs are passed along to users in the form of higher prices. That is easy to understand with respect to seizure of drugs, money, and other assets, but R&P extends the argument to monetary compensation for the nonmonetary risks. Just as miners and deep-sea divers are paid more than construction workers doing otherwise comparable work at relatively safer work sites above ground, R&P likewise attributes a portion of drug dealers’ monetary profits to compensation for the risks of arrest and incarceration and also the risk of injury or death at the hands of other criminals.

Twenty-five years later there is no reason to retreat from that fundamental insight, but it now seems incomplete. The traditional R&P argument remains sensible on average in the long run, but a lot of other things can happen on the way to that long run. This essay seeks
to describe additional ways that enforcement and also general market
dynamics can influence price, mechanisms that can trump the original
R&P mechanisms for long enough to be of consequence in at least
some markets. Though our focus is on the United States, we make use
of experiences and data from other Western nations, particularly Aus-
tralia and the United Kingdom.

The first section is descriptive. It presents time-trend data for three
aspects of cocaine and heroin markets: enforcement (as measured by
incarceration), prices (adjusted for purity and inflation), and consump-
tion, particularly as consumption relates to criminally active offenders.
Over the period since 1980 incarceration rates (per thousand people)
have risen by almost an order of magnitude, while the prices of both
cocaine and heroin have fallen substantially. The number of dependent
users fell between 1988 and 2000, the only period for which estimates
are available, and we conjecture that this decline may have continued.

Sections II through V are the heart of the essay. They present a
variety of models for understanding the relationship between enforce-
ment and price. Section II starts with the simple static equilibrium
model that was posited in R&P, explaining how, in the long run,
tougher enforcement might raise price. Section III then considers a
dynamic model that allows for disequilibrium as a market moves to-
ward a new, long-run equilibrium. The subsequent section then dis-
cusses models connecting enforcement to emerging drug markets,
while Section V considers enforcement during disequilibrium. Section
VI reviews empirical evidence supporting the explanatory value of
these models, while Section VII briefly presents some policy implica-
tions.

The static model of R&P is helpful in explaining some important
characteristics of the major drug markets. For example, the fact that
most of the earnings, though not the large individual incomes, are from
the low level of the distribution system may be explained by the fact
that retail dealer risks are distributed over many fewer grams of the
drug than are the risks of high-level dealers and smugglers.

However, there is a need for dynamic models to account for how
the market behaves in times of rapid change. For example, it is char-
acteristic of psychoactive drugs that an epidemic of initiation starts with
explosive growth for a period of months, if not a few years; supply may
lag for a number of reasons in adjusting and R&P does not do a good
job of accounting for price variation during this time. We also conjec-
ture that there are multiple equilibria in many drug markets and these
dynamic models show why it is easier to prevent a market at a lower
equilibrium (with few users and high price) from tipping to a high
equilibrium than it is to tip a mass market back down to a lower equi-
librium.

Most of the advantages of prohibition can be attained with modest
levels of overall enforcement coupled with targeting of dealers whose
behavior poses a particular risk to the community (e.g., use of juvenile
distributors, violence against competitors). Tougher enforcement
across the board does not, as a matter of historical observation, seem
to raise price much farther or restrict availability; it imposes rather
high individual and social costs. However, enforcement may have pre-
vented the emergence of mass markets for various new and attractive
drugs. Unfortunately, the very nature of emerging markets makes it
difficult to test this proposition empirically.

It is necessary to limit our domain. We say little about efforts to
control production in source countries simply because there has been
little development since 1986 in understanding that aspect; the pessi-
mistic arguments and judgments of R&P (namely, that interventions
in source countries would have minimal effect on retail markets in the
United States) hold up well with the historical record.2 Prices in source
countries account for only 1–2 percent of retail prices in developed
countries. So even if alternative development, crop eradication, or en-
forcement in source countries quintupled prices in source countries,
the effect on retail prices downstream could be modest. Indeed, prices
of coca leaves in Peru and opium in Afghanistan have gyrated dra-
matically with little or no corresponding observable effect on street
prices for crack in Los Angeles or heroin in London. Sometimes
source-country interventions reduce production in one country, but
unfortunately there seems to be no shortage of peasant farmers and
criminals in relatively lawless regions who are willing to take up the
slack.

The discussion of markets is restricted to what is necessary for un-
derstanding the effects of enforcement on prices. We say much less
about drug markets than did R&P; there is now a very large literature
on the markets themselves, which we have surveyed elsewhere (Babor

---

2 Paoli, Greenfield, and Reuter (2009) provide an extensive review for heroin source
and transit country control efforts. A briefer analysis that also includes cocaine and
amphetamine type stimulants (ATS) is provided in Reuter and Trautmann (2009).
et al. 2010, chap 5; see also Bushway and Reuter 2008). Moreover, some observations about markets—such as their generally competitive rather than monopolistic character—were new and controversial in 1986 and so required extensive discussion, whereas now they are widely accepted. We do not provide systematic discussion of availability because there are few data. Marijuana gets relatively little mention because it is less causally related to crime than are cocaine and heroin.

I. Drug Enforcement, Prices, and Consumption in the United States, 1980–2010

Over the last 30 years the most striking observation about drug markets is that the number of persons serving time in prison for drug offenses in the United States has risen steadily and substantially, while the prices of cocaine and heroin, adjusted for purity, have declined. The price decline was sharp during the 1980s and has been gradual since then. Since there is evidence of a shrinking cocaine and heroin market (discussed below), it is plausible that the risk of incarceration for a cocaine seller in 2005 was more than five times higher than in 1985, even though the price was only 25 percent as high (Fries et al. 2008). Methamphetamine prices have also trended down, with some notable spikes discussed below.

Superficially this is a paradox. Greater pressure on supply has been unexpectedly accompanied by declining, not rising, prices. Technically it is enforcement intensity, not the total number of imprisonments, that R&P predicts should drive up prices, and while variation in demand could play a role, in the rest of this section we argue that the basic paradox remains even after accounting for these issues. This challenge to R&P was a primary motivator of research over the last two decades to improve understanding of drug markets’ response to enforcement.

Critics of U.S. drug policy often report trends in total drug arrests to paint a picture of ever-increasing toughness, since total arrests have been increasing steadily since 1980. That is misleading, if not disingenuous. It is rarely sensible to combine marijuana arrests with heroin and cocaine arrests in any analysis. The substances are very different in terms of usage patterns, health risks, enforcement practices, and notably in the consequences of arrest. The vast majority of arrests for marijuana possession, which account for about half of all drug arrests,
result in no sentence of incarceration, even to local jail. These arrests present a very different policy issue than do the long terms of incarceration faced by sellers of cocaine, heroin, and methamphetamine.

The rapid increase in cocaine and heroin arrests during the 1980s reflects a deliberate policy choice, and it drove correspondingly large increases in imprisonment. The increase in marijuana arrests starting in the 1990s did not. The George W. Bush administration did make marijuana a priority, but arrests grew faster under the Clinton administration, and in neither period did the arrests lead to big increases in imprisonment. Some of the increase in marijuana arrests stemmed from quality-of-life policing, such as New York City’s policy of arresting people for using marijuana in plain view (Golub, Johnson, and Dunlap 2007).

There are different trends for arrests for different drugs and for sales versus possession. Heroin and cocaine arrests rose exponentially during the 1980s to peak in 1989. They fell back down when cocaine markets were disrupted in 1990, and never resumed a consistent upward trend; sales arrests in particular have been ebbing (down one-third from the 1990 peak) as markets have become less flagrant.

Marijuana arrests were stable during the 1980s, and then fell briefly before nearly doubling in the 1990s. There was another lull in growth from 1999 to 2002 before they resumed their upward trend. These substantial shifts in marijuana arrest patterns primarily reflect changes in policing practices, not underlying changes in production, distribution, or consumption patterns.

Figure 1 shows this by plotting the number of persons arrested annually for marijuana offenses and also for either cocaine or heroin (combined) from 1979 to 2007. Figure 1 shows only total marijuana arrests because in all years only 11–21 percent of those arrests were for sale as opposed to possession. For heroin and cocaine the figure shows sales offenses alone as well as the total because the proportion of the total that was for sales changed more over time. The Uniform Crime Reporting program (UCR) does not separate cocaine from heroin offenses, a curious tribute to tradition since there has been intense interest in each separately for at least 25 years (Federal Bureau of In-

---

3 The only study that we know of that traces the incarceration risks for marijuana possession arrestees is Reuter, Hirschfield, and Davies (2001) for three large counties in the state of Maryland. They report no sentence of jail time, but one-third spent at least one night in jail pretrial.
vestigation 2009); UCR data also lump methamphetamine into a heterogeneous “other” category, so figure 1 includes no line for methamphetamine, though by all accounts those arrests grew rapidly in the 1990s.

The arrest figures understate the increase in enforcement intensity over the last quarter century. Arrest is just a part of enforcement and is only a modest penalty by itself, with incarceration being the more substantial penalty to which supply and price trends should respond. However, figure 2 shows that the total number incarcerated for 1980–2002 increased, based on Caulkins and Chandler (2006). The story here is much simpler; the number of those locked up in all three types of correctional facilities (federal prison, state prison, and local jails) rose without cease throughout the period. The growth has been much faster for the state prison population than for local jail inmates. Whereas in 1985 the figures were about even for each, by 2002 the state figure was 225,000 people, whereas that for jails was only 140,000.

---

![Graph showing drug arrests (in thousands), 1979–2007. Source: Federal Bureau of Investigation (2009), Crime in the United States, various years.](image-url)
This may represent increasing severity of penalties meted out for drug offenders.

Sevigny and Caulkins (2004) used the 1997 Survey of Prison Inmates to show that the great majority of those who reported that they were incarcerated for drug possession offenses were involved in drug distribution, albeit often in minor roles. For some, the possession conviction involved a plea bargain. Also, of the 90,000 they estimated to be in state prison for possession (i.e., facing a sentence of more than 12 months), only about one-third said that their actual offense was possession of less than 10 retail units of drugs. One-sixth of them reported that they had a “major role” in drug selling. No analyses have been published for those locked up in jails for drug offenses; a higher share of them are presumably there for simple possession offenses, which rarely result in felony conviction, but the earlier statement that the bulk of incarceration, and the growth in the number, is dominated by selling offenses still holds here.

These arrest and imprisonment figures by themselves give no information about what has happened to the intensity of drug enforcement.
For that one needs to know the ratio of enforcement effort relative to the size of the drug market. For example, it is plausible that cocaine markets expanded even faster than the number of incarcerations for cocaine selling did during the first half of the 1980s (Reuter 1991), in which case the risk faced by a cocaine seller could have been less in 1985 than in 1981. However, that is unlikely for the remainder of the period. Indeed, it appears that total sales (quantity of cocaine and number of transactions) fell through the 1990s (ONDCP 2001). For our purposes, the extent of the decline is a second-order consideration. All we need is for the reader to accept that the stringency of enforcement against cocaine and heroin selling has risen sharply over the period since 1980.

How have prices responded to this increase in enforcement intensity? Cocaine, heroin, and methamphetamine prices are most usefully reported after adjusting for purity since much of the effective variation in price comes through changes in purity (e.g., Caulkins 1994; Rhodes et al. 2007). For example, if 1 gram of a powder that is 50 percent cocaine by weight costs $100, the purity-adjusted price would be $200 per pure gram.

