Prohibition vs. Peace

Juan Camilo Castillo[*] and Dorothy Kronick[†]

July 2017[‡]

Abstract

Why is prohibition so seldom enforced? We propose that governments face a tradeoff between prohibition enforcement and peace. Using a model in which policy affects dynamic interaction among those trafficking an illegal good, we show that enforcing prohibition can increase profits—and that higher profits drive violence among traffickers, who invest more in fighting over a more valuable prize. Jailing or killing traffickers, while politically popular, makes them short-sighted and splinters criminal organizations, both of which increase violence. While previous models of illegal markets focus either on supply reduction or on violence, we consider both, revealing why prohibition is self-limiting: the government achieves one enforcement goal only at the cost of another.

[*]Economics Department, Stanford University, jccast@stanford.edu

[†]Political Science Department, University of Pennsylvania, kronick@sas.upenn.edu

[‡]We are especially grateful to Daniel Mejía for his guidance when conceiving this project. We also thank David Bardey, Tulia Falleti, Leopoldo Fergusson, Guy Grossman, Ben Lessing, Ana María Ibáñez, Carlo Prato, Jake Shapiro, Dawn Teele, Hernán Vallejo, Alex Weisiger, and Andrés Zambrano for many helpful comments and suggestions.

Consider the common fate of three attempts at prohibition. In Mexico, a president who began his term with zealous anti-cocaine policies left office publicly questioning "the limits of the current prohibitionist approach;" Mexicans then elected a successor who had campaigned on retreating from the drug war (Calderón, 2012; Fabián, 2012; Lessing, 2015). In Brazil, politicians who passed a strict ban on mahogany exports then looked the other way when exports in "other tropical species" suddenly surged (Chimeli and Boyd, 2010). And in the United States, state governments that had rushed to ratify alcohol prohibition rapidly shrank from enforcement, repealing state enforcement legislation and abandoning seizures (García-Jimeno, 2016). A judge in Oklahoma claimed that "a candidate for sheriff would not possibly be elected if it were known that he intended to enforce Prohibition."[1]

Traditional literature on the political economy of enforcement would explain these reversals by pointing to weak state capacity: in this view, whatever the benefits of enforcing prohibition, the state simply lacks the fiscal or bureaucratic resources to do so (for a review, see Levitsky and Murillo, 2009). Accurate as this account might be for many countries, few would consider the United States a weak state. Another line of work would argue that prohibition is simply *ineffective* in the specific sense of Becker (1968): supply is unresponsive to changes in enforcement. But while ineffectiveness may have plagued certain prohibition policies—such as the aerial spraying of coca crops—other policies, such as interdiction, are known to reduce supply (Mejía and Restrepo, 2011, 2016).

How, then, to understand why *de jure* prohibition is so seldom enforced *de facto*? In this paper, we propose that governments face a tradeoff between enforcing prohibition, on the one hand, and low violence, on the other. While previous theories focus either on supply reduction or on vi-

[1] Franklin (1971), cited in García-Jimeno (2016).

olence reduction, we consider both, illuminating how supply reduction can increase violence. Efforts to shrink illegal markets are therefore self-limiting: the government achieves one enforcement goal only at the cost of another. Non-enforcement of prohibition can thus be viewed as a strategic choice, consistent with Acemoglu et al. (2013) and Holland (2016), who argue that non-enforcement of laws can serve electoral ends.

To arrive at this result, we specify a model in which government policy affects dynamic interaction among traffickers. In the absence of government intervention, traffickers might coexist in peace: repeated interaction creates incentives for dividing the business via negotiation rather than combat. But prohibition enforcement changes those incentives. Seizures of the prohibited good, we find, *can* successfully stem the flow of that good to consumers, but in doing so they increase prices; then, if consumer demand is sufficiently inelastic, profits increase, leading traffickers to fight over an increasingly valuable prize.[2]

Attacking organized crime bosses, while politically popular, further exacerbates violence. The risk of being captured or killed makes capos short-sighted, and leaderless organizations break up into smaller factions—both of which make peace harder to sustain. And because (as we show) supply does not depend on which traffickers control distribution, or on the number of active organizations, jailing El Chapo or Al Capone increases violence without even shrinking supply in consumer markets.[3]

We use qualitative evidence and a review of quantitative empirical studies to illustrate how these results result help explain prohibition enforcement, focusing especially on the cases of cocaine prohibition in the Amer-

[2]Becker et al. (2006) derive an elasticity threshold below which interdiction increases *revenues* in illegal markets; we derive a different threshold below which interdiction increases *profits*, since profits—not revenues—determine the incentives for violence. (Becker et al. (2006) do not explicitly consider violence as a policy outcome.)

[3]This result arises from the assumption of constant returns to scale in smuggling, an assumption we discuss in detail in Section 2. It does not depend on constant returns to scale in other activities (such as conflict).

icas and alcohol prohibition in the United States. The model predicts that reducing the supply of an illegal good increases violence when demand is sufficiently inelastic.[4] And indeed, when Colombian and U.S. authorities switched from the ineffective policy of coca crop eradication to the more efficient strategy of interdiction (drug seizures), lethal violence increased as a result (Castillo et al., 2016). Moreover, consistent with our prediction that targeting capos provokes violence, Calderón et al. (2015) establish that Mexico's aggressive kingpin strategy increased homicide rates. The violence caused by counternarcotics operations then fueled anti-drug-war sentiment—so much so that leaders from across Latin America questioned prohibition enforcement at a regional summit later dubbed the "Rebellion at Cartagena" (Guillermoprieto, 2012). We also discuss the case of the retreat from enforcing alcohol prohibition in the United States, which, as García-Jimeno (2016) documents, followed a similar political logic.

This paper is most closely related to formal work on illegal markets. One strand of this literature models trafficking organizations as profit-maximizing firms (Poret, 2003; Poret and Téjédo, 2006; Burrus, 1999; Baccara and Bar-Isaac, 2008).[5] While providing critical infrastructure for understanding trafficker behavior, these papers evaluate policy only in terms of supply reduction and trafficker revenues—without explicitly considering violence. Other work, most notably Lessing (2015), provides critical infrastructure for understanding the policy determinants of violence in drug markets—without explicitly considering traffickers' productive operations (that is, the purchase, transport, and sale of illicit goods).[6] We

[4]Estimates in Bachman et al. (1990), DiNardo (1993), and Saffer and Chaloupka (1999) suggest that the price elasticity of demand for cocaine is indeed below the threshold we derive; we discuss these estimates in Section 3.3.

[5]Poret (2003), Poret and Téjédo (2006), and Burrus (1999) model the horizontal and vertical structure of drug markets. Baccara and Bar-Isaac (2008) study the efficiency of organizational structures of criminal groups.

[6]Lessing (2015) and Lessing (2012) also focus primarily on cartel-state violence, rather than violence among cartels.

bridge these literatures by considering traffickers' profit motivations and violent behavior together, formalizing how government operations affect traffickers' *dynamic* incentives to divide the business peacefully. By doing so, and together with case evidence, the model clarifies why governments so often retreat from enforcing prohibition.

1 The tradeoff between prohibition enforcement and peace

The model we present in Section 2 formalizes our explanation for weak enforcement of prohibition: that policymakers face a tradeoff between prohibition enforcement and peace. Here we preview the results informally and explain how they relate to previous literature.

We model criminal trafficking organizations as profit-maximizing firms who compete for market share; they can divide the business via all-out war, or they can collude to reduce violence. Some of the determinants of collusion in our model mirror those of other models in international relations and industrial organization.[7] Both in our model and in the literature, higher discount factors and a smaller number of actors make it easier for treaties to hold. Both in our model and in the literature, information about rivals facilitates agreements: traffickers must understand their rivals' incentives in order to design incentive-compatible punishments. Both in our model and in the literature, theory predicts that actors always comply with a collusive agreement—meaning that deviation is off the equilibrium path, so that the punishment stage is never reached. Here as elsewhere, however, this unrealistic prediction can be reconciled with reality by allowing for imperfect information, which easily leads to violations of the agreement and

[7] These are vast literatures and we do not attempt to summarize them; see, e.g., Fearon (1995); Powell (2002); Fey and Ramsay (2007).

thereby war among traffickers (or price wars among firms, or international war among states); explicitly modeling the role of information and communication among traffickers could make for a valuable extension to the model.

While thus similar in many ways, the model considered here also departs from previous work. In industrial organization settings, firms usually cannot observe *which* rival deviated from a collusive agreement, since firms only observe prices. Among rival traffickers, however, traitors are easily identified, since deviating from a collusive agreement consists of ramping up conflict expenditure. This ability to identify defectors—together with the possibility of traffickers directly attacking a lone defector—enables *optimal punishment* that is infeasible among competing industrial firms, where "punishment" must take the form of price changes that harm all firms equally (whether they betrayed or not) (c.f. Abreu, 1983). Also, there is a normative difference between our analysis and related work in industrial organization: Collusive agreements among firms hurt society (relative to perfect competition), whereas collusive agreements among traffickers benefit society by lowering violence. Whereas policymakers in antitrust cases help the public by preventing collusion, prohibition enforcement often hurts the public by preventing collusion, as the examples in Section 5 illustrate.

The most important difference between our model and the existing literature, however, lies in the way in which a third party—in our case, a government enforcing prohibition—affects the level of violent conflict among rivals (in our case, traffickers).[8] We make two contributions in this vein: first, to understanding prohibition enforcement and conflict in a static set-

[8]Part of the international relations literature considers how third-party enforcement affects bargaining among states (e.g., Kydd, 2003). However, the role of the third party in this work is largely facilitating information revelation, rather than affecting the value of the prize over which rivals bargain, or their dynamic incentives to collude. See also Bueno de Mesquita (2005) on the effect of government crackdowns on terrorism.

ting; second, in extending the analysis to a dynamic setting.

In the static setting, previous work has noted that counternarcotics enforcement can reduce the supply of drugs, which, if demand is sufficiently inelastic, increases total revenue (Becker et al., 2006). In order to link this insight to violence, we derive conditions under which interdiction increases total *profits* (rather than total revenue); since traffickers fight over profits, not revenues, profits are the relevant outcome for studying violence.

We then extend the analysis to a dynamic setting, finding that supply reduction also increases violence when traffickers can use repeated interaction to divide the business less violently than in the one-period setting. In the dynamic context, however, prohibition enforcement does not fuel violence simply by increasing profits (and thereby the stakes of conflict); indeed, in a dynamic setting, higher profits affect the benefits *both* of complying with a collusive, lower-violence agreement *and* the benefits of deviating from that agreement. Instead, interdiction increases violence in a dynamic setting because it increases the benefit of deviating *relative* to the benefit of complying (Section 4.2.5).[9]

Analyzing the effect of prohibition enforcement on profits and conflict in a dynamic setting also produces a counterintuitive result: while rising profits affect the *scale* of violence when traffickers cannot sustain a peaceful equilibrium, rising profits do not make it harder to sustain a peaceful equilibrium in the first place. In a violent equilibrium, rising profits increase the benefit of deviating from a collusive, lower-violence agreement more than they increase the benefit of complying—but in a peaceful equilibrium, rising profits affect complying and deviating equally. As we show in Section 5, empirical findings are consistent with this result.