Figure 3 shows that the purity-adjusted price of cocaine declined over the period 1981–2007, expressed in constant dollars. Though there are occasional spikes, none lasted more than a few quarters and none were proportionately large. The big decline occurred in the 1980s, during a period of market expansion. There was not much change during the 1990s, but the decline from 2000 to 2007 was substantial in percentage terms, approximately 25 percent.

The sharp decline in price might be expected to generate a rise in consumption, as even drug users show demand sensitivity to price (Grossman 2005). Yet the available indicators, though not strong, show instead that total consumption has probably declined (ONDCP 2001). Certainly the number of past-year users has declined, at least as reflected in annual surveys of the household population, high school students, and young adults (Johnston et al. 2009; Substance Abuse Mental Health Services Administration [SAMHSA] 2009). It is much harder to measure total consumption. People can more reliably report whether they have used than how much they have used, and, more important, total consumption is dominated by the small subset of past-year users.

\(^4\) No estimates have been published after 2001, but an update through 2006 is in preparation.
who are heavy or dependent users. Such individuals are more likely to fall outside the surveys’ sampling frames (e.g., because they are homeless) or be nonrespondents.

Several teams have stitched together time series for demand by combining various complementary data sets and modeling assumptions (Everingham and Rydell 1994; ONDCP 2001; Caulkins, Behrens et al. 2004). The general finding is that consumption grew in the 1980s even though prevalence was falling because growth in the number of dependent users more than offset the steep declines in light or occasional users. Since the early 1990s there has been a slow ebbing of demand as the cohort with peak exposure ages, dies, or eventually gives up use. This is a slow process, and there is still some, albeit smaller, inflow of new users, so demand today is probably still within a factor of two of its peak and will not diminish rapidly any time soon.

Figure 4 shows past-month marijuana and cocaine and past-year heroin use as measured by the National Survey on Drug Use and Health since 1974. Prevalence increased during the 1970s and fell during the 1980s. Recorded cocaine and marijuana prevalence is running about 25 percent higher since 2001 than in the 1990s. That may be due to

---

Data obtained from SAMHDA’s online data analysis site, http://www.icpsr.umich.edu/cocoon/SAMHDA/DA53/00064.xml.
changes in survey methodology, including increased response rates after the introduction of $30 payments to incentivize participation in 2002. For example, the number of people reporting having ever used marijuana grew by 10.5 million between 2001 and 2002 even though only 3 million persons reported using for the first time in 2002 (SAMHSA 2003).

Figure 4 also illustrates the enormous differences in prevalence across substances. Marijuana is by far the most commonly used substance, and even past-year heroin use barely registers on the graph when that use is self-reported by the household population. Fewer than one in 200 persons over the age of 12 reported that they had used heroin in the past 12 months in 2007.

Figure 5 shows that the estimated numbers of chronic users of cocaine and heroin have declined in the period 1988–2000, in both cases

---

*Defined as having used on 8 or more days in the previous 30 days.
by almost exactly one-third. The heroin numbers are scaled by multiplying by three both to make the heroin trends easier to see and also to highlight how closely the cocaine and heroin trends parallel each other.

These estimates were produced by sophisticated modeling of some weak data sets, and reestimation throughout the 1990s led to substantial changes in the estimates for specific years. For example, the prior set of estimates (ONDCP 2000) found that there had been an increase for heroin in the mid-1990s, almost exactly canceling the decline from 1988 to 1993; that rebound is missing in the most recent published estimate.7 Thus these estimates should be seen as indicative rather than precise. However, for our purposes what is important is that results agree that there was no expansion in the size of the cocaine and heroin markets from 1988 to 2000. We can safely assume that the massive increase in incarceration outpaced increases in market size, and so indeed generated a very large rise in the legal risks faced by a dealer.

We now turn to the main focus of this essay, assessing four different models of how enforcement affects the purity-adjusted price of drugs. Section II addresses markets in static equilibrium, meaning that prices

---

7 Updated estimates were prepared for ONDCP but were never published. See Reuter (2008).
are stable at a level such that the quantity suppliers produce exactly matches the amount customers wish to buy at that price. More precisely, it updates the “risks and prices” framework for comparing two markets, each in equilibrium but differing in the intensity of enforcement. Such analysis attributes difference in price and consumption across the two markets to the difference in enforcement intensity.

The general conclusion from this “comparative statics” analysis is that illegality plus a modest level of enforcement can keep prices much higher and availability much lower than would be the case if there were no prohibition or enforcement. However, for an established market, increases in enforcement beyond that basic level (akin to the enforcement intensity in the United States in 1980) have only modest further effects on price, availability, and use. We take up these policy implications in the concluding section.

The next sections take account of important complications that this model does not consider. Section III discusses what happens when a market for a given substance in a given place and time can have more than one stable equilibrium level of price and use, for example, because the market must attain a certain minimum viable size in order to create general availability. At least in theory, aggressive enforcement may be able to tip a market from a high-level equilibrium to a low-level equilibrium or, more modestly, prevent a market at a low-level equilibrium from tipping up to a high-level equilibrium. We offer some examples of the latter but none of the former; the scarcity of instances of shifting a mass market to a new, small-market equilibrium points to the difficulty of that task.

Section IV extends the analysis to a market that is in the process of moving from one stable equilibrium to another, such as one tipping from a low-use to a high-use equilibrium. Even though price and quantity may be changing over time, they are on an equilibrium path in the sense that at any given moment in time, the market price balances supply and demand. Borrowing a term from thermodynamics, we call them quasi-static equilibria because the changes occur relatively gradually—meaning they take a year or more. The market always clears (meaning users willing to pay the current market price can locate a supplier), and no one is ever surprised by the next week’s average market price. The price may be trending higher or lower, but it does so smoothly. This model is important in helping understand price trends...
in the United States during the 1980s and more generally when a mass market for a drug first emerges.

In Section V, we turn attention to markets that are not in equilibrium, typically due to some shock or surprise. The extreme form of disequilibrium is a market that does not clear, meaning the drug simply is not available no matter how much buyers are willing to pay. The more common form of disequilibrium is when a shock to the system quickly moves the market clearing (purity-adjusted) price and quantity substantially away from the baseline price and quantity, but then price and quantity move again to an equilibrium (whether static or quasi-static). Usually the market returns to its original equilibrium, but in principle if the shock is large enough it could “tip” the market to a different equilibrium. “Quickly” in this context means purity-adjusted price is pushed away from its original equilibrium within a few weeks or months, most often through an adjustment to purity, not (just) a change in the price per raw gram, unadjusted for purity.

A comparison to oil and gasoline markets may be useful. When oil prices spiked to $147 in July 2008, that price did not reflect a long-run equilibrium. However, motorists were always able to buy gasoline, albeit at very high prices. In contrast, during the 1973 oil embargo, sometimes gasoline was simply not available. Both events reflect shocks that pushed prices away from their long-run equilibrium, but only in 1973 were there instances of the market not clearing. It is rare for markets not to clear, but price spikes and troughs are not uncommon in commodity markets, whether legal (e.g., natural resources and farm products) or illegal.

II. Drug Enforcement’s Effect on Long-Run Equilibrium Prices: Risks and Prices Theory

Application of the standard competitive model to drug markets is now common in introductory economics texts (e.g., Frank and Bernanke 2004). The most basic concepts in economics are those of demand and supply. A demand curve maps the relationship between price and the quantity consumers are willing to purchase. The supply curve maps the relationship between price and the quantity that producers (distributors) are willing to produce (provide). The market clears at the price at which the quantity demanded by consumers equals the quantity that distributors are willing to supply, that is, at the point where the supply
and demand curves cross. Interventions can be thought of in terms of how they affect the supply curve or the demand curve and the resulting equilibrium price and quantity.

Consider what the model suggests are the effects of restricting supply, for example, by incarcerating more drug dealers. The standard analysis is that because dealers will only sell in return for more money, the supply curve shifts up and to the left, so that at any given price less is supplied. As figure 6 shows, this increases the market price and reduces the quantity of drugs sold.

Both the magnitude of the reduction in quantity sold and the resulting effect on dealers’ revenues depend on the elasticity of demand, which is related to the slope of the demand curve. When the demand curve is relatively flat (high elasticity of demand, as in the left-hand panel of fig. 1), quantity consumed falls a lot, so dealers’ revenues decline. If the demand curve is steep (low elasticity, right-hand panel), quantity falls by less, proportionately, than prices rise, so restricting supply actually increases dealers’ gross revenues.

Some drug-related consequences are driven by consumption, including overdose, dependence itself, and crime caused by intoxication. Others are driven more directly by spending on drugs; those include corruption, the economic-compulsive crime that users commit to raise funds, and systemic crime between dealers (Goldstein 1985). Hence, the elasticity of demand is perhaps the single most important parameter.
determining how effective restricting supply is at ameliorating drug-related harms.

Economists have done extensive work on estimating the price elasticity of demand for various drugs, particularly cocaine and marijuana, including the cross-elasticity between illegal drugs and cigarettes or alcohol (e.g., Cameron and Williams 2001; Grossman 2005; Jofre-Boignet and Petry 2008). Almost all serious studies find that consumption responds to changes in price. However, the various studies using different methods and different measures of use produce a range of estimates of just how responsive consumption is to price changes.

Still, the overall logic model should be clear. Supply control drives up prices. Higher prices reduce consumption. The change in consumption, together with the change in price, determines the change in spending. The change in consumption, together with the change in spending, determines the change in drug-related social costs.

“Risks and prices” theory provides the foundation for this logic model by elaborating the simple textbook model of how enforcement imposes costs on drug suppliers. To simplify, enforcement acts almost like a tax. For example, if year in and year out law enforcement seizes one-third of all drugs, the drug distribution system will respond by shipping more drugs, 1.5 kilograms for every kilogram to be consumed. However, the ongoing cost of replacing those seized drugs makes it more expensive to deliver a given quantity to users. Drug dealers are in business to make money, so they pass those costs on to users in the form of higher prices, under plausible assumptions about the structure of drug markets.

Indeed, it is hard to identify any large costs of delivering drugs to final users other than those directly related to illegality and enforcement. One indication of that proposition is that the share of cocaine’s retail price accounted for by growing and refining it is less than 1 percent. A gram of a drug that sells for over $100 on the streets of Chicago (in units of less than one gram) leaves Colombia at a cost of

---

8 The price elasticity of demand is the percentage reduction in demand caused by a 1 percent increase in price; except under exceptional conditions it is negative. There is a similar price elasticity of supply, usually positive. The cross-elasticity of marijuana and alcohol is the percentage change in the demand for marijuana caused by a 1 percent increase in the price of alcohol; the cross-elasticity will be positive if the marijuana and alcohol are substitutes and negative if they are complements. Demand is labeled “inelastic” if the absolute value of the elasticity is smaller than one, since it means that a rise in price will lead to a rise in total revenues.

The risk-compensation aspect is somewhat hidden in this decomposition. For example, importation (12 percent of the total) is almost all risk compensation. FedEx will deliver a kilogram of a legal good from Bogotá to Miami for less than $50. In contrast, the price of a landed kilogram of cocaine in Miami is around $15,000, compared to an export price of less than $1,500. That difference between $50 and $13,500 is a consequence of the current prohibition system. (It does not represent monopoly rents; smuggling services are provided in a competitive market.)

Labor costs are high in part because illegality forces the distribution system to operate in inefficient ways. Drugs are diluted and packaged by hand. Simple machines could do that work at a small fraction of the cost (as is done for mass-produced goods such as sugar) except that enforcement makes it risky (i.e., expensive) to maintain dedicated, fixed capital equipment. Sellers need to have protection against robbery, look out for police, and work with small, easily exhausted stock because it is too risky to maintain large quantities at the point of sale.

Two studies provide some data on the rate of sales per dealer. Reuter, MacCoun, and Murphy (1990) estimate that a daily cocaine retailer sold a median of $3,600 worth of drugs each month in 1988 in Washington, DC. The same individuals were estimated to have spent 66 hours per month selling and to earn a median of $7 per hour in the licit labor market. They sacrificed about $462 (66 times $7) in legitimate earnings per $3,600 sold. As Caulkins and Reuter (1998) show, this yields a labor time cost of about 13 percent of the retail value to just compensate the seller’s time. Levitt and Venkatesh (2000), working with the financial records of a drug selling gang in Chicago in the mid-1990s, estimated that the average sales volume per dealer per month was only $1,000 and that the wages per “foot soldier” were about $470. The implied number of sales transactions per day per dealer was only two, suggesting that the costs of time compensation might be higher than estimated in the Reuter et al. study, though the legitimate earnings of this younger and more marginalized population may have been lower in the Levitt and Venkatesh study.

For policy purposes, given the effort that goes into interdicting drugs

---

This assumes that they could have found paid employment for the extra hours.
on their way to the United States, a key analytic issue is how changes in the import price affect retail prices. The R&P model can help understand what parameters are central for that purpose. Assume that a kilogram of heroin sells for $50,000 at import in El Paso and the equivalent of $500,000 per kilogram when sold at retail units of 200 milligrams in Chicago. Now imagine that more effective interdiction boosts the import price by 50 percent to $75,000. How would this affect the retail price of heroin in Chicago?

One theory of vertical price relationships, which Caulkins (1990) has termed the “additive model,” argues that the import price is essentially a raw material cost (Reuter and Kleiman 1986). Thus, the wholesaler who previously bought heroin at $50,000 and now pays $75,000 has had roughly $25,000 added to the per-kilo costs. The actual cost increase will be somewhat more than $25,000 per kilo, due to seizures, thefts, and other losses: wholesalers have to buy more than one kilo of drug for each kilo they sell.

In a competitive market, the wholesaler simply passes these increased costs along to the next stage of the distribution chain. The buyer at this stage will thus face an increase of $25,000+ in costs, which will be passed along to the next market level. Eventually, the $25,000+ cost increase reaches consumers, and the end result is that the retail price of heroin increases by $25 per gram (or somewhat more, when all the losses and seizures along the distribution chain are factored in), which is only about 5 percent of the initial retail price. Doing this exercise for cocaine and imported marijuana in the United States, Reuter, Crawford, and Cave (1988) projected, based on some assumptions that were reasonable but not empirically tested, that a doubling of the import price results in only a 10 or 20 percent rise in the retail price. Or, put another way, it would take a quintupling of import prices to effect a doubling of retail prices.

Of the few empirical analyses of interdiction, most have assumed the additive model. And because the replacement cost of seized cocaine and heroin is a small fraction of their final retail price—as little as 1 percent in source countries and usually no more than 20 percent at the point of entry into the final market country—these analyses have concluded that the potential contribution of interdiction to the reduction of drug abuse is small, or at least that it is not cost effective compared to domestic enforcement and treatment (Rydell and Everingham 1994). Although the additive model is conceptually compelling, some scholars
have noted that it does not fit very well with some historical price data
drawn from the lower end of the distribution chain (Caulkins 1990; Boyum 1992).

Some historical price data appear to be more consistent with what
Caulkins has called the “multiplicative model” of vertical price rela-
tionships, which holds that a change of a certain percentage in price
at one stage of production or distribution brings about a similar per-
tcentage change at subsequent stages. The idea behind the multiplica-
tive model is that many of the costs of doing business in the drug trade,
such as the risk of employees and other dealers stealing drugs, are more
strongly related to the value of drugs bought and sold than to the
weight or quantity trafficked. Thus, the multiplicative model predicts
that if the import price of heroin rises 50 percent, from $50,000 to
$75,000, then suppliers’ costs all the way down the distribution chain
will also increase by 50 percent. So retail prices will also rise 50 per-
cent, from $500,000 to $750,000, providing a much more favorable
assessment of interdiction.

Cocaine prices in the 1980s and early 1990s followed a remarkably
consistent multiplicative relationship between kilogram level and retail
prices (Caulkins 1990, 1994). Likewise, Rhodes, Hyatt, and Scheiman
(1994), Crane, Rivolo, and Comfort (1997), and DeSimone (1998) ar-
gue for a multiplicative model. However, the data that indicated a mul-
tiplicative model were far from conclusive (DeSimone 2006). Except
during occasional shortages, prices consistently declined during this
period, and so it is possible that the factors causing the decline operated
at all levels of the market. In other words, it may be that declines in
retail prices were not so much caused by the decline in import prices,
but rather that both import and retail prices were influenced by other
factors. For example, the growth in the cocaine industry internation-
ally, and the development of crack markets domestically, may have cre-
ated economies of scale that lowered the costs of wholesale and retail
operations.

In short, the nature of vertical price relationships in drug markets is
an open question. The answer may lie somewhere between the additive
and multiplicative models. Indeed, Caulkins (1990) and Boyum (1992)
have speculated that the multiplicative model may hold when the
drugs’ price per unit is very high (e.g., for retail cocaine and heroin
transactions in the United States and Europe), but that the additive
model may apply when prices are less extreme (e.g., for higher-level
TABLE 1
Components of the Costs of Cocaine, 1990

<table>
<thead>
<tr>
<th>Component Activity</th>
<th>Percentage of Total Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wholesale price in Colombia</td>
<td>1</td>
</tr>
<tr>
<td>Importing of drug</td>
<td>12</td>
</tr>
<tr>
<td>Retail labor</td>
<td>13</td>
</tr>
<tr>
<td>Higher-level labor</td>
<td>~3</td>
</tr>
<tr>
<td>Drug and asset seizure</td>
<td>8–11</td>
</tr>
<tr>
<td>Money laundering fees</td>
<td>2–4</td>
</tr>
<tr>
<td>Packaging, processing, and inventory costs</td>
<td>~2</td>
</tr>
<tr>
<td>Compensation for risk of prison</td>
<td>24</td>
</tr>
<tr>
<td>Compensation for physical risk</td>
<td>33</td>
</tr>
<tr>
<td>Total</td>
<td>~100</td>
</tr>
</tbody>
</table>


It is clear that interdiction imposes considerable costs on drug traffickers; import prices are much higher than they would be if borders were unpolicied. What is uncertain, besides the import-retail price linkage, is whether sizable changes in the interdiction budget would have commensurate effects on import prices. It may be that a 50 percent reduction in interdiction funding would be inconsequential because the remaining effort would be sufficient to induce traffickers to continue to take costly avoidance actions. However, it is also not clear whether even current interdiction spending can preserve the current markups in the long run given increasing globalization of commerce (Stares 1996; Costa Storti and de Grauwe 2009).

We conclude this discussion of the simple risks and prices framework with four observations. First, risk compensation due to violence affects price determination, but that role tends to get short analytic shrift, probably due to the paucity of data available. However, the original R&P noted that it is a risk for which drug dealers need to be compensated. As noted in table 1, Reuter, MacCoun, and Murphy (1990) estimated that compensation risk for violent victimization accounted for one-third of the retail price of cocaine in Washington in the late 1980s. That, however, was a period when the markets were near their peak and had attracted many young men who had high propensities for violence and may not be generalizable across time. Violent crime...
in the United States generally has declined markedly since the early 1990s, and drug market violence may have fallen substantially as well. The aging of cocaine and heroin users, as indicated in treatment data (e.g., Trunzo and Henderson 2007), suggests that these markets are now populated by older participants, who are probably less violent than they were when younger. For example, in 1992 data on treatment admissions for cocaine, 40 percent of these clients were under the age of 30. By 2006, that figure had dropped to 26 percent, while the fraction of clients over the age of 40 rose from 15 percent to 47 percent between 1992 and 2006. Furthermore, Caulkins, Reuter and Taylor (2006) offer a theoretical model of drug dealing in which the removal of violent dealers, either through incarceration or homicide, might drive down equilibrium prices. Thus, one force driving down cocaine and heroin prices in the United States may be a decline in compensation for physical risk. However, there is little that can be said empirically on this matter.

Second, David Boyum (1992) offers an important insight, namely, that drug dealing “firms” (almost) never have negative accounting profits. As a consequence of illegal drug markets, the normal Darwinian selection winnowing out inefficient firms, namely, that they lose money, works very weakly. For example, a large rise in the sentence for selling more than one kilogram of cocaine, that should in principle make selling at the current price unattractive and lead some dealers to exit the business and others to raise the wholesale price, may only have those effects with a long lag if it takes time for those in the drug-selling business to learn about this change.

Third, R&P presumes people will move in and out of drug dealing in response to balancing the risk-adjusted return from the dealing against their potential earnings from other activities (whether legal or not). Since R&P was published, advances in behavioral economics and behavioral decision theory have shown repeatedly that people do not always act in ways that maximize their expected gain, even in relatively simple laboratory settings. Decision heuristics routinely fail in certain contexts. People misjudge probabilities (Kahneman, Slovic, and Tversky 1982), discount future events “too heavily” (Laibson 1997), are distracted by irrelevant information, and fail to anticipate how happy or unhappy they will feel in future hypothetical states (Loewenstein, O’Donoghue, and Rabin 2003). These limitations on human decision
making are particularly pronounced for decisions to use or to sell drugs, since both offer immediate rewards with delayed, uncertain consequences. Dependent users’ decisions are further clouded by the cravings triggered by withdrawal (Badger et al. 2007). Caulkins and MacCoun (2005) suggest a partially decoupled relationship between drug enforcement and prices. R&P may still apply in the long run as people do gradually move in or out of the market after they finally recognize that it is in their interest to do so, but the adjustment periods may be lengthy.

Fourth, from the perspective of a dealer, enforcement outcomes are random events, akin to fires and traffic accidents in daily life. R&P stresses the idea that greater enforcement intensity increases the expected cost of enforcement for dealers, but enforcement also creates variation in outcomes across dealers because some get arrested but others do not. The point is perhaps best illustrated with some stylized numbers. Suppose enforcement creates for a certain category of dealer a 50 percent risk of receiving a prison term for which those dealers would demand $500,000 in compensation. In other words, in this hypothetical example we imagine that for these dealers $500,000 is the bare minimum they would need to receive—and get to keep—in order to voluntarily serve a prison term for selling. R&P predicts that all dealers in that category will demand $250,000 in additional monetary compensation, since $250,000 equals the 50 percent probability times the $500,000 consequence. However, not all dealers “pay” an enforcement cost of $250,000. Indeed, none do. After the uncertainty about which 50 percent of dealers will be punished, it turns out that half of the dealers are winners and half are losers. The lucky 50 percent made an additional $250,000 without going to prison; the other half are losers who may have made $250,000, but that does not compensate them for the long prison term they received. Both winners and losers at the drug-dealing game are bad for society. Winners make visible, bad role models for impressionable youth who have weak legitimate economic alternatives, and losers suffer in obvious ways. Levitt and Venkatesh (2000) investigate this tournament quality of dealers’ wages empirically, and find that there are many losers and just a few big winners, just as with a familiar state lottery and representing the well-known optimism bias that influences so much behavior.
III. Enforcement’s Ability to “Tip” Markets from One Long-Run Equilibrium to Another

Drug markets often experience rapid changes, in particular, periods of explosive growth. Risks and prices is a static framework not well suited to accommodate that phase of market development, which is of particular interest for policy purposes. Are there methods for preventing the growth? For epistemological reasons it is hard to identify law enforcement successes here. Markets that remain small attract neither general public attention nor good measurement.