The fact that profits only affect violence when traffickers cannot sustain peace underpins the danger of jailing or killing high-profile traffickers.

[9] This result (on relative benefits of deviating and complying) hinges on the assumption that investment in the conflict has decreasing marginal returns.

The risk of being captured or killed makes capos shortsighted and splinters criminal organizations—both of which make peace harder to sustain. But once a peaceful equilibrium is infeasible, the scale of violence begins to respond to profits. Targeting kingpins and ramping up interdiction—which often go hand-in-hand as part of a crackdown—therefore constitute a fatal combination: taking out capos breaks a peaceful equilibrium, and then rising profits fuel violence.

Our model also relates to work that considers how to design policy so as to reduce violence in illegal markets. As Kleiman (2011) argues and Lessing (2015) formalizes, governments could minimize violence among traffickers by "concentrat[ing] enforcement efforts against dealers selected for their contribution to violence" (Kleiman, 2011, p. 96). As Kleiman acknowledges, however, this would require governments abandoning the notion of "a world free of drugs" (p. 101)—in other words, it would require governments to choose peace over enforcing prohibition. Our model clarifies this tradeoff.

Certainly, our model does not seek to explain all or even most of the observed variation in prices, quantities, and violence in illegal markets. Many forces shaping these markets—such as demand shocks driven by changing drug fashions—lie outside our scope. Likewise, enforcement decisions depend not only on the factors considered here—that is, the government's desire to reduce supply, and the resulting violence—but also on political allegiances; perhaps the clearest example is U.S. tolerance for the drug trafficking of Manuel Noriega in Panama (see, e.g., Kornbluh, 1991). Rather than comprehensively model all of the political and economic forces affecting illegal markets, our model seeks to illuminate one explanation for the difficulty of sustaining prohibition enforcement. The following sections present the model, and Section 5 relates the model to case evidence from a number of illegal markets, focusing on the examples of cocaine prohibition in the Americas and alcohol prohibition in the United States.

2 The trafficking industry

To better understand the self-limiting nature of prohibition, we write down a dynamic game involving trafficking organizations and the government. While our model in principle applies to any illegal market, for concreteness we frame the discussion in terms of the international cocaine market. Drug cartels seek to maximize profit, which requires two types of activities: *productive* activities, in which they buy drugs from producers and then move those drugs to consumer markets, evading government enforcers along the way; and *military* activities, in which they fight other cartels for the routes used to transport drugs.[10,11] We characterize each of these activities in turn.

2.1 Buying, shipping, and selling illegal goods

The core components of cartels' productive behavior can be described in one sentence: Cartel $i \in I$ buys a quantity x_i of illicit drugs in producer markets at a price p_p, transports those drugs through R_i routes it controls, and sells them in consumer markets at a price p_c. We assume that a fixed number n of cartels participate in drug trafficking; n is fixed because incumbents' military power constitutes a barrier to entry: incumbent cartels will wipe out any group that tries to enter.

[10] In reality, as Lessing (2015) and others have described in detail, cartels engage in many types of activities other than those described here: For example, they may directly lobby governments. We exclude these political activities from the model not because they are uninteresting, but because including them would complicate the model without illuminating our core results.

[11] We use the term *cartel* not because drug trafficking organizations collude to increase prices, but rather because early Colombian traffickers described themselves as *carteles* and the term thereafter became widely used and anglicized; it is thus conventional, and also less cumbersome than "drug trafficking organizations."

The government seizes a fraction of each cartel's drug shipments; the size of the fraction depends on government expenditures on interdiction, which we denote e. By *interdiction*, we mean any operation targeting the productive behavior of cartels: for example, seizing drugs in transit, patrolling routes, or seizing boats or airplanes used to move drug shipments. *Interdiction* therefore excludes counternarcotics activities such as capturing or killing cartel leaders.

The amount the cartel sells in consumer markets, q_i, is an increasing function of the amount it purchases (x_i) and of the routes it controls (R_i), and a decreasing function of the government's interdiction expenditures (e): $q_i = q(x_i, R_i, e)$.[12] We make the simplifying assumption that all cartels are equally efficient, so that this function holds for every cartel. Considering the consequences of asymmetry could make for an interesting extension to the model.[13]

We assume that the production function $q(x_i, R_i, e)$ has constant returns to scale in routes R_i and quantity purchased x_i. A better way to understand this assumption is to define the *survival rate* $w_i = w(x_i, R_i, e)$, which denotes the fraction of cartel i's drugs that reach consumer markets (so that $q(x_i, R_i, e) = w(x_i, R_i, e)x_i$). We assume that this survival rate only depends on the ratio of routes to the quantity of drugs purchased, $\frac{R_i}{x_i}$, which is a measure of route saturation. In other words, increasing route ownership R_i and quantity x_i in the same proportion does not change the fraction of drugs that make it to consumer markets.[14]

[12] We make two standard assumptions about the production function $q(x_i, R_i, e)$: (1) That it is twice-differentiable, increasing in both factors of production, and decreasing in enforcement ($\partial q/\partial x_i > 0$, $\partial q/\partial R_i > 0$, and $\partial q/\partial e < 0$); and (2) that the marginal productivity of both factors of production is decreasing ($\partial^2 q/\partial x_i^2 < 0$ and $\partial^2 q/\partial R_i^2 < 0$).

[13] Our dynamic model has a multiplicity of equilibria. The analysis is simplified by choosing the lowest symmetric equilibrium. In an asymmetric model we would have to use some other equilibrium selection criterion, and the details of the comparative statics would depend on how the chosen equilibrium changes as the model parameters change.

[14] An implicit assumption is that all routes are equally productive. This may seem at odds with reality, where some routes are better for traffickers than others, but it can be

This assumption does not imply that *all* cartel operations have constant returns to scale. We assume constant returns to scale only in one specific activity: the smuggling of drugs from producers to consumers. In contrast, we do *not* assume constant returns to scale in cartels' attempt to control routes (Section 2.2), a major component of cartel operations.

Shrinking the quantity of drugs reaching consumers is not the only consequence of interdiction. It also reduces the marginal productivity of both routes (R_i) and quantity purchased (x_i). In other words, interdiction decreases the fraction of every additional unit of drugs purchased that reaches consumer markets.[15] We state and prove some additional useful properties of the production function q in Appendix A.1.

Each cartel controls a small share of the total market, so it has no market power and takes both upstream and downstream prices as fixed.[16] All traffickers together, however, account for an important share of the total drug trade, so the total quantity of drugs supplied affects prices.

We denote the elasticity of demand for drugs in the consumer market ϵ_c.[17] We assume that prices in the producer market are fixed, which corresponds to an elasticity $\epsilon_p = \infty$. This greatly simplifies the analysis without sacrificing any important insights: in Appendix D, we relax this assumption, allowing $\epsilon_p \in (0, \infty)$, and the main results hold.

Cartels' productive behavior determines both $X = \sum_{i \in I} x_i$, the aggregate quantity of drugs bought from producers, and $Q = \sum_{i \in I} q_i$, the aggregate supply of drugs to consumers. It is this Q that the government would like to minimize.

---

rationalized by redefining R_i as the quantity of drugs that can be smuggled with some fixed survival rate.

[15] I.e., $\partial^2 q / \partial e \partial x_i < 0$ and $\partial^2 q / \partial e \partial R_i < 0$.

[16] The Herfindahl-Hirschman index for Mexican cartels, for example, is around 0.15, suggesting that this is a reasonable assumption. This calculation is based on the data from Castillo et al. (2016).

[17] This is the elasticity of residual demand, after taking into account the demand satisfied by other sources of drugs not being analyzed.

2.2 Fighting for trafficking routes

In order to move illegal goods from producers to consumers, traffickers must control routes. There is a continuum of drug trafficking routes normalized to one, i.e., $\sum_{i \in I} R_i = 1$. In order to control these routes, cartel i invests g_i in fighting; g_i includes the salaries of gunmen, the cost of guns, and losses associated with dead gunmen, among other costs. At the end of the conflict, the proportion of routes held by each cartel is determined by a contest success function $R_i(g_i, g_{-i})$ that depends both on own conflict expenditure and on the total amount $g_{-i} = \sum_{j \neq i} g_j$ spent by all other cartels. We can think of aggregate spending in the conflict, $G = \sum_{i \in I} g_i$, as a proxy for the level of violence.[18]

Four intuitive and standard assumptions pin down the form of the contest success function R_i that determines each cartel's share of routes:[19]

Lemma 1. *The contest success function R_i has the following form:*

$$R_i(g_i, g_{-i}) = \frac{g_i}{g_i + g_{-i}}. \tag{1}$$

This implies the following properties:

1. *R_i is increasing in own expenditure in the conflict $\left(\frac{\partial R_i}{\partial g_i} > 0\right)$.*

[18]Some papers model investment in the conflict and violence as separate quantities, especially in conflicts between countries (Powell, 1993; Jackson and Morelli, 2009). Countries first decide to invest in arms races, and then they decide whether to attack. We believe that our simplification that investment and violence are the same is more realistic in the context of competing cartels because it is harder to observe others' investment, and arms races would thus be pointless.

[19]These assumptions are: (1) That the function is continuous and homogenous of degree zero; this captures the intuition that if all cartels increase conflict expenditures proportionately, the distribution of routes controlled does not change (for a detailed discussion, see Skaperdas, 1996). (2) That the function is the same for every cartel, meaning that all cartels are a priori equally efficient in their use of military resources. There might be institutional reasons for heterogeneity across cartels—for instance, cartels might have differential ties to corrupt government officials—but studying that heterogeneity is beyond the scope of this paper. (3) That investing nothing in conflict means controlling zero routes unless all cartels have zero expenditure. (4) That every route is controlled by exactly one cartel.

2. *The marginal productivity of conflict expenditure is decreasing and goes to zero in the limit* $\left(\frac{\partial^2 R_i}{\partial g_i^2} < 0,\ \lim_{g_i \to \infty} \frac{\partial R}{\partial g_i} = 0\right)$.

Proof. See Appendix A. □

Since cartel i sells a quantity q_i of drugs in the consumer market at a price p_c, it ultimately obtains profit given by

$$\pi_i = p_c \cdot q(x_i, R_i(g_i, g_{-i}), e) - g_i - p_p \cdot x_i \tag{2}$$

While cartels make two decisions—the quantity of drugs to buy, x_i, and how much to invest in conflict, g_i—only conflict expenditure g_i affects strategic interaction among cartels. Since cartels are price takers, the quantity of drugs sold by cartel i does not affect its rivals.[20,21]

3 Interdiction, the supply of drugs, and profit

Before considering violence outcomes, we analyze the relationship between the government's interdiction expenditure e, the total supply of drugs reaching consumer markets Q, and aggregate productive profit. Three key findings emerge. First, the total quantity of drugs bought from producers and sold to consumers does not depend on *which* cartels control routes; redistributing routes from cartel i to cartel j makes no difference for the aggregate supply of drugs. Second, increasing interdiction expenditure reduces the aggregate supply of drugs reaching consumers. And third, the relationship between interdiction and total profit depends crucially on the

[20] If cartels had some market power, interaction through drug quantities becomes relevant, as in a traditional Cournot oligopoly.