Many models of markets have multiple stable equilibria separated by an intermediate “tipping point” at which a modest push effects rapid change toward one or another equilibrium. This presents interesting opportunities to time and scale enforcement for the greatest effect. Unfortunately, markets may grow beyond the tipping point quickly, before the associated use has been identified as a “problem situation.”

Tipping points have been popularized by Gladwell’s (2000) book, but multiple equilibria have long been a mainstay of models of drug markets and drug epidemics (Kleiman 1988, 1993; Baveja et al. 1993; Caulkins 1993; Tragler, Caulkins, and Feichtinger 2001). The scientific notion of a tipping point is more precise than what Gladwell describes (Grass et al. 2008), but the essence is similar.

Stable equilibria are conditions that the system will tend to stay in naturally, without a great deal of control effort being exerted. Many systems have only one such equilibrium: the classic static market described in figure 1. Others systems have more than one condition that tends to persist over time.

Models from development economics offer a familiar example (Skiba 1978). Some macroeconomists speak of very poor countries being stuck in a “poverty trap.” What they mean is that, even should the income gap between affluent countries and middle-income countries tend to diminish over time through “convergence” or technological catch-up, prospects for the world’s “bottom billion” may be less favorable (Collier 2007). Those countries, according to some schools of thought, are trapped in a different, much poorer stable state than the stable high-income state enjoyed by developed countries.

If there are two or more states the system can remain in stably over time, and if at the moment the system is near a balance point in between two of them, the “tipping point,” then even a little shove can have very large effects if it tips the system into a trajectory to approach
one equilibrium instead of another (what in lay jargon is referred to as a “butterfly effect”).

For drug markets, the basic intuition is that when few people are using or selling the drug, it is relatively easy to keep the drug from spreading, so that is one (low-use) equilibrium. Reasons include the following: in a “thin” market, sellers and buyers have a hard time locating each other, behaviors that are uncommon are more likely to be socially stigmatized, and a modest level of enforcement effort can create high risks for each of the relatively small number of people over whom that effort is distributed (Kleiman 1993). The other extreme is a high-volume equilibrium, when the drug is very widely used. That situation may also be stable because having one more or one fewer person use might have very little effect on the behavior of others. Potential users will be offered the drug multiple times anyway and, if they choose to use, will be able to locate multiple suppliers (Riley 1997). So there are not feedback effects that amplify movements away from the high-volume equilibrium. In between, however, there may be a tipping point where the market is still of modest size, but close to reaching a critical mass that will enable it to spread widely.

Kleiman (2009) lays out the general theory behind multiple equilibria, but some of the key ideas are captured in his (1993) paper on “enforcement swamping.” Drug enforcement’s effect on the risk of trafficking a kilogram and, hence, the resulting price per kilogram depends not on the total level of enforcement effort but rather on the enforcement intensity per kilogram sold. When the market is small, a given level of enforcement effort can make selling quite risky, but if the market grows tenfold while enforcement resources grow only fivefold, then the enforcement risk experienced per dealer or per kilogram sold falls by 50 percent.

Dynamic analysis of drug markets is a relatively recent phenomenon, and empirically validating and parameterizing models is very difficult. However, in almost all models constructed to date, the tipping point is not halfway between the low- and high-use equilibria; it is much closer to the low-level equilibrium. That is, only with very low levels of use are those low levels stable. If use rises to even a modest fraction (perhaps one-tenth or one-fifth) of the long-run steady state, then that is enough to place the system on a trajectory approaching a higher-use steady state.
If one believes drug markets operate in this way, the implications are as follows: If a market is at a low-level equilibrium, it should take little enforcement effort to hold it there, and that effort may be a tremendously cost-effective investment, even if the drug does not seem to be very threatening, precisely because it is not widely used. Conversely, if a market is at a high-level equilibrium, suppressing use is difficult because one is pushing against the tide. Moderately strenuous efforts might reduce consumption somewhat for a time, but as soon as the extra efforts are relaxed, use will rapidly rise back to that stable, high-level equilibrium. A very aggressive effort sustained long enough might eventually tip the market back to its low-level equilibrium, but it would take great determination. Finally, if one happens to catch the market after it has begun moving toward the high-level equilibrium but before it has approached that equilibrium, then it might be wise to act very quickly and decisively, because it is much easier to stuff a genie that is halfway out of the bottle back in than it is to corral a genie after it has escaped completely.

There are very few examples of entire countries being able to put a genie back in the bottle. Communist China’s suppression of its opium market in the 1950s is perhaps the last example. Paoli, Greenfield, and Reuter (2009) note that the irredentist movement controlling the dominant opium producer in Myanmar managed to reduce production there drastically in the period 1998–2007. It is not clear whether suppression of the U.S. cocaine and heroin markets between 1900 and 1930 counts. That might instead be seen as an original imposition of prohibition that drove users out of the market who would not participate in a black market, rather than an increase in enforcement tipping a large, already illegal market down to a small market equilibrium.

There are more instances at the local level, where the two equilibria are not between levels of use but rather between levels of flagrancy in dealing (each supporting roughly the same amount of use). Surreptitious dealing is one possible equilibrium. Even if there are 1,000 dealers in a city, if all are hidden from the police and the public eye, and you become the only dealer brazenly selling on a street corner, then your drug dealing career will be very short.

---

10 Both episodes involved use of coercion not compatible with even a modest level of civil rights and democratic process. We know of no similar success on the part of a democratic government.
Unfortunately, another equilibrium is one in which all or most dealers sell brazenly in public markets. Such selling does not make them immune to arrest. Indeed, when there are flagrant street markets, the police sometimes can make more arrests than the rest of the criminal justice system could handle. However, the arrest risk per dealer may be tolerably low if physical presence in the street market increases opportunities to obtain customers. The risk is a bit akin to small fish swimming in schools. It is easy for big fish to see the school, so schooling does not deter predation. But, from the perspective of an individual fish, it decreases the individual likelihood that it gets eaten.

Enforcement does have instances of success tipping large black markets from high flagrancy equilibrium into the lower equilibrium. There are a variety of case studies of aggressive “crackdowns” suppressing at least the flagrancy of dealing activity (Zimmer 1987; Kleiman 1988; Caulkins, Larson, and Rich 1993; Braga et al. 2001), and a complementary set of theoretical papers on the topic (Baveja et al. 1993; Caulkins 1993; Tragler, Caulkins, and Feichtinger 2001). More generally, the idea that markets can be tipped from brazen to covert modes of operation is a plausible interpretation of the changes in local U.S. markets between the 1980s and 2000s when flagrant, place-based markets supporting anonymous transactions were largely replaced by transactions arranged through some electronic contact (cell phone, beeper, etc.).

Perhaps enforcement has many more instances of success tipping small black markets into low equilibrium, or expunging them entirely. After all, very many substances are prone to abuse. Yet a rather modest number of substances account for most use of purely illegal drugs (as opposed to diverted pharmaceuticals). These standbys (marijuana, opiates, cocaine/crack, and certain of the amphetamine type stimulants) may be widespread because they are more appealing than other drugs, but it is also possible that law enforcement has instead succeeded in preventing an otherwise robust market from forming for these other substances as well: PCP, GHB, Rohypnol, LSD, ketamine, methcathinone, peyote, and mescaline. If so, then law enforcement might be very effective at controlling those smaller markets, and we have no way to know because the prevention of a potential outcome, or counterfactual and nonevents, are intrinsically hard to document or disprove.
IV. Enforcement’s Effects on Emerging Markets

The general conclusion of the R&P model is that after simple illegality, backed by enough enforcement to realize the “structural consequences of product illegality” (Reuter 1983), drives prices quite high, expanding drug law enforcement against an already established market has little further effect on the price of illegal drugs. The purpose of this section is to note that this conclusion may not apply to new or emerging markets.

That is not to say that there is strong evidence that enforcement is able to drive up prices in emerging markets. Rather, there is little empirical evidence one way or another. However, there are deductive arguments suggesting that enforcement might be relatively more successful at keeping prices high in small or emerging markets, at least for a time. We discuss those hypotheses after first clarifying what we mean by an emerging market.

A. Concept of an Emerging Market

Drug use emerges over time in ways that reflect what is sometimes called an “epidemic cycle.” Initially, drug use is rare, but then something on the supply (increased availability) or demand side (e.g., a counterculture or youth movement) kicks off a period of rapid expansion in drug use. Drug use diffuses by word of mouth, with existing users introducing friends and associates to the substance (Ferrence 2001). Such “contagious” spread tears through networks quite literally to create exponential growth. That rapid growth cannot go on indefinitely. There are competing explanations for what typically brings an end to the rapid growth, including exhaustion of the pool of individuals susceptible to initiation and lagging negative feedback that begins once enough users escalate to dependent or problematic use to give the drug a reputation for being dangerous (Musto 1987; Behrens et al. 1999, 2000; Rossi 2001).

Whatever the mechanism, initiation rates for many substances drop substantially from their initial peaks. For some drugs, prevalence (the percentage of the population using) also spikes up above (“overshoots”) its long-run equilibrium. As time goes on, the mix of users shifts as well, with a growing number of dependent users. Nondependent users generally remain in the majority by number, but dependent users consume so much more per capita that they come to dominate consumption by quantity of drug. In the United States today, dependent users...
account for much more than half of the consumption of all the major illegal drugs (heroin, cocaine/crack, methamphetamine, and cannabis). For example, by ONDCP (2001) estimates, the relatively small number of cocaine users who are “hard core” account for 84 percent of total spending on the drug; the corresponding figure for heroin is 93 percent. Once initiation has passed its peak and demand becomes dominated by dependent users, drug use is less likely to vary sharply over time; dependent users make stable customers.

The markets can likewise evolve in terms of both size and sophistication. Obviously, a growing number of drug users implies a growing total quantity supplied. Larger drug markets may be more efficient and more robust to enforcement actions. While the exact threshold of size needed for an emerging market to resist enforcement efforts is not clear, the qualitative result is that although it may be cheaper to find and arrest people when markets are large, the cost effectiveness of arresting an additional dealer may be higher when markets are small.

Also, drug production and distribution can become more sophisticated over time, with smugglers constantly developing new and more sophisticated methods (submarines, tunnels under the border, embedding drugs within other materials, etc.). Drug enforcement also innovates. Indeed, drug smuggling is often depicted as something of an “arms race” between smugglers and law enforcement. However, there is some reason to believe that smugglers’ risk of arrest is particularly high the first few times that an organization tries a particular route or tactic (Caulkins, Burnett, and Leslie 2009), and enforcement’s early edge might tend to erode over time.