[21] We do not model bargaining explicitly, an element that might be important based on the international relations literature (see, for instance, Wagner, 2010, chapters 4 and 5). We do, however, model the principle elements that allow agreements to be enforceable—the same forces that might allow the outcome of bargaining to be enforeceable.

price elasticity of demand for drugs. Previous work derived an elasticity threshold above which interdiction increases total *revenue* in illegal markets (Becker et al., 2006); here, we derive a threshold above which interdiction increases total *profit* in illegal markets. The thresholds are different because cartels' *costs* change with interdiction, and understanding the effect of interdiction on *profit* (as opposed to revenue) is essential for understanding the effect of interdiction on violence, as we show in the following section.

3.1 Market shares don't affect supply

To see why the aggregate supply of drugs is independent of the distribution of routes across cartels, suppose that in some equilibrium cartel i controls a quantity of routes given by $\hat{R}_i$. Since the quantity of drugs that cartel i purchases from producers does not affect other cartels, cartel i chooses that quantity in order to maximize profit:

$$x_i^* = \underset{x_i}{\operatorname{argmax}} \left[p_c q(x_i, \hat{R}_i, e) - g_i - p_p x_i \right], \tag{3}$$

The optimal quantity x_i^* can be found from the following first-order condition:

$$\underbrace{p_c \frac{\partial q}{\partial x_i}}_{MgBx_i} = \underbrace{p_p}_{MgCx_i} \tag{4}$$

which equates marginal benefit and cost. We assume that the ratio of retail to wholesale prices, $\frac{p_c}{p_p}$, is large enough that there is an interior solution (otherwise, the illegal drug market would not exist).[22] Then it is straightforward to see that:

[22]The solution is bounded since the marginal productivity drops to zero as x_i goes to infinity: $\partial r/\partial x = -R/x^2$, and $\partial q/\partial X = w + x \times (\partial w/\partial x) = w + x \times (\partial w/\partial r)(\partial r/\partial x) = w - r \times (\partial w/\partial r$. And w, r and $\partial w/\partial r$ drop to zero as x_i goes to infinity. Also note that cartels choose x_i and g_i simultaneously, so this is only a necessary condition at equilibrium.

Lemma 2. *Aggregate demand from the producer market and supply to the consumer market are determined by*

$$p_c \frac{\partial q(X, 1, e)}{\partial x} = p_p \qquad Q = q(X, 1, e), \tag{5}$$

which are independent of the distribution of routes.

The share of drugs bought and sold by each cartel is equal to the fraction of routes it controls:

$$x_i = R_i X \qquad q_i = R_i Q \tag{6}$$

Proof. See Appendix A. □

Lemma 2 is a very general result. It is a consequence of the production technology having constant returns to scale in routes R_i and drug purchases x_i, and of the fact that the conflict among cartels is a zero-sum game that does not alter the total number of routes. The result holds regardless of the level of violence, and it has a clear policy implication: government operations that target the balance of power among cartels may affect violence, but these operations do not affect the total supply of drugs reaching consumers.

3.2 Interdiction does affect supply

To understand why *interdiction* does reduce the total supply of drugs reaching consumers, it is useful to first define *aggregate productive profit* as the difference between total drug revenue and the total cost of purchasing drugs from producers, $\pi^A = p_c Q - p_p X$, without taking into account conflict expenditure. (As opposed to *aggregate profit*, which would be $\pi^A - G$.) Defining aggregate productive profit in this way allows us to rewrite each cartel's profit function:

Lemma 3. *Cartel i's profit can be restated as*

$$\pi_i = \pi^A R(g_i, g_{-i}) - g_i \tag{7}$$

Proof. See Appendix A. □

which emphasizes that cartels invest resources in fighting (g_i) in order to control a share of routes $R(g_i, g_{-1})$, thereby obtaining that same share of the aggregate productive profit π^A.

The effect of interdiction on the aggregate quantity of drugs purchased from producers can be found by applying the implicit function theorem to Equation (5):

$$\frac{\partial X}{\partial e} = \left(\frac{-\frac{\partial q}{\partial X}}{\frac{1}{Q\epsilon_c}\left(\frac{\partial q}{\partial X}\right)^2 + \frac{\partial^2 q}{\partial X^2}} \right) \left[\overbrace{\frac{1}{\epsilon_c}\frac{\partial \log q}{\partial e}}^{\text{(a)}} + \overbrace{\frac{\partial \log \frac{\partial q}{\partial X}}{\partial e}}^{\text{(b)}} \right] \tag{8}$$

The first term is positive, so the sign of the effect is determined by the term in square brackets.

Two mechanisms work in opposite directions. Term (a) captures the fact that interdiction decreases the supply of drugs, thereby increasing prices and encouraging cartels to buy more drugs to sell them to consumers. In term (b), interdiction reduces the marginal productivity of drugs purchased, with the opposite effect. Which effect dominates depends on whether demand is inelastic enough that effect (a) is larger.

While the effect of interdiction on drugs purchased (X) is ambiguous, the effect of interdiction on the supply of drugs to consumers (Q) is not. The intuition is that supply of drugs to consumers, Q, can only increase if cartels purchase more drugs from producers as interdiction intensifies. But if cartels purchase more drugs from producers as interdiction intensifies, it must be because prices are increasing—which implies that supply Q is falling.[23] Therefore it is always the case that:

Proposition 1.

- $\frac{\partial Q}{\partial e} < 0$*: Increasing interdiction reduces the supply of drugs.*
- $\frac{\partial Q}{\partial n} = 0$*: The number of cartels has no effect on the supply of drugs.*

3.3 Elasticity, interdiction, and productive profit

While the previous section establishes that interdiction decreases the supply of drugs to consumers, it remains unclear whether interdiction increases or decreases productive profit $\pi^A = p_c Q - p_p X$. This relationship between interdiction and productive profit is critical: As we explain in the rest of this paper, changes in productive profit drive changes in violence, both in a one-period equilibrium and in a dynamic model.

The effect of interdiction on productive profit can be decomposed into its effects on costs and its effects on revenues:

$$\frac{\partial \pi^A}{\partial e} = p_c \left(1 + \frac{1}{\epsilon_c}\right) \frac{\partial Q}{\partial e} - p_p \frac{\partial X}{\partial e} \tag{10}$$

Substituting in $\frac{\partial Q}{\partial e}$ and $\frac{\partial X}{\partial e}$ from Equations (8) and (9) allows us to derive a threshold for the elasticity of demand such that productive profit increases if $\epsilon_c > \hat{\epsilon}_c$: that is, $\hat{\epsilon}_c$ is the elasticity threshold above which interdiction drives up revenue (by increasing prices) faster than it increases costs. The

[23] To see this formally, consider the sign of the individual derivatives in:

$$\frac{\partial Q^e}{\partial e} = \frac{\frac{\partial^2 q}{\partial X^2}\frac{\partial q}{\partial e} - \frac{\partial q}{\partial X}\frac{\partial^2 q}{\partial X \partial e}}{\frac{1}{Q\epsilon_c}\left(\frac{\partial q}{\partial X}\right)^2 + \frac{\partial^2 q}{\partial X^2}}. \tag{9}$$

expression for this elasticity threshold is:

$$\hat{\epsilon}_c = -1 - \underbrace{\frac{\left(\frac{\partial q}{\partial X}\right)^2}{\frac{\partial q}{\partial e}\frac{\partial^2 q}{\partial X^2}}}_{\text{(a)}} \underbrace{\left(\frac{\partial \log q}{\partial e} - \frac{\partial \log \frac{\partial q}{\partial X}}{\partial e}\right)}_{\text{(b)}} \tag{11}$$

Crucially, the threshold $\hat{\epsilon_c}$ differs from the threshold of -1 that determines how aggregate *revenue* changes with supply. Aggregate revenue constitutes a good measure of the harm caused *to consumers* by illegal drugs, which is why Becker et al. (2006) focus on the -1 threshold in their analysis of the welfare consequences of prohibition. While Becker et al. (2006) do not explicitly consider violence, subsequent studies have taken the -1 threshold out of context, concluding that it also determines how cartel *profits*—and thus conflict—change with interdiction.

Proposition 2. *The comparative statics on aggregate productive profit are:*

a) If $\epsilon_c < \hat{\epsilon}_c$, then $\frac{\partial \pi^A}{\partial e} < 0$: If demand is sufficiently elastic, interdiction reduces aggregate productive profit.

b) If $\epsilon_c > \hat{\epsilon}_c$, then $\frac{\partial \pi^A}{\partial e} > 0$: If demand is sufficiently inelastic, interdiction increases aggregate productive profit.

The threshold in Equation 11 is the -1 threshold plus a correction for the extent to which the sensitivity of *profits* to interdiction is higher or lower than the sensitivity of *revenue* to interdiction. Since profits and revenues differ only because of costs, $p_p X$, the correction depends on how interdiction affects wholesale drug purchases X. Here, the intuition is just as in Section 3.1: Interdiction causes a decrease in supply and an increase in prices, in response to which cartels buy more drugs, driving up their costs. On the other hand, interdiction lowers the marginal productivity of drugs, leading cartels to buy fewer drugs—thereby decreasing costs $p_p X$. Which effect dominates determines the sign of the correction to the -1 threshold: if the former effect dominates, interdiction drives costs up, making profits *less* sensitive to interdiction than revenue is sensitive to interdiction; if the latter effect dominates, interdiction drives costs down, making profits *more* sensitive to interdiction.[24]

The sign of the correction is ambiguous without further assumptions about the functional form of q. One obvious assumption is that $q(x, R, e)$ should never exceed one: a cartel cannot sell more drugs than it buys. This rules out many functional forms (such as Cobb-Douglas or CES in (x, R)). Although this assumption, in and of itself, does not pin down the sign of the correction, in Appendix C we show that the correction is negative for a wide variety of functional forms that satisfy the assumption. This strongly suggests that the relevant threshold for policy evaluation is less than -1.[25]

As one specific example, consider a production function that follows a standard contest-success function in which the survival rate is a function of the inverse route saturation $r_i = \frac{R_i}{X_i}$ and interdiction e:

$$w(r, e) = \frac{r}{r + \varphi e} \tag{12}$$

In this case, the threshold is $\hat{\epsilon}_c = -\left(1 + \frac{\sqrt{\gamma}}{2(1-\sqrt{\gamma})}\right)$, where $\gamma = \frac{p_p}{p_c}$. (Appendix C graphs $\hat{\epsilon}_c$ as a function of γ). This expression makes clear that $\hat{\epsilon}_c < -1$, meaning that the effect of interdiction on marginal productivity *always* exceeds the effect on productivity.

The sensitivity of profit to interdiction thus exceeds the sensitivity of

[24]This can be seen in Equation 11: the sign is determined by $\frac{\partial \log q}{\partial e} - \frac{\partial \log \frac{\partial q}{\partial X}}{\partial e}$, which is the difference of the two opposing effects.

[25]More specifically, in Appendix C we argue that the threshold is less than -1 unless the production technology exhibits very uneven behavior as inputs increase—for example, if it initially grew faster in x than a Cobb-Douglas technology and then suddenly became saturated as q approaches x. Even in that case, however, the elasticity threshold would only be greater than -1 for a limited range of x.

revenue to interdiction, meaning that the range of elasticities for which interdiction spurs violence is therefore greater than the range suggested by the -1 threshold. The exact magnitude of the correction depends on the ratio of producer to consumer prices; plugging in the numbers provided in Reuter (2004) (p. 130) yields a corrected threshold of -1.1. Given that many empirical estimates of the price elasticity of cocaine fall between -0.9 and -1 (Gallet, 2014), this correction is significant: without it, we might conclude that enforcement has only negligible effects on conflict; with it, we understand why enforcement can fuel large-scale violence. The correction is also relevant because a number of elasticity estimates lie between -1 and -1.1—below the original threshold, but above ours. Overall, the fact that elasticity estimates tend to cluster close to the original -1 threshold underscores the importance of the correction.

4 Interdiction, beheadings, and violence

The results so far have considered the determinants of the aggregate supply of drugs and of aggregate productive profit, finding that (a) the aggregate supply of drugs to consumers is independent of the distribution of route ownership across cartels; (b) the aggregate supply of drugs to consumers declines with interdiction; and (c) aggregate productive profit increases with interdiction as long as demand is more inelastic than a threshold $\hat{\epsilon}_c$, which differs from the threshold of -1 that determines how total revenue changes with interdiction.

We now turn to the principal outcome of interest in this paper: violent conflict. The equilibrium violence level in the stage game of cartels' conflict over routes provides a benchmark for understanding the equilibria in the game with repeated interaction.

4.1 Stage-game Nash equilibrium (SGNE)

To derive cartels' best response functions, recall that cartel i's problem can be written as:

$$\max_{g_i} \pi_i = \pi^A R(g_i, g_{-i}) - g_i \tag{13}$$

where π^A is aggregate productive profit, $R(g_i, g_{-1})$ is cartel i's share of routes and therefore its share of productive profit, and g_i is cartel i's conflict expenditure. This leads to the following first order condition:

$$\underbrace{\pi^A \frac{\partial R}{\partial g_i}}_{MgBg_i} = \underbrace{1}_{MgCg_i} \tag{14}$$

meaning that cartels equate the marginal benefit of conflict expenditure to its marginal cost (one).[26]

First order-conditions (14), one for each cartel, give the best-response functions for g_i in terms of the quantities g_{-i} chosen by all other cartels. The Nash equilibrium of the stage game occurs when the first-order conditions are satisfied simultaneously for all cartels. The symmetry of the problem means that the unique one-period Nash solution has $g_i = g_j = g^N$ and $x_i = x_j = x^N \quad \forall i, j \in I$, and $g_{-i} = (n-1)g^N$, meaning that every cartel controls an equal share of routes $R_i = \frac{1}{n}$.[27]

Proposition 3. *Under a symmetric stage-game Nash equilibrium, the comparative statics on the level of violence are as follows:*

[26] Note that there cannot be a corner solution with $g_i = 0$ and $x_i > 0$, since the marginal productivity of expenditure in the conflict tends to infinity if all cartels spend zero resources in the conflict. Likewise, there cannot be a solution with g_i unbounded because the marginal productivity of conflict expenditure goes to zero as g_i goes to infinity. Thus, every cartel arrives at an interior solution. We check the second-order conditions in Appendix B.

[27] The solution is unique because the marginal productivity of g_i is strictly decreasing; existence was established above.

a) *If* $\epsilon_c < \hat{\epsilon}_c$*, then* $\dfrac{\partial G^N}{\partial e} < 0$*: If demand is sufficiently elastic, interdiction reduces the level of violence.*

b) *If* $\epsilon_c > \hat{\epsilon}_c$*, then* $\dfrac{\partial G^N}{\partial e} > 0$*: If demand is sufficiently inelastic, interdiction increases the level of violence.*

c) $\dfrac{\partial G^N}{\partial n} > 0$*: An increase in the number of cartels increases violence.*

Proof. See Appendix A. □

In other words, in the stage game, the effect of interdiction on violence mirrors the effect of interdiction on the size of aggregate productive profit π^A: if interdiction increases the prize π^A, cartels invest more in fighting over it; if interdiction shrinks the prize, violence declines. The conflict intensifies as the stakes increase.

4.2 Repeated interaction and collusion

Of course, cartels do not interact in a one-period setting; they interact repeatedly over time, which creates room for non-violent solutions to the conflict over routes—and indeed, cartels often operate in relative peace (Duran-Martinez, 2015; Lessing, 2015). While an extensive literature on the economics of conflict describes how opposing parties can sustain a peaceful equilibrium through repeated interaction, previous work has not considered how the behavior of a third party—that is, an agent outside or above the conflict, in this case the government—affects the likelihood of peace. This is fundamental to understanding the relationship between counternarcotics policy and violence. In this section, we therefore analyze the consequences of third-party behavior for conflict in a dynamic setting.

4.2.1 Beheadings and impatience

First, note that cartel i's total profits are the discounted sum of the profit obtained in each period:

$$\Pi_i = \sum_{t=0}^{\infty} \beta^t \pi_{i,t} \tag{15}$$

where $\pi_{i,t}$ is the profit obtained by cartel i in period t, and $\beta \in (0, 1)$ is the discount factor. The discount factor, in turn, depends both on a monetary discount factor (related to the interest rate), which we call δ, and the probability p that the current leader of the cartel will still be in charge in the next period, such that $\beta = \delta p$. The probability p of a leader staying in power depends on the government: Policies aimed at capturing or killing leaders decrease p, thereby decreasing the value of the future for current cartel leaders (i.e., making leaders more impatient).

4.2.2 The feasibility of optimal punishment

We consider three of the many equilibria that can arise in a repeated game. The baseline equilibrium repeats the stage-game Nash equilibrium perpetually, with profit $\Pi^N = \pi^N/(1 - \beta)$ for each cartel. This is the equilibrium that arises when there is no cooperation between cartels. The comparative statics are exactly as in the SGNE, which means that Proposition 3 holds.

The other two equilibria arise from punishment strategies: if some cartel deviates from a low-violence agreement by increasing its conflict expenditure, all other cartels punish it in the next period. The two punishment strategies we consider are *Nash reversion*, in which cartels return to the SGNE, and *optimal punishment*, in which cartels punish deviators as harshly as possible (Abreu, 1983). While we include *Nash reversion* for its simplicity and usefulness as a point of comparison, we focus on *optimal punishment* because it achieves the greatest possible payoff for cartels. In this context, optimal punishment implies ensuring that any cartel that deviates receives zero profits from that moment on, as in Rotemberg and Saloner (1986), since enforcing negative profits would encourage the deviator to exit from the industry.

Assumption 1. *There exists a subgame perfect mechanism by which zero profits can be imposed on any given cartel through coordination of all other cartels. We call this mechanism* **optimal punishment***.*[28]

In most oligopolies, optimal punishment is infeasible: in part because firms only observe prices, so they know that some firm deviated but do not know *which* firm deviated; and in part because punishment consists of starting a price war that does not target a specific firm, so that compliers are punished along with deviators. But drug cartels—and traffickers in other illegal goods—are different. First, *which* firm deviates from the collusive equilibrium is easy to observe, since deviating consists of violent actions. Second, cartels can punish by ganging up against the deviator, thus coordinating a harsh punishment that does not hurt compliers. And third, the punishment can take a form that is outside the game: instead of punishing by increasing conflict expenditure, cartels can punish by directly attacking the leadership of the deviating cartel. For a cartel leader considering deviating from a peaceful equilibrium, the threat of other cartels killing him serves as a powerful deterrent.

4.2.3 The possibility of a peaceful equilibrium

In an ideal agreement, cartels split routes evenly without any investment in conflict. Each cartel controls $\frac{1}{n}$ routes, and, from Lemma 3, each cartel

[28]Once optimal punishment exists, it is straightforward to justify that it is subgame perfect as long as the net present value from punishing is nonzero. If any cartel decides not to join the punishment, every other cartel would gang up against it, so it would be suboptimal for cartels not to punish the deviator.

obtains profits

$$\pi^a(0) = \frac{1}{n}\pi^A \tag{16}$$

where $\pi^a(0)$ stands for the profit obtained under an agreement with zero conflict expenditure.

If cartels invest nothing in conflict, they cannot defend their routes. In this world, any single cartel can then make an arbitrarily small conflict investment η and gain control of all routes. For a single period, the deviator takes all the aggregate productive profit, $\pi^d(0) = \pi^A - \eta = n\pi^a(0) - \eta$, where $\pi^d(0)$ denotes the profit when deviating from an agreement with zero conflict expenditure. From the next period on, the deviator will receive profits π^p due to others' punishment. With Nash reversion, $\pi^p = \pi^N$, and with optimal punishment, $\pi^p = 0$. The *incentive constraint* (IC) is thus $\frac{1}{1-\beta}\pi^a(0) \geq \pi^d(0) + \frac{\beta}{1-\beta}\pi^p$. Isolating β yields:

Proposition 4. *A peaceful collusive equilibrium can be sustained if* $\beta \geq \dfrac{n-1}{n - \frac{\pi^p}{\pi^A}}$. *For optimal punishment, this reduces to* $\beta \geq \dfrac{n-1}{n}$.

Proposition 4 shows that cartels can indeed coexist without any violence—if the number of cartels is sufficiently small, and if cartel leaders sufficiently value future earnings. Even though an individual cartel could seize all routes for one period, the consequent punishment will decrease its long-run profits, which deters deviation. And since the potential gain from deviating is greater when there are many cartels, the discount factor required to sustain peace increases with the number of cartels.

Perhaps surprisingly, interdiction does not affect the feasibility of peace when cartels impose optimal punishment. The reason is that interdiction affects the returns to colluding ($\frac{\pi^A}{n}$) and the returns to deviating (π^A) equally, thus not tilting cartels' incentives one way or the other. In the next section, however, we establish that interdiction *does* affect the scale of violence when peace cannot be sustained.