A counterpart of sophistication at the production and smuggling levels can be “professionalization” of retail distribution, meaning an increasing share of demand is supplied by people who derive much of their income from drug distribution. Professional selling can be distinguished from distribution within social networks, where the retail transaction occurs between people who interact for reasons other than the distribution of drugs and the supplier has some other principal source of income.\footnote{By this definition, stereotypical street markets for crack or heroin involve professional sellers, whereas most marijuana users obtain their drugs within social networks. Indeed, Caulkins and Pacula (2006) report that 89 percent of household survey respondents report obtaining their marijuana most recently from a friend or relative. However, the small numbers of users who obtain their marijuana from professional sellers tend to be disproportionately heavier users.}
There is little written about the supply side of nascent or emerging drug markets. Thus our description is largely inferential. It does suggest, however, that emerging markets may be more fragile to various kinds of enforcement efforts.

B. Why Enforcement Might Be More Effective Against Small or Emerging Markets

There are three distinct reasons why enforcement might be relatively more effective at constraining small and/or emerging markets than it is at addressing larger, more established markets: differential effects on dealers’ costs, dynamic growth, and epidemic effects.

1. Differential Effects on Dealers’ Costs. In the comparative statics (“risks and prices”) framework outlined above, enforcement affected retail prices by increasing dealers’ costs. Enforcement can also increase dealers’ costs in an emerging (dynamic or “quasi-static”) market, and may do so more efficiently.

Late in the epidemic, there may be a steady stream of convicted drug traffickers being released from prison. If, as is likely, their prison record makes it difficult to secure legitimate employment, they may see few options other than a return to dealing. Such “barriers to exit” from dealing might make replacing newly incarcerated dealers easier for an established market than in an earlier one. The released traffickers can act as replacements to new arrestees, creating a revolving door where investment in arrests simply maintains the status quo.

A more subtle factor is that the distribution network is more robust in larger markets. Drug distribution networks are embedded within a larger social network.12 The larger the drug market, the denser is the embedded distribution network. Denser networks have more redundant linkages, making it harder for enforcement to isolate a subnetwork through judicious elimination of certain nodes or arcs. In the language of engineers who study network reliability in telecommunications and other fields, dense networks are more “robust” to disruption. Also, the denser the network, the better the information flows are likely to be and the more like a competitive market it becomes. If a denser network means that buyers at various levels know more sellers, then they are

---

12 Social networks can be visualized as a sea of dots—called “nodes”—connected by lines (“arcs”) indicating that the two nodes at either end of that arc interact with each other. In the embedded distribution network, an arc would represent a supplier-customer relationship.
better able to force those sellers to compete on price, driving the price bar down farther on.

2. Greater Effects When Supply Is Constrained. Distinct from these cost effects is the possibility that, early on in a drug epidemic, supply capacity at some market levels might simply not be able to keep up with exponential growth in demand. So eliminating some of the supply network’s scarce throughput capacity (e.g., by arresting a particular seller) might cause a greater disruption than would the same arrest 5 years later, when the market has stabilized and has excess supply capacity.

Constrained capacity is unlikely at lower distribution levels where it is relatively easy for outsiders to break into the business. However, not just anyone can decide to be a successful international smuggler or high-level drug dealer. The main barrier to entry does not appear to be skill per se. Studies of incarcerated sellers (e.g., Reuter and Haaga 1989; Matrix Knowledge Group 2007; Decker and Chapman 2008) often report that drug dealing requires few specialized skills. Even high-level domestic distribution is essentially just brokerage activity. One needs a source and customers, but there is next to no physical processing of the drugs.

However, a certain amount of tacit knowledge seems to be required, and the primary source of this tacit knowledge is by working with an existing dealer. Existing dealers can spread this knowledge to others intentionally (e.g., by introducing a friend to the business when demand exceeds what one person can supply) or unintentionally (e.g., when a former employee begins to sell on his or her own).

This suggests that expanding higher-level supply capacity in an emerging market should not be thought of as drawing on a limitless pool of labor that rapidly flows in to bid away wage differences whenever profits from dealing exceed profits available from legitimate work. Rather, when selling is profitable (e.g., because there is excess demand) then existing dealers spawn new dealers, akin to the way existing users are the principal vector for inducing initiation of new users. Hence, when the current number of sellers at these higher market levels is very small, the inflow of new dealers is also small. In short, there may be “word of mouth” diffusion of drug dealing throughput capacity, at least at the rate limiting market level.

The key insight is that if drug use spreads faster (is “more contagious”) than is drug selling, then during the explosive stage of a drug
epidemic, the ratio of demand to supply may grow very large.\textsuperscript{13} In particular, if the rate of growth for both is proportional to the current numbers, as in standard diffusion models, then the ratio of customers to supply would tend to grow exponentially over time.

The ratio of customers to sellers cannot grow exponentially forever. At some point the relative shortage of distribution capacity limits the rate at which drug use can spread. When supply capacity limits the spread of use, eliminating sellers at that constrained market level could directly constrain growth in use. Later on, once use has hit a plateau and distribution capacity has had time to catch up, eliminating the same number of sellers might have much less effect. More precisely, the “risks and prices” deterrence effect exists both early and late, but the incapacitation effect of leaving insufficient distribution network throughput capacity only pertains early in an epidemic.

There is no appreciable incapacitation effect late in an epidemic because most dealers in mature markets could quickly increase the quantity distributed. If the demand were there, they could simply ask their supplier for more, and the supplier would in turn ask his or her supplier for more, up the distribution chain. That is, in equilibrium, lost network capacity can be made up not only by bringing in new dealers but also by having the remaining dealers sell more per unit time.

3. Multiplier Effects. The previous subsections gave reasons why it is easier for enforcement to constrain supply earlier in a drug epidemic. It is also plausible that a given constraining of supply has a bigger effect on drug use when the constraining happens early in an epidemic. That is, if one could conceptualize one “unit” of supply reduction (e.g., an increase in the market clearing price by a given amount for a given duration of time), then early in an epidemic it may not only be easier (cheaper) for enforcement to produce that effect, but also that unit reduction in supply might yield a larger reduction in drug consumption.

This could happen if initiation and light use are more price responsive than is dependent use because early in an epidemic, initiation is

\textsuperscript{13} In symbols, let \(U(t)\) and \(S(t)\) denote the number of users (demand) and selling or throughput capacity, respectively, where we specifically mean the capacity at the most constrained market level (e.g., the import level). If growth is proportional to current numbers, then both demand and throughput capacity grow exponentially (\(U(t) = e^{\alpha t}\) and also \(S(t) = e^{\beta t}\), and so the ratio of customers to supply would also grow exponentially over time since \(U(t)/S(t) = e^{\alpha t - \beta t}\) and \(\alpha > \beta\).
high and dependent use is low (Caulkins, Behrens et al. 2004). Also, an intervention that directly prevents one person from using may indirectly affect the use of others. The magnitude of this so-called social multiplier effect can vary over the course of a drug epidemic (Caulkins et al. 2002). For example, Caulkins, Behrens et al. (2004) observe that one proxy for infectiousness, the ratio of new initiations per current user, changed quite dramatically over the course of the U.S. cocaine epidemic, and simulations of drug epidemics reach the same conclusion (Winkler et al. 2004; Caulkins et al. 2009). Hence, early on, when drug use is spreading rapidly, eliminating one initiation may prevent a cascade of subsequent initiations in the same way that preventing one case of an infectious disease can avert additional secondary infections.

C. Summary

Drug markets evolve over time. Drug use often follows an epidemic cycle, and drug distribution networks mature. A modeling literature (e.g., Tragler, Caulkins, and Feichtinger 2001) and a variety of armchair arguments suggest that supply control may be more effective earlier in this process than later, after drug use has become endemic and the markets have matured.

Hence, although we are fairly confident in the central conclusion of the preceding sections, namely, that it is difficult for enforcement to suppress endemic use that is supported by a large, well-established market, we recognize that this conclusion may not apply to small or emerging markets. Unfortunately, there is limited potential to replace the many conjectures in this section with empirical evidence. The great bulk of data and analysis pertain to drugs whose use is endemic and whose markets are mature.

Also, as in the previous section, there is a problem of measuring dogs not barking. How can we measure instances in which enforcement tempers or delays an emerging market? Consider, for example, methamphetamine use in Missouri. Methamphetamine is widely used in Missouri today, but it was not 15 years ago, even though methamphetamine was already popular on the West Coast. It is hard to imagine that Missourians in 1995 intrinsically liked meth less than did Californians estimated from use self-reported in surveys tend to be higher than those based on measures associated with heavy or dependent use (e.g., Dave 2008). Evidence on the elasticity of initiation relative to the elasticity of nondependent use is weak, but a recent meta-analysis (Cicala 2006) reports cocaine participation elasticities that are slightly higher than elasticities for current users.
nians in 1995 or Missourians in 2010. More plausibly, the big differences in use are attributable to differences in availability and price.

Legal consumer goods tend to diffuse very quickly, so it is possible that enforcement deserves credit for sparing Missouri from some number of meth-related deaths and other problems for a decade or more. More generally, it seems entirely plausible that prohibition plus enforcement can at least temporarily constrain some emerging drug markets, even though there is next to no direct empirical evidence of quality one way or another.

V. Reducing Supply through Disequilibrium Effects

Introductory economic texts focus on equilibrium in markets, yet markets—including drug markets—can also be out of equilibrium. Disequilibrium presents challenges and opportunities both analytically and for enforcement choices. The question of interest here is, how can and how has enforcement affected disequilibrium prices in drug markets? Ideally we would also like to know whether such enforcement is a cost-effective form of drug control, but by and large the literature lacks empirical estimates of cost effectiveness.

A. Historical Examples of Market Disruptions

Markets for illicit drugs are frustratingly resilient, but there have been significant disruptions. In one extreme case, purity-adjusted prices at least tripled and overdoses fell by as much as 80 percent within 3 months during the Australian heroin drought (Moore et al. 2005). There is a modest literature debating whether it was supply control or other circumstances that created that shock (e.g., Weatherburn et al. 2003; Degenhardt, Reuter et al. 2005; Wood et al. 2006a, 2006b; Woodak 2008), and a larger literature documenting the resultant effects on outcomes such as overdose, crime, and the use of other substances in Australia (Baker et al. 2004; Longo et al. 2004; Degenhardt, Conroy et al. 2005).

Supply shocks include not only literal shortages, when customers are physically unable to locate a supplier, but also large spikes in (purity-adjusted) prices, most often taking the form of troughs in purity. (If

15 One exception to this is discussion of the cobweb cycle for agricultural markets, where a high price in season 1 leads growers to plant large quantities in season 2, which leads to low prices and thus decisions to plant smaller quantities in season 3, and so on.
the price spike is large enough to induce additional suppliers to enter the market, then by definition the market is not in equilibrium.)

Supply shocks and disequilibrium are the exceptional cases. At any given time, prices in most drug markets are stable or exhibit some long-run trend (Caulkins, Behrens et al. 2004), and the great majority of users usually report success at being able to locate a drug supplier (National Institute of Justice 2003). However, there are enough exceptions to merit study.