4.2.4 Violent equilibria

Even if peace cannot be sustained ($\beta < \frac{n-1}{n}$), we would not expect cartels to wage all-out war as they do in the SGNE. Rather, we would expect them to arrive at some intermediate symmetric outcome in which they agree on spending $\bar{g} < g^N$ on the conflict, after which each cartel ends up controlling the same share of routes as in the SGNE ($R^a = \frac{1}{n}$) but with higher profit. The logic is that non-zero conflict expenditure $\bar{g} > 0$ raises the benefits of colluding relative to the benefits of deviating: if no cartel invests in conflict, any one of them can grab all routes with infinitesimal expenditure. Punishment and $\bar{g}$ thus work together as deterrents against deviating.

In Appendix A, we show formally that

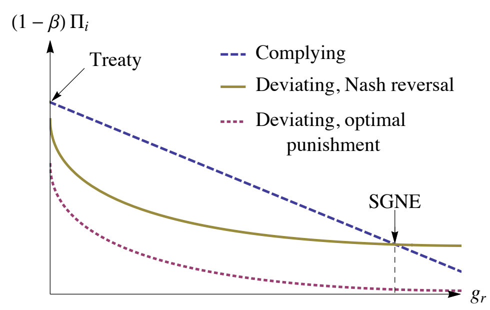
Proposition 5. *An equilibrium with less violence than the SGNE (i.e., $\bar{g} < g^N$) always exists, both with Nash reversion and with optimal punishment.*

The proof (which also appears in Appendix A) establishes that there is always a $\bar{g} < g^N$ that satisfies the new incentive compatibility constraint, which follows from modifying a general theorem stating that any SGNE can be improved with Nash reversion.

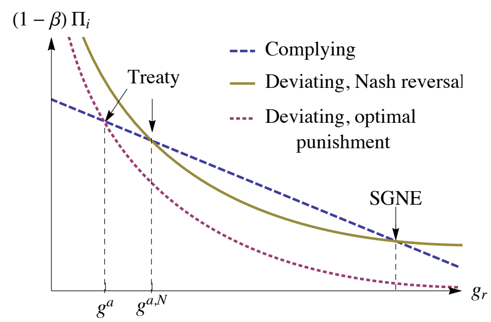
The level of violence that obtains in this equilibrium can be understood graphically. In order to maximize profits, cartels agree on the minimum conflict expenditure that ensures incentive compatibility; we denote this minimum incentive-compatible expenditure $g^{a,N}$ for the minimum under Nash reversion and g^a for the minimum under optimal punishment.[29]

Figure 1 compares the possible equilibria under high and low discount factors β. Note first that, if cartels use optimal punishment, the profits from deviating are $(1-\beta)\pi^N$ lower than with Nash reversion; this is why Figure 1 plots the profits from deviating under optimal punishment as a

[29] These quantities fulfill the incentive-compatibility constraints with equality, such that $\pi^a(g^{a,N}) = (1-\beta)\pi^d(g^{a,N}) + \beta\pi^N$ and $\pi^a(g^a) = (1-\beta)\pi^d(g^a)$.



(a) High β: The future is important enough that with zero investment in the conflict the agreement can be sustained.



(b) Low β: deviating is relatively more profitable, so cartels would deviate if they were in a peaceful equilibrium, but there still exist levels of deterrent expenditure $\bar{g} < g^N$ that enable an agreement.

Figure 1: Possible equilibria under high and low discount factors

downward shift of the curve plotting profits from deviating under Nash reversion.

With a high discount factor (Figure 1a), the profit under complying and the profit under deviating are equal when conflict expenditure is zero: Cartels value the future so much that complying beats deviating, even when other cartels invest nothing in fighting for routes. This is the peaceful equilibrium. But with a low discount factor (Figure 1b), the profit from complying and the profit from deviating are equal at some conflict expenditure greater than zero but less than the SGNE level: In Figure 1b, g^a and $g^{a,N}$ are the minimum levels of conflict expenditure for which the incentive-compatibility constraint is satisfied, under optimal punishment and Nash reversion, respectively. Clearly, violence under optimal punishment is lower than violence under Nash reversion.[30]

[30]This implies a third possibility, which is that for intermediate values of β a peaceful equilibrium exists with optimal punishment but not with Nash reversion.

4.2.5 The effect of prohibition enforcement in a violent equilibrium

Having established the existence of a violent equilibrium in a repeated game, we now analyze how interdiction affects violence in that equilibrium: Under what conditions does interdiction intensify violent conflict among cartels? Under what conditions does interdiction mitigate violent conflict among cartels?

To answer these questions, note first that aggregate *productive* behavior—that is, cartels' purchase and sale of drugs—in the repeated game is identical to that of the stage game, such that the supply of drugs reaching consumers is still given by Lemma 2. In other words, under an agreement with less violence than the SPNE, interdiction still reduces supply—but neither the number of cartels nor the discount factor affect supply.[31]

In what follows, we assume that cartels use optimal punishment rather than Nash reversion, because optimal punishment yields an equilibrium with higher profits. The incentive constraint that determines the level of conflict is then

$$\pi^a = (1 - \beta)\pi^d. \tag{17}$$

How, then, do government operations affect cartels' profits under compliance with the collusive, low-violence agreement relative to their profits under deviation from this agreement? This is the key to understanding how violence must change in order for the incentive constraint in Equation 17 to hold with equality. We find that:

Proposition 6. *If the discount factor is such that peace cannot be sustained (i.e., $\beta < \frac{n-1}{n}$), the comparative statics on the level of violence under optimal punishment are:*

[31] More formally, (1) $\partial Q^a/\partial e < 0$: interdiction reduces supply; (2) $\partial Q^a/\partial n = 0$: The number of cartels has no effect on supply; and (3) $\partial Q^a/\partial \beta = 0$: The discount factor has no effect on supply.

a) If $\epsilon_c < \hat{\epsilon}_t$*, then* $\dfrac{\partial G^a}{\partial e} < 0$*: If demand is sufficiently elastic, interdiction reduces violence.*

b) If $\epsilon_c > \hat{\epsilon}_t$*, then* $\dfrac{\partial G^a}{\partial e} > 0$*: If demand is sufficiently inelastic, interdiction increases violence.*

c) $\dfrac{\partial G^a}{\partial n} > 0$*: An increase in the number of cartels increases the level of violence.*

d) $\dfrac{\partial G^a}{\partial \beta} < 0$*: More forward-looking cartels decreases the level of violence.*

Proof. See Appendix A. □

The first two statements in Proposition 6—about the response of violence to enforcement, $\frac{\partial G^a}{\partial e}$—echo similar results from the SGNE analysis, but the intuition behind them is very different. In the stage-game analysis, inelastic demand implied that interdiction increased violence simply because interdiction increased productive profit π^A, thereby raising the stakes of the conflict. In the repeated game, in contrast, interdiction increases violence because it increases the profit from deviating *more* than it increases the profit from complying.

To see this, recall the expressions for each cartel's profit under complying with a low-violence agreement (π^a) and for the profit under deviating (π^d): $\pi^a = R^a \pi^A - \bar{g}$ and $\pi^d = R^d \pi^A - g^d$. Because the contest success function has diminishing returns, when a deviating cartel increases conflict expenditure from $\bar{g}$ to g^d, its share of routes increases by a smaller proportion: $\frac{R^d}{R^a} < \frac{g^d}{\bar{g}}$. This implies that the profit margin is lower when deviating than when complying. So while any increase in aggregate profit π^A causes the same percent increase in *revenues* for deviators and for compliers, that same increase in π^A causes a larger percent increase in *profits* for deviators than for compliers. *This* is the mechanism that links interdiction to higher violence (results a) and b) in Proposition 6), *not* the simple fact of interdiction increasing aggregate productive profit.

To illustrate, imagine that there are ten cartels, and that initially total productive profit π^A is \$100. Imagine further that the deterrent expenditure required to sustain a low-violence equilibrium is \$1; since the equilibrium is symmetric, each cartel controls 1/10 of routes, earning \$10 in productive profit ($R^a \pi^A$) and \$9 in overall profits ($\pi^a = R^a \pi^A - \bar{g}$).

Now consider a cartel weighing whether to deviate, increasing conflict expenditure in order to double its share of routes from 1/10 to 1/5. Because of the declining returns to conflict expenditure, the would-be deviator must more than double expenditure in order to double its route share—for the sake of example, say that it must spend \$5 to control 1/5 of routes. For one glorious period, then, the deviator would earn $\pi^d = \$20 - \$5 = \$15$, a 67% increase over its complying profits of $\pi^a = \$9$; it would then earn zero profits thereafter. Since we defined \$1 as sufficient deterrent expenditure to sustain a low-violence equilibrium, we already know that the cartel values the future sufficiently not to deviate. The one-period 67% increase in profits just isn't worth it.

Now imagine that total productive profit increases to $\pi^A = \$200$. In this new world, our cartel earns $\pi^a = \$20 - \$1 = \$19$ by complying. But by deviating, the cartel would earn $\pi^d = \$40 - \$5 = \$35$. Just as before, revenues double—but *profits* from deviating now increase by 84%, significantly more than 66%. Even with the same discount factor β, then, this might be enough to tempt deviators—in which case deterrent expenditure, and thus violence, would have to increase in order to sustain a new agreement.

Similarly, the interpretation of the result that more cartels increase violence (result c)) differs from the interpretation of the analogous result in the SGNE. In the repeated game, a larger number of cartels increases the rewards of deviating, since a deviating cartel could take over the routes of all other cartels. Cartels must therefore raise equilibrium conflict expenditure—and violence—in order to deter potential deviators.

Result d) is perhaps the most straightforward: cartels that place higher value on the future are more easily deterred by the threat of future punishment, allowing cartels to reduce deterrent conflict expenditure and thereby to reduce violence.

5 Prohibition enforcement vs. peace: empirical evidence

The model makes two closely related predictions about the consequences of counternarcotics operations. The first prediction is that government operations aimed at drug cartels' productive operations—such as seizing drug shipments along drug routes, or capturing boats or airplanes used for drug transport—can indeed reduce the supply of drugs reaching consumers, but only at the cost of increased violence.[32] The second prediction is that government operations aimed at capturing or killing high-profile cartel bosses has only a minimal (if any) effect on the supply of drugs reaching consumers, while also increasing violent conflict.

Here, we use case evidence to discuss how these predictions help us make sense of the evolution of why governments so often retreat from prohibition enforcement, and of why the most effective policies paradoxically fuel political opposition.[33] We focus on two cases: cocaine prohibition in the Americas, and alcohol prohibition in the United States.

[32] If demand is sufficiently inelastic; estimates suggest that it is (see Footnote 3).

[33] While the model focuses on policy ideal types—interdiction that disrupts productive operations without affecting fragmentation or leaders' time horizons; "beheadings" that break up trafficking organizations and make leaders impatient without affecting productive operations—our discussion recognizes that, in reality, governments invest in a combination of these two types of policies.