In the United States, every major drug has witnessed significant market disruptions. There was the 1969 marijuana shortage (McGlothlin, Jamison, and Rosenblatt 1970; Gooberman 1974; Craig 1980); the heroin drought of the early 1970s associated with the Turkish opium ban, a literal drought, and the “French connection” case (DuPont and Greene 1973); multiple rounds of methamphetamine shortages associated with precursor controls (Cunningham and Liu 2003, 2005, 2008; Dobkin and Nicosia 2009); and the cocaine shortage of 1989–90 as well as a possible, briefer one in 1995 associated with the Peruvian air bridge interdiction (Crane et al. 1997; Caulkins et al. 2009).

The earlier cocaine shortage apparently stemmed from a combination of U.S. efforts and the “war” between the Colombian government and the Medellin-based traffickers. It led to a sharp (50–100 percent at its peak) increase in cocaine prices that lasted about 18 months (Crane, Rivolo, and Comfort 1997; Caulkins, Pacula et al. 2004). The “Peruvian air bridge” refers to a period in which the Peruvian Air Force, assisted by U.S. agencies, began to intercept planes flying coca paste from Peru to Colombia; that had previously been the dominant source of cocaine, but the markets have since innovated around that tactic. It is harder to identify instances of true market gluts with excess supply, but some authors have mentioned the possibility in Australia with respect to heroin (Dietze and Fitzgerald 2002) and MDMA (Hall, spillane, and Camejo 2000).

Two facts limit historical data on market disruptions. First, price reports are usually annual; market disruptions, by contrast, are relatively brief, usually lasting less than one year. Second, most sources report prices not adjusted for purity, and often most of the effective price change comes about through changes in purity (Caulkins, Pacula et al. 2004; Caulkins 2007; Caulkins et al. 2009). Still, table 2 summarizes price changes associated with some of the more prominent
### Table 2

<table>
<thead>
<tr>
<th>Country</th>
<th>Drug</th>
<th>Baseline Year</th>
<th>Baseline Price ($)</th>
<th>Disruption Year</th>
<th>Disruption Price ($)</th>
<th>Change in Price (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>Cocaine</td>
<td>1989</td>
<td>190</td>
<td>1990</td>
<td>235</td>
<td>+24</td>
</tr>
<tr>
<td>United States</td>
<td>Crack</td>
<td>1989</td>
<td>198</td>
<td>1990</td>
<td>255</td>
<td>+28</td>
</tr>
<tr>
<td>United States</td>
<td>Meth</td>
<td>1994</td>
<td>160</td>
<td>1995</td>
<td>254</td>
<td>+59</td>
</tr>
<tr>
<td>United States</td>
<td>Meth</td>
<td>1997</td>
<td>178</td>
<td>1998</td>
<td>256</td>
<td>+44</td>
</tr>
<tr>
<td>Australia</td>
<td>Heroin</td>
<td>2000</td>
<td>750</td>
<td>2001</td>
<td>2,250</td>
<td>+200</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2002</td>
<td>1,500</td>
<td>+100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2003</td>
<td>1,500</td>
<td>+100</td>
</tr>
</tbody>
</table>

Source.—Caulkins, Pacula et al. (2004b) for U.S. data and Moore et al. (2005) for Australia.

market disruptions that have occurred in times and places with relatively good price monitoring.

We focus on purity-adjusted prices as an outcome that combines enforcement’s effects on purity and price into a single metric. Thus, halving purity at a given price per gram and doubling the price per gram when purity is held constant are seen as equivalent changes in the effective price for users. This is sensible from a theoretical perspective. Diluents and adulterants typically account for a trivial share of the dealers’ cost of producing and distributing drugs (as shown in table 1), and users primarily care about pure milligrams of the principal intoxicant. It also has empirical support. Purity-adjusted prices often correlate much more strongly with other indicator series such as treatment admissions and emergency department mentions than do prices not adjusted for purity (e.g., Hyatt and Rhodes 1995).

### B. Mechanisms of Supply Disruption

Given this description of the types of market shocks, the next question is, how can policy go about creating them? There are at least two ways: eliminating in one fell swoop a large proportion of suppliers at some critical market level, and eliminating for all suppliers a common operating practice or tactic that forces everyone to adapt. The former

---

16 Diluents such as mannitol are psychopharmacologically inert, so they merely dilute the mixture; adulterants such as caffeine have some psychopharmacological effect. The distinction is more important for some amphetamine type stimulants (e.g., when ecstasy is “cut” with amphetamine) than for cocaine or heroin.
is the goal of crackdowns in which multiple arrests are executed simultaneously (Kleiman 1988). A good example of the latter would be imposing controls on precursor chemicals used in the drug manufacturing process (Cunningham and Liu 2003, 2005, 2008). Other examples include eliminating the Peruvian air bridge (Crane, Rivolo, and Comfort 1997) and Operation Intercept’s closing of the U.S.-Mexican border to marijuana smugglers (and many others) in 1969.

There are no clear examples of the simultaneous arrest of large numbers of dealers disrupting a market. The challenges faced by this tactic when employed at the retail level become clear when one considers the numbers. Roughly 1 million Americans sell cocaine at retail every year (Caulkins and Reuter 1998), so an average metropolitan area with a population of 1 million (i.e., the size of Raleigh, North Carolina, or Birmingham, Alabama) might be expected to have about 3,300 cocaine sellers. Police departments consider crackdowns to be large if they simultaneously arrest a few hundred people at a time, which would still be only about 10 percent of retail sellers in such a metropolitan area.

Strategies that credibly threaten arrest without actually arresting very large numbers may be more effective (e.g., Braga et al. 2001) and avoid overwhelming the criminal justice system (see Zimmer 1987). David Kennedy has effectively promoted such efforts; for example, in High Point, North Carolina (a city of 100,000), police accumulated credible evidence for cases against well-established drug sellers and used that evidence to persuade them to quit the business (Kennedy 2009). Similarly, arrests or threat of arrests are likely more effective in small or isolated markets. Grimm (2009) offers a credible account of the drying up of LSD supply around 2001 as the result of the incarceration of a few key producers.17

Parallel arguments apply with respect to disrupting crop production for the crop-based drugs (cocaine/crack, marijuana, and opiates including heroin). The number of market participants at a given market level decreases as one moves up from retail to upstream activities, but only to a point. The level of the greatest market concentration (i.e., smallest number of participants and organizations) may be at export (e.g., from Colombia) and smuggling and transshipment countries (e.g., through

17 Grimm’s (2009) account stresses the role of social networks in the distribution of LSD. Grateful Dead concerts were important events for distributors, and the death of Jerry Garcia in 1995 created a major problem for them, exacerbated when the band Phish stopped touring in 2000 (p. 7). The final blow seems to have been the imprisonment of the largest producer in 2001.
Mexico). It is not even clear whether these narrower parts of the distribution system constitute true “bottlenecks”; even for smuggling of cocaine and heroin there are many competent potential entrants.

However, what is clear is that still further up the distribution chain the funnel begins to broaden out again (Babor et al. 2010, chap. 5). There are many small-time processing labs that convert coca leaves into paste and paste into base. The number of farmers cultivating crops is another order of magnitude larger still. For Afghanistan it is estimated that over 2 million individuals were involved in poppy growing and opium harvesting in the middle of this decade (Paoli, Greenfield, and Reuter 2009).

It is perhaps not surprising that source-country control efforts have rarely had any discernible effect on the quantity of drugs available for consumption downstream in high-income countries. The most notable exceptions are the Turkish opium ban, which in conjunction with the breaking of the French connection contributed to a heroin drought in the United States in the early 1970s, and the Taliban’s opium eradication campaign in Afghanistan, which appears to have affected prices in European markets (Paoli, Greenfield, and Reuter 2009). Even for cocaine, whose production is concentrated in one continent and has long been the focus of U.S. efforts, coca farming did not begin to decline until 2002 (United Nations Office on Drugs and Crime [UNODC] 2004), and prices continued to decline at least until 2007, as shown in figure 3.18

For countries with defensible borders, such as Australia and Canada, interdiction is different than source-country control in this regard, even though both are “international” programs. Interdiction has been credited with occasional, significant, short-term disruption of U.S. cocaine markets, notably through 1989–90 and mid-1995 (Crane, Rivolo, and Comfort 1997), and it is the most plausible explanation for the Australian heroin drought (Degenhardt, Reuter et al. 2005).

Note that seizures in and of themselves should not generally be expected to disrupt the market unless they are extremely large since usually suppliers can easily replace the lost drugs at wholesale costs (Reuter 1988). If the seizure is associated with dismantling an organization that had a substantial share of market throughput capacity, then one might

---

18 Peruvian production began to fall in the mid-1990s, and Bolivian production in the late 1990s, but until recently expansion in Colombia made up the difference. U.S. government figures are slightly different but tell the same basic story.
observe a correlation between seizures and price changes, as Smithson et al. (2004, 2005) found for the Canberra (Australia) metropolitan area. However, inasmuch as seizures are just a proxy for an unmeasured, true causal variable (destruction of throughput capacity), such a relationship will not always be apparent (see DiNardo 1993; Yuan and Caulkins 1998).

Three large heroin seizures in 2000 involving a multinational smuggling organization serving Australia and Canada instruct us in this regard. The first seizure (126 kilograms seized in January 2000 in the Bangkok airport) is not claimed to have affected markets in either Australia or Canada. Indeed, it evidently did not put even that particular smuggling organization out of business. The second (~100 kilograms on September 2 in Vancouver) did not lead to detectable changes in Vancouver’s heroin markets over the next 30 days as compared to the previous 30 days (Wood et al. 2003). However, Royal Canadian Mounted Police investigation of the seizure provided evidence that led back to a Fijian front company (Prime Success Enterprises) that was then dismantled in October. That operation led to the third seizure (357 kilograms in October), which, in the AFP account, “rendered ineffective a sophisticated concealment methodology, identified a legitimate cargo stream, and removed some very important facilitators [smugglers]” (Hawley 2002, p. 48). This has been credited by some (e.g., United Nations Office for Drug Control and Crime Prevention 2002), though by no means all (see Wodak [2008] for a contrary view), with precipitating the Australian heroin drought by convincing Asia-based trafficking syndicates to look for easier distribution markets.

C. Effects of Short-Term Disruptions

It is easy to dismiss market disruptions as inconsequential, since in most cases prices return to equilibrium levels within a year or two (Dobkin and Nicosia 2009). The price increases do not last because suppliers adapt by modifying tactics and operations or by replacing lost resources (Reuter 1988; Reuter, Crawford, and Cave 1988). For example, at one time or another over the last 25 years, four different regions have been the principal supplier of heroin to the United States (Mexico, South America, Southwest Asia, and Southeast Asia). Similarly, in the late 1970s Colombia quickly replaced Mexico as the principal supplier of marijuana to the United States in response to paraquat
spraying and fears of adverse health effects of using sprayed marijuana (Kleiman 1992).

However, even transient price increases can have meaningful effects. Lives are saved even if overdoses decline only temporarily, as hospital data suggest they did after the 1989/1990 cocaine supply disruption, the 1995 interdiction of the Peruvian air bridge, and the imposition of controls on chemical precursors for methamphetamine (Crane, Rivolo, and Comfort 1997; Cunningham and Liu 2003; Dobkin and Nicosia 2009).