5.1 Cocaine prohibition in the Americas

In the 1990s and early 2000s, the Colombian and U.S. governments focused counternarcotics efforts on eradicating coca crops (Mejía et al., 2017; Mejía and Restrepo, 2016). But crop eradication proved highly ineffective: producers responded by moving to locations where aerial spraying was more difficult, and by improving crop yields (thus requiring less land to produce the same amount of coca). Journalists, academic researchers, and the public criticized the crop eradication policy; one writer later compared it to "trying to drive up the price of fine art by raising the cost of paint" (Wainwright, 2016). Mejía et al. (2017) concluded that spraying one additional hectare of coca reduces cultivation by 0.022–0.3 hectares, a reduction far too small to make aerial spraying a cost-effective enforcement policy. In response, the Colombian and U.S. governments switched strategies, investing instead in drug interdiction and laboratory destruction. Mejía and Restrepo (2016) estimate the marginal cost (to the government) of reducing retail cocaine supply via interdiction as less than one-fifth of the marginal cost of reducing supply via crop eradication.

One might have thought that a fivefold increase in the effectiveness of supply-reduction efforts would buoy public support for the war on drugs. But successful supply reduction had an unintended consequence: higher prices and increasing profits for the Mexican drug cartels that bring Colombian cocaine to consumers. Figure 2a plots (purity-adjusted) cocaine prices in the United States against the net cocaine supply coming from Colombia. While our model does not attempt to explain long-term trends in price or quantity, or to account for all variation in price or quantity, the negative correlation in Figure 2a illustrates empirically one of our theoretical findings: that investing in interdiction can reduce supply and thereby increase price.

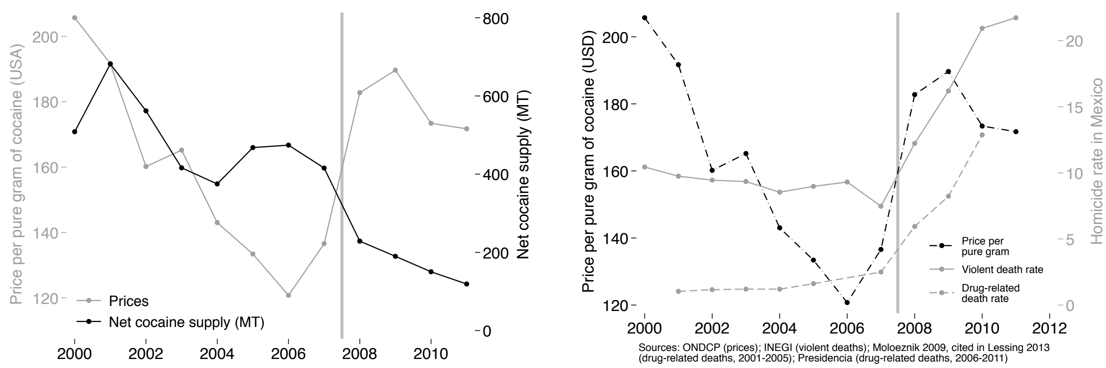
Simultaneously, violence in Mexico doubled (Figure 2b). Castillo et al.

Figure 2: Cocaine prohibition and violence in Mexico

Figure (a), drawn from Castillo et. al. 2016, suggests that more effective counternarcotics enforcement in Colombia reduced the net supply of cocaine reaching the United States, increasing prices and thereby the profits of Mexican cartels. Figure (b) suggests that violence in Mexico increased simultaneously; Castillo et al. (2016) find that higher profits indeed fueled violence in Mexico.

(a) Cocaine supply and prices

(b) Prices and violence



(2016) provide evidence that this relationship was causal: using monthly data on drug seizures in Colombia and municipality-level monthly homicide data from Mexico, they exploit variation in Mexican municipalities' exposure to drug trafficking in order to evaluate the effect of the supply reduction on violence. They estimate that a 1% decrease in the supply of cocaine (because of increased seizures) drove a 0.119%–0.161% increase in homicide rates in the Mexican municipalities most exposed to drug trafficking.

Of course, cocaine seizures in Colombia—and the consequent negative supply shock—were not the only policy change affecting drug-related violence in Mexico in the mid-2000s. In December 2006, Mexican President Felipe Calderón launched what was called a "kingpin strategy" or "decapitation strategy," aimed at arresting or executing high-profile leaders of drug cartels (most infamously "El Chapo," Joaquín Guzmán Loera) (The New York Times, 2016). Using an original data set on the timing and location of these captures, Calderón et al. (2015) find that removing cartel kingpins and lieutenants "has exacerbating short-term effects not only on DTO-related violence, but also on homicides that affect the general population."[34,35]

The model in this paper clarifies the mechanisms driving the increase in lethal violence in Mexico. Prior to 2006, Mexican cartels ran a massive cocaine trafficking operation with very few drug-related homicides (Figure 2b)—this is what our model characterizes as a peaceful equilibrium. During this period, as predicted by our model, violence did not appear to respond to changes in profits (see Appendix E). Calderón's crackdown, which began in late 2006 and targeted cartel leaders, splintered the cartels and made capos more short-sighted. The peaceful equilibrium broke down. At the same time, a surge in profits—driven by Colombian policy changes—increased the returns to fighting *relative* to the returns to complying, requiring cartels to spend more on conflict in order to sustain a new, more violent equilibrium.

The wave of violence driven by attacking kingpins and by supply reduction fanned the flames of anti-prohibition sentiment. Upon taking office as President of Guatemala in January of 2012, Otto Perez Molina—a graduate of the Pentagon's School of the Americas and a U.S. ally—called for legalizing drugs and a cease-fire with drug traffickers (Guillermoprieto, 2012). At a regional summit later that year, the reaction from other Latin American heads of state was "strikingly favorable"; President Juan Manuel Santos of Colombia, who hosted the meeting, included legalization on the official agenda, and the summit was later dubbed "The Rebel-

[34]Dell (2015) also finds that Calderón's crackdown increased violence among traffickers, though she attributes this primarily to the fact that government attacks "weaken the incumbent DTO, creating incentives for rival DTOs to violently wrest control of a territory while the incumbent is vulnerable," rather than to conflict among splinters of the original group.

[35]Cunningham (2013) provides empirical evidence that civil war is more likely when opposition movements are internally fragmented. While the setting is different, the logic is related to ours: fragmented groups make peace harder to sustain.

lion at Cartagena," in reference to the rejection of U.S. anti-cocaine policies (ibid). That same year, Mexico elected a presidential candidate who campaigned on abandoning his predecessor's military offensive against drug cartels (Fabián, 2012), and even that predecessor—Felipe Calderón—publicly called for reevaluating supply-reduction policies (Calderón, 2012). In the United States, too, spending priorities shifted away from supply reduction (Ingraham, 2016), and President Barack Obama referred to illegal drug use as "a health problem, not a criminal problem" (Pratt, 2016).

5.2 Alcohol prohibition in the United States

That alcohol prohibition in the United States fueled violence is firmly established both in the academic literature and in Hollywood films (e.g., Miron, 1999; Owens, 2011, 2014). We also know that the violence stemmed largely from conflict over the illegal alcohol market (ibid). Equally apparent is that gangs sometimes divided the alcohol business peacefully: Okrent (2010) provides numerous examples both of gang treaties ("you take the north side, I'll take the south") and also of "escalating arms races" among competing criminal organizations (p. 275).

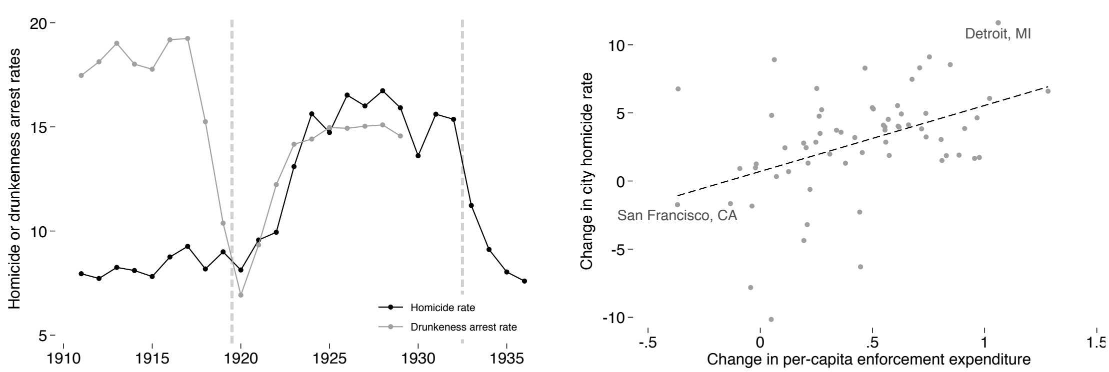
What accounts for this variation? García-Jimeno (2016) collected data on the intensity of the enforcement of Prohibition, which varied dramatically both across cities (and states) and over time. In Okrent's simplification, local enforcement "took on one of two humors—either a vigor that outshone federal efforts or something close to torpor" (p. 255). Using the enforcement data together with local homicide data, and in the context of a dynamic structural model, García-Jimeno estimates the elasticity of crime to prohibition enforcement, finding "that the Prohibition-related homicide rate was increasing in the level of law enforcement" (p. 513).

To illustrate, Figure 3b plots change in the homicide rate against change in per-capita law enforcement expenditures in 67 cities, using García-Jimeno's replication data.[36] (We define *change* as the mean values for all years *during* Prohibition minus the mean for all years *before* Prohibition.)[37] Consider two examples. In San Francisco, one of a few cities where violence actually *decreased* during Prohibition, Okrent describes "you've-gotta-be-kidding-me disdain" among local enforcement authorities and a federal Prohibition administrator who "was a self-professed drinker eventually indicted for diversion of seized liquor" (p. 258). While wine grape prices initially skyrocketed due to increased demand from home fermenters across the country, grape and wine production increased unchecked until prices fell again (p. 180). And since wine distribution was also unrestricted, San Francisco became, in the view of one Sonoma County winemaker, "the only place in the United States where the distribution of wine was practiced without guns" (p. 258). At the other end of the spectrum, per-capita enforcement expenditures increased more in Detroit than almost anywhere else—and rum running on the Detroit river turned the waterfront into a war zone (Young, 1929; Okrent, 2010, p. 320–321).

Figure 3: Alcohol prohibition and violence in the United States

García-Jimeno (2016) finds that, during alcohol Prohibition in the United States, "that the Prohibition-related homicide rate was increasing in the level of law enforcement" (513).

(a) Prohibition fueled violence . . . (b) . . . in proportion to enforcement intensity



[36]Note, however, that this graph does not appear in his paper, and that this cross-sectional comparison in no way captures his empirical approach.

[37]Prohibition starts before the Eighteenth Amendment in a number of cities and states.

While previous work establishes the connection between Prohibition enforcement and violent crime, our model illuminates the intervening mechanisms: that enforcement reduced the supply of alcohol, driving prices up and increasing gangs' incentives to fight rather than abide by treaties.[38] Capturing and killing leaders of trafficking organizations—most famously, imprisoning Al Capone in 1931—may have exacerbated the problem by making traffickers more impatient and by splintering their organizations. Certainly, contemporary observers and historians share the impression that (as we find) these beheadings did nothing to reduce the supply of alcohol; after Capone's imprisonment, "rival mobs soon satisfied Chicago's thirst" (Okrent, 2010, p. 245).