Furthermore, some effects of transient disruptions may be long-lasting. Tighter supplies can induce users to enter into drug treatment (Weatherburn and Lind 2001), and some of those individuals may not return to the former patterns of use, even once supply is restored. Likewise, Day, Degenhardt, and Hall (2006) document the Australian heroin drought’s effects on initiation. As Moore (1990) notes, most people initiate within a narrow band of ages. If they pass through that “window of vulnerability” during a supply shortage, their initiations may be prevented, not merely delayed.

However, not all of the effects of market disruptions are positive. For example, transient shortages have been blamed for increasing needle sharing and, hence, the spread of HIV (Maher and Dixon 2001), and they can lead both users and sellers to substitute one drug for another (Cunningham, Liu, and Muramoto 2008). Kleiman (1992) wonders whether successful interdiction of marijuana (which is relatively bulky and hard to conceal) induced smugglers to substitute into cocaine. Some have raised similar concerns about the Australian heroin drought inducing substitution into amphetamine use (Longo et al. 2004; but see Snowball et al. [2008] for an opposing view).

Perhaps the most common concern pertains to the effect on users’ spending on drugs. Recall that some drug-related consequences are driven by consumption (e.g., overdose); others are driven by spending on drugs (e.g., economic-compulsive and systemic crime). Constraining supply shifts back the supply curve, leading to higher prices and lower consumption. Whether drug spending goes up or down depends on how great a percentage decline in consumption is caused per 1 percent increase in price, a parameter that is called the elasticity of demand.

The elasticity of demand is always negative. (Use goes down when price goes up.) If the elasticity is smaller than 1.0, in absolute value,
then constraining supply will actually increase, not decrease, spending on drugs (and thus drug dealers’ revenues).

Everything in the paragraph above applies equally to supply restrictions that are sustained (shifting the long-run equilibrium) as well as shocks. However, even if the elasticity of demand is large (in absolute value) in the long run, drug use may be less price responsive over the course of a transient price spike. The few empirical studies that investigate this possibility for illegal drugs seem to support this concern.\(^{19}\)

We close by mentioning two more speculative potential effects of market shocks, both of which would enter on the positive side of the ledger. First, there are mathematical models of drug epidemics incorporating endogenous nonlinear feedback for which a temporary change in drug use can have lasting effects (e.g., Behrens et al. 1999, 2000; Winkler et al. 2004; Bultmann et al. 2008). There are also four noteworthy instances of drugs spreading rapidly up until a major market disruption and then stabilizing thereafter: the mid-1970s Turkish opium ban and French connection cases coinciding with the end of the U.S. heroin epidemic (DuPont and Greene 1973), the 1989/1990 U.S. cocaine market disruption marking the end of the 1980s expansion in cocaine-related problems (Crane, Rivolo, and Comfort 1997), the Australian heroin drought of 2001, and—though less documented—the Taliban opium ban and marked 2003 disruption in U.K. heroin purity roughly coinciding with the end of a long period of increasing heroin dependence in the United Kingdom (Reuter and Stevens 2007). Absence of a counterfactual makes it very difficult to know whether these market disruptions were actually what broke the momentum of some epidemic-type positive feedback loop, but the coincidence is intriguing.

Second, an early observation in drug policy was that increasing nondollar costs of drug use was generally more appealing than increasing the dollar price (Moore 1977; Kleiman 1988). Both discourage use, but only higher dollar prices increase market revenue per gram consumed. Hence, a consumption reduction created by increasing so-called search

\(^{19}\) Dave (2008) estimated a long-run elasticity of arrestees testing positive for heroin that was almost twice the corresponding short-run elasticity. Becker, Grossman, and Murphy (1994) and Saffer and Chaloupka (1999) obtained similar results for cigarettes and for cocaine, respectively. Although perhaps less relevant to the modern era, van Ours (1995) used 1922–38 records from the Dutch East Indies Opium Regie to estimate the short-term and long-run elasticity for opium to be \(-0.70\) and close to \(-1.0\), respectively. Liu et al. (1999) similarly found that opium use in Taiwan between 1895 and 1945 had short- and long-run elasticities of \(-0.48\) and 1.38, respectively.
time costs is more valuable than an equally large reduction in consumption created by driving up price.

We neglected this possibility in the comparative statics discussion above because even though regular users do spend considerable time locating a seller (Rocheleau and Boyum 1994), that may be due primarily to the structural consequences of product illegality, not something that scales linearly in the toughness of enforcement. Two observations supporting that speculation are that heavy users (who account for the majority of demand for most drugs) report knowing multiple alternative suppliers (e.g., Riley 1997), and that a mature market’s distribution network tends to have multiple redundant linkages. (See Caulkins [1998] for some illustrative calculations.)

However, reports of market disruptions ranging from Operation Intercept to the Australian heroin drought describe users having trouble locating supplies (Gooberman 1974; Degenhardt, Conroy et al. 2005; Degenhardt, Reuter et al. 2005). Hence, there may be greater potential for disequilibrium shocks to increase search times than for routine enforcement pressure to drive up search times on an ongoing basis.

D. Cost Effectiveness of Disrupting Drug Markets

The discussion above underscores how little is understood about how to quantify the benefits of disrupting drug markets. Unfortunately, the situation is equally bleak with respect to the other half of a cost-effectiveness ratio. There are no serious estimates of the cost of producing market disruptions.

Some law enforcement operations have fairly predictable production functions. For example, suppose police investigating retail selling operations in a certain city produce an average of 20 arrests per officer per year. In that case, one can reasonably expect that assigning another officer to that activity will increase arrests by about 20 per year. The actual number may be higher or lower since the marginal officer’s productivity may differ from the average, and it could increase or decrease somewhat from year to year as police develop new tactics and the market adapts. However, in some sense taxpayers can just “order” a certain number of additional low-level drug arrests by increasing police budgets by a certain amount.

Market disruptions, however, cannot just be ordered. It seems plausible that the greater the level of enforcement effort, the greater the likelihood that such a disruption will occur. However, historically most
TABLE 3

The Price of Cocaine at Some Market Levels between Farm Gate and Retail

<table>
<thead>
<tr>
<th>Stage</th>
<th>Raw Price ($)</th>
<th>Purity (%)</th>
<th>100% Pure ($)</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Farm gate</td>
<td>800</td>
<td>100</td>
<td>800</td>
<td>Colombia</td>
</tr>
<tr>
<td>Export</td>
<td>2,200</td>
<td>91</td>
<td>2,400</td>
<td>Colombia</td>
</tr>
<tr>
<td>Import/wholesale (kg.)</td>
<td>14,500</td>
<td>76</td>
<td>19,000</td>
<td>Los Angeles</td>
</tr>
<tr>
<td>Mid-level/wholesale (oz.)</td>
<td>19,500</td>
<td>73</td>
<td>27,000</td>
<td>Los Angeles</td>
</tr>
<tr>
<td>Typical retail price</td>
<td>78,000</td>
<td>64</td>
<td>122,000</td>
<td>U.S.</td>
</tr>
</tbody>
</table>


Market disruptions have been the result of a convergence of fortuitous circumstances, not all of which are under policy makers’ control. So it is not useful to think of market disruptions created per million dollars. That inherently limits the ability to bring the drug control tactic of creating temporary market disruptions within a cost-effectiveness framework or to compare it in such terms to other drug control strategies.

VI. Empirical Evidence

Above we mentioned specific data to illustrate the various models of how enforcement affects prices, but in this section we present some broader evidence supporting the view that drug markets can be understood with these economic models.

A. Markups across Levels of the Distribution System

We start by noting how cocaine prices increase along the distribution chain from coca growing to retail price, as presented in table 3. An essentially identical chain applies to heroin and to other Western countries. What is relevant for our purposes is that most of the markup occurs at the bottom of the distribution chain. Though the large fortunes are made by controlling the higher levels of the trade (smuggling, wholesaling), these levels account for a small share of the retail price. Risks and prices theory accounts for that nicely, as suggested by the following hypothetical calculation.

At the bottom of the system, very large numbers of retailers and low-level wholesalers make modest incomes (Reuter, MacCoun, and
Murphy 1990; Levitt and Venkatesh 2000). Smugglers and high-level wholesalers (selling ounces of heroin or kilos of cocaine) have much higher incomes. For purposes of calculation, suppose that smugglers demand as compensation for the risks of enforcement 100 times as much income per year ($5 million per year) as does a typical retail seller (only $50,000 per year). That demand might be rooted in some combination of greater objective risk of receiving a long sentence if caught and a higher value of freedom, for example, because smugglers might be forgoing better alternatives in the legitimate labor market.

Even so, if the smuggler typically handles 500 kilograms of cocaine each year (five shipments of 100 kilograms each) and the retailer sells a total of 0.5 kilograms (2 grams each day for 250 days of the year), then the smuggler spreads his or her risk over 1,000 times as many kilograms of cocaine. Even with a greater valuation of freedom and longer expected sentence, the smuggler’s risk compensation cost per gram is one-tenth that of the retailer.20 As the observed markups show, and as R&P predicts, the risk compensation costs per gram are much higher at the lowest level of the trade than at the higher levels. While these specific figures are obviously entirely speculative, their general relationship is not. Absolute markups are highest at the bottom of the distribution system because risks are distributed over such small quantities.

B. Differences across Countries

There are persistent differences in the purity-adjusted price across countries. We focus here on the United States versus the United Kingdom, another country with price and purity data over a substantial period of time. Within those markets, we compare U.S. cocaine with U.K. heroin since those are the dominant drugs in the respective countries.

Over very broad ranges of market levels, price per pure gram can be well modeled as being proportional to pure quantity raised to an exponent (Caulkins and Padman 1993; Caulkins 1994; Crane, Rivolo, and Comfort 1997). An exponent of 1.0 would indicate that the transaction value is simply proportional to quantity, so the price per pure

---

20 It would be slightly more realistic to speak of a smuggling organization, not a single individual smuggler, but the point carries through. If the organization collectively demands income of $5 million per year and spreads it over 500 kilograms, the risk compensation per gram is the same as if an individual smuggler did so.
TABLE 4
Price Markup Coefficients in the United States and United Kingdom

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Retail Mid-Low</td>
<td>Mid-Level Top-Level</td>
</tr>
<tr>
<td>Cannabis</td>
<td>.817</td>
<td>.573</td>
<td>.820</td>
</tr>
<tr>
<td>Sinsemilla</td>
<td>.850</td>
<td>.716</td>
<td>.787</td>
</tr>
<tr>
<td>Cannabis resin</td>
<td></td>
<td>.751</td>
<td>.813</td>
</tr>
<tr>
<td>Hashish</td>
<td>.850</td>
<td>.661</td>
<td>.833</td>
</tr>
<tr>
<td>Cocaine</td>
<td>.869</td>
<td>.731</td>
<td>.787</td>
</tr>
<tr>
<td>Crack</td>
<td>.827</td>
<td>.661</td>
<td>.833</td>
</tr>
<tr>
<td>Heroin</td>
<td>.834</td>
<td>.531</td>
<td>.718</td>
</tr>
<tr>
<td></td>
<td>.83 brown</td>
<td>.764</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.84 white</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source.—Caulkins et al. (2009).

gram is the same across market levels. In reality, the exponents are less than 1.0. The smaller the exponent, the greater the markup as drugs move down the distribution chain. Inasmuch as the United States has particularly tough domestic enforcement, one would expect price regressions to produce smaller exponents in the United States than abroad. Table 4, reproduced from Caulkins, Burnett, and Leslie (2009), shows that this tends to be the case. Predictably, Moore et al.’s (2005) heroin prices across market levels in Australia imply an exponent (0.82) closer to that of the United Kingdom than to the United States.