Garcia-Jimeno documents voter response to the violent consequences of Prohibition, finding that "learning proved key in modulating its effects by leading communities to partially offset the policy through weakened law enforcement choices" (García-Jimeno, 2016, p. 514). In other words, as "any doubt about Prohibition's impact on the spread of violent crime had melted away" (Okrent 321), voters increasingly supported candidates and referenda that limited Prohibition enforcement. This is precisely the consequence implied by our model: enforcement of prohibition creates violence, the public balks, the government pulls back.

6 Conclusion

Why do governments so often permit *de facto* the sale of goods that are *de jure* prohibited? To help answer this question, we extend the formal analysis of criminal trafficking organizations from a single-period game to

[38]To the best of our knowledge, there are no local data on alcohol prices, but prices certainly increased overall; one estimate put sales at $3.6 billion in 1926, "almost precisely the same as the entire federal budget that year—army, navy, and every other government function included" in that year (Okrent p. 274).

a repeated-interaction, dynamic setting. In the model, traffickers are profit-maximizing organizations that buy illicit goods from producers, move them to consumer markets, and sell to consumers; transporting illicit goods requires both controlling routes—for which they must compete with other traffickers—and evading government forces trying to seize shipments.

If traffickers are able to use incentives created by repeated interaction in order to divide routes (and thus profits) peacefully, they can coexist with lower levels of violence than those predicted by the single-period interaction analyzed in previous papers. But government operations darken the prospects of a low-violence equilibrium. By attacking organized crime bosses, the government shortens their time horizons and spurs fragmentation—both of which, in the model, make it more difficult for criminal organizations to coordinate peace. Moreover, attacking capos largely does not affect the supply of the illegal drug to consumers; the model shows that the aggregate *productive* behavior of traffickers is independent of which organizations control transport routes. By attacking traffickers' productive operations, on the other hand, the government does successfully shrink the supply of the illegal good reaching consumers—but this policy, too, has unintended consequences. Previous papers have noted that shrinking the supply of an illegal good raises prices and, if demand is sufficiently inelastic, increases total revenues. Our model extends the analysis from total revenues to total profits, finding that interdiction can increase profits even when demand is slightly elastic (because it affects cartels' costs in addition to their revenues). Higher profits, in turn, stoke violence—not only because cartels invest more in fighting over a more valuable prize, but also because enforcement increases the benefits of deviating from a collusive, low-violence agreement *relative* to the benefits of complying with such an agreement. To the best of our knowledge, this effect of enforcement on traffickers' dynamic incentives has not been explored in other work.

Using case evidence, we then suggest that these connections between counternarcotics operations and violence help account for changes in the enforcement of anti-cocaine policy in Latin America and anti-alcohol policy in the United States—changes puzzling in light of other political economy explanations such as state weakness or ineffectiveness of available enforcement technology. Our explanation, instead, emphasizes the paradox of prohibition policies: as they become more effective at reducing supply, so they become more difficult to sustain politically.

References

Abreu, Dilip, "Repeated games with discounting: A general theory and an application to oligopoly.," *Ph.D. thesis, Princeton University*, October 1983.

Acemoglu, Daron, James A Robinson, and Rafael J Santos, "The Monopoly of Violence: Evidence from Colombia," *Journal of the European Economic Association*, January 2013, *11*, 5–44.

Baccara, Mariagiovanna and Heski Bar-Isaac, "How to Organize Crime," *Review of Economic Studies*, 2008, *75* (4), 1039–1067.

Bachman, Jerald G., Lloyd D. Johnston, and Patrick M. O'Malley, "Explaining the Recent Decline in Cocaine Use among Young Adults: Further Evidence That Perceived Risks and Disapproval Lead to Reduced Drug Use," *Journal of Health and Social Behavior*, 1990, *31* (2), 173–184.

Becker, Gary S, "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, March 1968, *76* (2), 169–217.

Becker, Gary S., Kevin M. Murphy, and Michael Grossman, "The Market for Illegal Goods: The Case of Drugs," *Journal of Political Economy*, 2006, *114* (1), 38–60.

Bueno de Mesquita, Ethan, "The Quality of Terror," *American Journal of Political Science*, 2005.

Burrus, Robert T., "Do Efforts to Reduce the Supply of Illicit Drugs Increase Turf War Violence? A Theoretical Analysis," *Journal of Economics and Finance*, 1999, *23* (3), 226–234.

Calderón, Felipe, "Mexican president Felipe Calderon United Nations General Assembly Address," C-Span September 2012.

Calderón, Gabriela, Gustavo Robles, Alberto Díaz-Cayeros, and Beatriz Magaloni, "The Beheading of Criminal Organizations and the Dynamics of Violence in Mexico," *Journal of Conflict Resolution*, 2015, *59* (8), 1455–1485.

Castillo, Juan Camilo, Daniel Mejía, and Pascual Restrepo, "Scarcity without Leviathan: The Violent Effects of Cocaine Supply Shortages in the Mexican Drug War," *Working paper*, 2016.

Chimeli, Ariaster B. and Roy G. Boyd, "Prohibition and the Supply of Brazilian Mahogany," *Land Economics*, 2010, *86* (1), 191–208.

Cunningham, Kathleen Gallagher, "Actor Fragmentation and Civil War Bargaining: How Internal Divisions Generate Civil Conflict," *American Journal of Political Science*, 2013, *57* (3), 659–672.

Dell, Melissa, "Trafficking Networks and the Mexican Drug War," *American Economic Review*, 2015, *105* (6), 1738–79.

DiNardo, John, "Law enforcement, the price of cocaine and cocaine use," *Mathematical and Computer Modeling*, 1993, *17* (2), 53–64.

Duran-Martinez, Angelica, "To Kill and Tell? State Power, Criminal Competition, and Drug Violence," *Journal of Conflict Resolution*, 2015, *59* (8), 1377–1402.

Fabián, Jordan, "Mexico President Enrique Pena Nieto Charts New Drug War Plan," *ABC News*, December 2012.

Fearon, James D., "Rationalist Explanations for War," *International Organization*, 1995, *49* (3), 379–414.

Fey, Mark and Kristopher W. Ramsay, "Mutual Optimism and War," *American Journal of Political Science*, 2007.

Franklin, Jimmie Lewis, *Born sober, prohibition in Oklahoma, 1907–1959*, University of Oklahoma Press, 1971.

Gallet, Craig A., "Can Price Get the Monkey Off Our Back? A Meta-Analysis of Illicit Drug Demand," *Health Economics*, 2014, *23* (1), 55–68.

García-Jimeno, Camilo, "The Political Economy of Moral Conflict: An Empirical Study of Learning and Law Enforcement under Prohibition," *Econometrica*, 2016, *84* (2), 511–570.

Guillermoprieto, Alma, "Drugs: The Rebellion in Cartagena," June 2012.

Holland, Alisha C., "Forbearance," *American Political Science Review*, 2016, *110* (2), 232–246.

Ingraham, Christopher, "The radical way Obama wants to change the drug war," *The Washington Post*, February 2016.

Jackson, Matthew O. and Massimo Morelli, "Strategic Militarization, Deterrence and Wars," *Quarterly Journal of Political Science*, 2009, *4* (4), 279–313.

Kleiman, Mark, "Surgical Strikes in the Drug Wars," *Foreign Affairs*, 2011.

Kornbluh, Peter, "Comment on the Invasion of Panama," *The New Yorker*, 1991.

Kydd, Andrew, "Which Side Are You On? Bias, Credibility, and Mediation," *American Journal of Political Science*, 2003, *47* (4), 597–611.

Lessing, Benjamin, "The Logic of Violence in Criminal War: Cartel-State Conflict in Mexico, Colombia, and Brazil," *Ph.D. Thesis, University of California, Berkeley*, 2012.

_ , "Logics of Violence in Criminal War," *Journal of Conflict Resolution*, 2015, *59* (8), 1486–1516.

Levitsky, Steven and María Victoria Murillo, "Variation in Institutional Strength," *Annual Review of Political Science*, June 2009, *12* (1), 115–133.

Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green, *Microeconomic Analysis*, Oxford University Press, 1995.

Mejía, Daniel and Pascual Restrepo, "The war on illegal drugs in producer and consumer countries: A simple analyticial framework," *Documento CEDE 2011-02 - Universidad de los Andes*, 2011.

_ **and** _ , "The economics of the war on illegal drug production and trafficking," *Journal of Economic Behavior Organization*, 2016, *126*, 255 – 275.

_ **,** _ **, and Sandra V. Rozo**, "On the Effects of Enforcement on Illegal Markets: Evidence from a Quasi-Experiment in Colombia*," *The World Bank Economic Review*, 2017, *31* (2), 570.

Miron, JA, "Violence and the U.S. prohibitions of drugs and alcohol," *American Law and Economics Review*, 1999, *1* (1), 78.

Moloeznik, Marcos Pablo, "Principales efectos de la militarización del combate al narcotráfico en México," *Renglones*, 2009.

Okrent, Daniel, *Last Call: The Rise and Fall of Prohibition*, Scribner, 2010.

Owens, Emily G., "Are underground markets really more violent? Evidence from early 20th century America," *American Law and Economics Review*, 2011, *13* (1), 1–44.

_ , "The American Temperance Movement and Market-Based Violence," *American Law and Economics Review*, 2014, *16* (2), 433.

Poret, Sylvaine, "Paradoxical effects of law enforcement policies: the case of the illicit drug market," *International Review of Law and Economics*, 2003, *22*, 465–493.

_ **and Cyril Téjédo**, "Law enforcement and concentration in illicit drug markets," *European Journal of Political Economy*, 2006, *22*, 99–114.

Powell, Robert, "Guns, Butter, and Anarchy," *The American Political Science Review*, 1993, *87* (1), 115–132.

_ , "Bargaining theory and international conflict," *Annual Review of Political Science*, 2002, *5* (1), 1–30.

Pratt, Timothy, "Obama: 'Drug addiction is a health problem, not a criminal problem'," *The Guardian*, March 2016.

Reuter, Peter, "The Political Economy of Drug Smuggling," in Menno Vellinga, ed., *The Political Economy of the Drug Industry: Latin America and the International System*, University Press of Florida, 2004.

Rotemberg, Julio J. and Garth Saloner, "A Supergame-Theoretic Model of Price Wars during Booms," *American Economic Review*, 1986, *76* (3), 390–407.

Saffer, Henry and Frank Chaloupka, "The Demand for Illicit Drugs," *Economic Inquiry*, 1999, *37* (3), 401–411.

Skaperdas, Stergios, "Contest Success Functions," *Economic Theory*, 1996, *7*, 283–290.

The New York Times, "Mexico's Kingpin Strategy Against the Drug Cartels," *The New York Times*, February 2016.