C. Changes over Time

Drug price trends over time also support the idea that the main driver of retail price is events in the final market country. When examining heroin and cocaine price trends in the United States (e.g., Fries et al. 2008) and in Europe (e.g., Farrell, Mansur, and Tullis 1996), the striking observation is that cocaine prices in one location are more strongly correlated over time with heroin prices in that same location than they are with cocaine prices in the other location. Similarly, heroin prices in Europe are more highly correlated with cocaine prices in Europe than with heroin prices in the United States.

That is hard to reconcile with models of price determination that
NOTE.—Smaller values indicate greater markups as the drugs move down the distribution chain.
focus on global production or other events in source countries. It is much easier to reconcile with the idea that retail prices are determined primarily by risks and other factors in the final market country. Indeed, it is even suggestive that there is some balancing of supply and demand across drugs within a domestic country. That is a plausible topic for further research given that most demand stems from polydrug users and most domestic drug enforcement policy shifts lump hard drugs together (symbolized by the fact that in the United States the arrest data report cocaine and heroin arrests in aggregate, not separately).

There remains the question of general trends. Although there are no quotable standard data sources, the discussion in Section I makes clear that at least in the mid-1980s the intensity of enforcement, as measured by the expected prison time per unit of cocaine sold, has apparently increased substantially in the United States. During that period the purity-adjusted price of cocaine has fallen, substantially in the 1980s and more gradually since. The slow ebbing since 1990 might be attributed to a decline in demand and declining risks of violence, but not the sharp decline in the 1980s.21 The conflicting price and enforcement trends of the 1980s would be a puzzle if the markets had been in long-run equilibrium.

However, the 1980s were clearly a time of quasi-static, not static equilibrium. The cocaine epidemic was still unfolding, with drug consumption expanding markedly as the mix of users shifted to include more dependent users. Distribution routes and tactics were evolving; Miami was a principal port of entry in 1980, but by 1990 had been supplanted by the Southwest border. Likewise, retail distribution evolved from relatively affluent people supplying friends in nightclubs and other discreet settings to flagrant street corner markets “staffed” by dealers for whom dealing was their principal form of employment. In short, we were still in the process of tipping from the low cocaine equilibrium to the high cocaine equilibrium. Furthermore, since enforcement swamping pertains to total drug volume (which is dominated by cocaine in the United States), not just or primarily drug-specific market volume, there is a simple explanation for why the explosion in cocaine use could have eroded price markups for heroin. Two pieces of empirical support for this are that by the 1990s the risks and prices

21 Demand refers to the quantity that users desire to purchase at a given price; it is a relationship between price and quantity. Consumption refers to actual quantity purchased at the observed price.
story roughly added up (Caulkins and Reuter 1998), and Kuziemko and Levitt (2004) estimate empirically that in the absence of expanding toughness, U.S. cocaine prices would have fallen even farther.

VII. Conclusion

The simple sound bite summary of the foregoing is this: prohibition plus a base level of enforcement can drive prices up far above the legalized price, but for most established markets, expanding enforcement beyond that base level is a very expensive way to purchase further increments in price. Overall the United States is far into the region of diminishing returns; toughness could be cut with modest effects on prices and use. Alternately, toughness could be focused on the forms of dealing that are most violent or otherwise noxious, rather than feeling compelled to treat all dealers operating at a particular volume or market level as equivalent. We did note that the dynamics of drug markets might offer specific opportunities, a matter we return to at the end of this section.

We start our elaboration of the sound bite by drawing in two topics that have only been at the margin of this analysis—enforcement against marijuana and against users. Marijuana presents a distinctive policy problem (Room et al. 2010). Contrary to popular belief, marijuana toughness is already low. In 1997 there were only about 27,000 offenders in state prison for marijuana offenses (Sevigny and Caulkins 2004), while the number of marijuana sellers is in the millions. Indeed, on the basis of self-reports to the National Survey on Drug Use and Health, it is estimated that 1.1 million individuals sell some of their most recent marijuana acquisition (Caulkins and Pacula 2006), and estimates based on self-report to a government-run survey are almost certainly underestimates. Hence, marijuana distributors can certainly expect to spend at least 40 years (1.1 million / 27,000) selling per year of incarceration, and the actual figure is probably close to 100. By comparison, Reuter, MacCoun, and Murphy (1990) estimated that the street sellers of cocaine and heroin spent about one-fifth of their time incarcerated. The upshot is that even if easing up on toughness

\[\text{Thursday Aug 26 2010 04:52 PM CJ v39 390103 MEL}\]
against cocaine, heroin, and methamphetamine is a plausible strategy for substantially reducing incarceration rates in the United States (Caulkins and Reuter 2006), there are no such easy savings from reducing marijuana toughness.

Likewise, the imprisonment risk just for users of the “expensive drugs” (cocaine, heroin, and methamphetamine) who are not also even peripherally involved in distribution accounts for 10 percent or less of drug-related imprisonment, resulting in a risk of only about one and a half days in prison per year of use. For marijuana it is about one hour in prison per year of use, even if all incarceration for polydrug use involving marijuana is attributed entirely to marijuana, and only 30 minutes of it is attributed to the other drug (Caulkins and Sevigny 2005). Though many of the seven hundred thousand individuals who are arrested each year for simple marijuana possession were only users, the average time served per arrest is already less than 2 days. So for users, it might be possible to reduce arrests greatly, but there is little scope for reducing imprisonment dramatically by cutting prison terms for users who were not also involved in distribution.

The policy implications for enforcement against the domestic suppliers of cocaine, heroin, and methamphetamine are as follows: markets can stably remain at either a low- or high-volume equilibrium. If the market is negligible (e.g., with GHB or methamphetamine in New England), use enforcement to keep it there. That is feasible and in some instances can prevent very serious harms. The consequences of a crackdown are unlikely to be harsh at the population level since few individuals are yet involved in distribution or production.

If the market is past its tipping point and is currently in epidemic expansion toward a higher-level steady state, use enforcement to delay that expansion; eradicating the market may not be realistic, but it is worth fighting a holding action rather than giving up on enforcement entirely. This policy of “grudging toleration” may be relevant to cocaine in parts of Europe today.

An important implication of the discussion surrounding these markets that are in quasi-static equilibrium is that the endogenous market forces may overwhelm enforcement efforts, even if those enforcement efforts are serving a valuable role. Hence, even if prices fall and use expands, that is not prima facie evidence that enforcement was mismanaged or counterproductive.

Established markets force a discrete choice. One might try to effec-
tively “eradicate” the market by tipping it back to low levels of use, but that takes enormous effort and patience; probably it is not feasible for most established drugs, at least in countries with strong traditions of freedom.

Practically speaking, one must accommodate most established markets. The marginal value of enforcement intensity beyond those basic levels is probably very low. That means doing enough enforcement to keep everyone’s heads down, both to realize the structural consequences of product illegality and because brazen selling is more harmful to society per kilogram sold than is surreptitious selling. Ongoing toughness may also increase the likelihood that one stumbles into a significant market disruption, although that takes considerable luck. Inasmuch as the United States has expanded enforcement intensity far beyond the levels needed to achieve the structural consequences of product illegality, it has the ability to reduce enforcement intensity with only modest adverse consequences on use and dependence.

Concretely, one way to cut back cocaine and probably heroin toughness without large adverse effects would be to reduce the number of drug prisoners to 250,000 rather than 500,000 (Caulkins and Reuter 2006). This would hardly be going soft on drugs. It would still be a lot tougher than the Reagan Administration ever was. It would ensure that the United States still maintained a comfortable lead over any other Western nation in its toughness toward drug dealers, to put it in the most cynical terms. Furthermore, not all incarcerated drug-law violators are equal. The minority who are very violent or unusually dangerous in other ways may be getting appropriate sentences, and with less pressure on prison space, they might serve more of their sentences. Deemphasizing sheer quantity of drug incarceration could usefully be complemented by greater efforts to target that incarceration more effectively.

There is no magic formula suggesting a halving of drug incarceration as opposed to cutting it by one-third or two-thirds. The point is simply that dramatic reductions in incarceration are possible without entering uncharted waters of permissiveness, and the expansion to today’s unprecedented levels of incarceration seems to have made little contribution to the reduction in U.S. drug problems.

Drug treatment as an alternative to incarceration has become a standard response, but it is not really an alternative. It is a separate policy that stands or falls on its own merits. We can compare the cost effec-
tiveness of locking up drug sellers to the cost effectiveness of treating drug users, but no agency or decision maker divides a constrained budget between these two policy choices. Furthermore, many of the most promising approaches to dealing with drug abusing offenders rely on frequent drug testing backed with the threat of enforcement, and Hawken (2010) notes how such coerced or mandated abstinence programs (Kleiman 2009) may usefully prioritize scarce treatment slots via what she calls “behavioral triage.” Likewise, enforcement has a key role to play in managing drug market harms (Caulkins and Reuter 2009). So backing away from long sentences for dealers in no way implies relegating enforcement to a minor role in drug policy.

A democracy should be reluctant to deprive its citizens of liberty, a reluctance reinforced by the facts that imprisonment falls disproportionately on poor minority communities and that many U.S. prisons are nasty and brutalizing institutions. Further, there is growing evidence that the high incarceration rates have serious consequences for communities. A recent study suggests that differences in black and white incarceration rates may explain most of the sevenfold higher rate of HIV among black males as compared to white males. If locking up typical dealers for 2 years rather than one has minimal effect on the availability and use of dangerous drugs, then a freedom-loving society should be reluctant to do it.

Yet we are left with an enforcement system that runs on automatic, locking up increasing numbers on a faded rationale despite the high economic and social costs of incarceration and its apparently quite modest effects on drug use. Truly “solving” the nation’s drug problem, with its multiple causes, is beyond the reach of any existing intervention or strategy. But that should not prevent decision makers from realizing that money can be saved and justice improved by simply cutting in half the number of people locked up for drug offenses.

Drug courts are well known (Belenko 2001). Hawaii’s Opportunity Probation with Enforcement (HOPE) and South Dakota’s 24/7 Sobriety programs are in some sense similar but empower other criminal justice personnel to do the supervision that judges do in drug courts, which greatly enhances the potential to scale these program up (Kleiman and Hawken 2004; Caulkins and DuPont 2010).
REFERENCES


How Drug Enforcement Affects Drug Prices


Degenhardt, Louisa, Elizabeth Conroy, Stuart Gilmour, and Linette Collins.


Grass, Dieter, Jonathan P. Caulkins, Gustav Feichtinger, Gernot Tragler, and
How Drug Enforcement Affects Drug Prices PROOF 55


Thursday Aug 26 2010 04:52 PM   CJ v39   390103   MEL


How Drug Enforcement Affects Drug Prices


Thursday Aug 26 2010 04:52 PM    CJ v39    390103    MEL
———. 2009. Data tables published on line, table 8.40B. http://www.oas.samhsa.gov/NSDUH/2K8NSDUH/tabs/Sect8peTabs1to43.htm#Tab.