Wagner, R Harrison, *War and the state: The theory of international politics*, University of Michigan Press, 2010.

Wainwright, Tom, *Narconomics* How To Run a Drug Cartel, Random House, February 2016.

Young, James, "Rum War Forces Mass on the Detroit Front," *The New York Times*, 1929.

Appendix A Formal statements and proofs of propositions and lemmas

A.1 Properties of the production and survival functions

Lemma. *The production function $q(x_i, R_i, e)$ has the following properties:*

1. *It is homogeneous of degree one in (x_i, R_i).*
2. *It is concave in (x_i, R_i).*
3. *Routes and drugs are complementary production factors ($\frac{\partial^2 q}{\partial x_i \partial R_i} > 0$).*

The survival rate $w(x_i, R_i, e)$ has the following properties:

4. *It is increasing in the inverse route saturation rate r_i ($\frac{\partial w}{\partial r_i} > 0$).*
5. *The marginal productivity of r_i is decreasing ($\frac{\partial^2 w}{\partial r_i^2} < 0$).*

Proof. 1 is straightforward to check. For 2, the conditions on the first derivatives, on both second derivatives, and on the cross derivatives (which we will soon state) imply that the function is quasiconcave. Quasiconcavity and homogeneity of degree one imply concavity. For 3, $\frac{\partial^2 q}{\partial x_i \partial R_i} = x\frac{\partial^2 w}{\partial x_i \partial R_i} + \frac{\partial w}{\partial R_i}$. By the chain rule, $\frac{\partial w}{\partial R_i} = \frac{\partial w}{\partial r_i}\frac{\partial r_i}{\partial R_i}$ and $\frac{\partial^2 w}{\partial x_i \partial R_i} = \frac{\partial^2 w}{\partial r_i^2}\frac{\partial r_i}{\partial R_i}\frac{\partial r_i}{\partial x_i} + \frac{\partial w}{\partial r_i}\frac{\partial^2 r_i}{\partial x_i \partial R_i}$. The derivatives of r_i can be readily calculated. Substituting everything in the initial expression for the cross derivative of q yields $\frac{\partial^2 q}{\partial x_i \partial R_i} = -\frac{R_i}{x_i^2}\frac{\partial^2 w}{\partial r_i^2}$, which is positive due to the decreasing marginal productivity of r_i. 4 and 5 can be easily checked by finding the derivatives of w with respect to R_i, holding x_i fixed. □

A.2 Lemma 1: Functional form of the contest success function

Proof. If m cartels spend g and every other cartel spends zero, $1 = \sum R_i(g_i, g_{-i}) = mR(g, (m-1)g) = mR(1/m, 1-1/m) \implies R(1/m, 1-1/m)$. If p cartels spend g and one cartel spends qg, the first cartels get $R(g,(p-1+q)g) = R(1/(p+q), 1-1/(p+q)) = 1/(p+q)$ routes, so the other cartel gets $R(q/(p+q-1), 1-1/(p+q-1)) = 1-p/(p+q) = q/(p+q)$. This pins down the function for every rational $s \in (0,1)$ as $R(s, 1-s) = s$. By continuity, this also has to be true for every real $s \in (0,1)$. And by homogeneity of degree zero, $R(g, G-g) = R(g/G, 1-g/G) = g/G$. □

A.3 Lemma 2: Aggregate supply to consumer markets

Proof. For two different cartels i and j, $\frac{\partial q(x_i^*, R_i^*, e)}{\partial x_i} = \frac{\partial q(x_j^*, R_j^*, e)}{\partial x_j}$. Since q is homogeneous of degree one, its derivative is homogeneous of degree zero, so $\frac{\partial q(x_i^*/R_i^*, 1, e)}{\partial x_i} = \frac{\partial q(x_j^*/R_j^*, 1, e)}{\partial x_j}$, and since this derivative is strictly decreasing, $\frac{x_i^*}{R_i^*} = \frac{x_j^*}{R_j^*}$. Thus, $\frac{x_i^*}{x_j^*} = \frac{\hat{R}_i}{\hat{R}_j} = \frac{q_i^*}{q_j^*}$, where we used homogeneity of q again for the last step. Summing over i yields $\frac{X}{x_j^*} = \frac{1}{\hat{R}_j} = \frac{Q}{q_j^*}$, so $X = \frac{x_i^*}{\hat{R}_i}$. Substituting in (3), taking into account the homogeneity of q, yields the condition for X. For Q, $Q = \sum_i q(\hat{R}_i X, \hat{R}_i, e) = \sum_i \hat{R}_i q(X, 1, e) = q(X, 1, e)$. For the shares of production, $q_i = q(x_i, \hat{R}_i, e) = \hat{R}_i q(X, 1, e)$. □

A.4 Lemma 3: Defining cartel profit

Proof. Cartel i invests g_i in the conflict, buys an amount of drugs $x_i = R(g_i, g_{-i})X$ and sells an amount $q_i = R(g_i, g_{-i})Q$. Substituting in (2) yields $\pi_i = p_c q_i - p_p x_i - g_i = (p_c Q - p_p X) R(g_i, g_{-i}) - g_i$. □

A.5 Proposition 3: Enforcement and violence in the SGNE

Proof. Individual expenditure is determined by (14). Enforcement has no direct effect on $\frac{\partial R_i}{\partial g_i}$, whereas g^N has no direct effect on π^A. This, and the implicit function theorem, lead to $\frac{\partial g^N}{\partial e} = -\frac{\frac{\partial \pi^A}{\partial e}}{\frac{\partial}{\partial g^N}\left(\frac{\partial R}{\partial g_i}\right)}$. The denominator is negative: $\frac{\partial^2 R}{\partial g^N \partial g_i} = \frac{\partial^2 R}{\partial g_i^2} + (n-1)\frac{\partial^2 R}{\partial g_{-i}\partial g_i} < 0$. Thus, the sign of the effect of enforcement is the same as the sign of the effect on π^A.

For the effect of the number of cartels, also take (14). Since π^A does not depend on the number of cartels, $\frac{\partial R_i}{\partial g_i}$ cannot depend on it either, so its derivative with respect to n must be zero: $\frac{\partial^2 R}{\partial n \partial g_i} = \left[\frac{\partial^2 R}{\partial g_i^2} + (n-1)\frac{\partial^2 R}{\partial g_i \partial g_{-i}}\right]\frac{\partial g^N}{\partial n} + g^N \frac{\partial^2 R}{\partial g_i \partial g_{-i}} = 0$. Isolating $\frac{\partial g^N}{\partial n}$ yields $\frac{\partial g^N}{\partial n} = -g^N \frac{\partial^2 R}{\partial g_i \partial g_{-i}}\left[\frac{\partial^2 R}{\partial g_i^2} + (n-1)\frac{\partial^2 R}{\partial g_i \partial g_{-i}}\right]^{-1}$.

In order to find the comparative statics on aggregate violence G, we use the fact that $\frac{\partial G^N}{\partial n} = g^N + n\frac{\partial g^N}{\partial n}$ to obtain

$$\frac{\partial G^N}{\partial n} = g^N \left[\frac{\partial^2 R}{\partial g_i^2} - \frac{\partial^2 R}{\partial g_i \partial g_{-i}}\right]\left[\frac{\partial^2 R}{\partial g_i^2} + (n-1)\frac{\partial^2 R}{\partial g_i \partial g_{-i}}\right]^{-1}$$

whose sign is undetermined, since it depends on whether $\frac{\partial^2 R}{\partial g_i^2}$ or $\frac{\partial^2 R}{\partial g_i \partial g_{-i}}$ is greater. However, it is negative in a symmetric equilibrium:

$R(g, (n-1)g) = \frac{1}{n}$, so

$$\begin{aligned}\frac{d^2 R(g,(n-1)g)}{\partial d g^2} &= 0 \\ &= \frac{\partial^2 R(g,(n-1)g)}{\partial g_i^2} + 2(n-1)\frac{\partial^2 R(g,(n-1)g)}{\partial g_i \partial g_{-i}} \\ &\quad + (n-1)^2 \frac{\partial^2 R(g,(n-1)g)}{\partial g_{-i}^2}\end{aligned}$$

Let $S(g_i, g_{-i}) = 1 - R(g_i, g_{-i})$, the fraction of routes held by all cartels other than i. In a symmetric equilibrium $S = (n-1)R$, and $g_{-i} = (n-1)g_i$, so $\frac{\partial^2 R}{\partial g_{-i}^2} = \frac{\partial^2 (1-S)}{\partial g_{-i}^2} = -\frac{\partial^2((n-1)R)}{\partial((n-1)g_i)^2} = -\frac{1}{n-1}\frac{\partial^2 R}{\partial g_i^2}$. Substituting above yields $(n-2)\frac{\partial R(g,(n-1)g)}{\partial g_i^2} = 2(n-1)\frac{\partial R(g,(n-1)g)}{\partial g_{-i}^2}$, so $\frac{\partial^2 R}{\partial g_i^2} < \frac{\partial^2 R}{\partial g_i \partial g_{-i}}$ and $\frac{\partial G^N}{\partial n} > 0$. □

A.6 Proposition 5: Existence of an equilibrium with less violence than the SGNE

Note that if each cartel spends $\bar{g} < g^N$ on conflict expenditure, each cartel's profit is:

$$\pi^a(\bar{g}) = \pi^A R^a - \bar{g} = \frac{1}{n}\pi^A - \bar{g} \tag{18}$$

Productive behavior is the same as in the stage game, so the only difference between this and the SGNE is lower violence (and lower cost of violence). Comparing the profit from colluding to the profit in the SGNE yields $\pi^a(\bar{g}) = \pi^N + g^N - \bar{g}$.

Each cartel's profit would increase if cartels could agree to reduce $\bar{g}$, since the agreement is such that each cartel controls the same share of routes regardless of the level of $\bar{g}$. But if $\bar{g}$ is too low, the incentive to deviate overwhelms the incentive to collude: grabbing a large share of routes for one period becomes too cheap. Thus, $\bar{g}$ cannot be arbitrarily low; in equilibrium, $\bar{g}$ will be as low as possible while still functioning as a deterrent.

Since the benefit of deviating from the cooperative strategy only lasts for one period, the deviating cartel would want to take as much profit as possible in that one period (given that all other cartels spend $\bar{g}$). Thus, the profit obtained by the deviator i is:

$$\pi^d(\bar{g}) = \max_{g_i} \left[\pi^A R(g_i, (n-1)\bar{g}) - g_i\right] \tag{19}$$

The first order condition is the same as for the SGNE, Equation (14), but with expenditure by other cartels evaluated at $g_{-1} = n\bar{g}$.[39]

Solving (19) gives the optimal expenditure for the deviator g^d, which

[39] As in the SGNE, we stated the problem as two maximizations over different variables, but the cartel solves a joint maximization problem over both variables. Thus, we must check the optimality of the joint maximization problem, but it leads to the exact same second order conditions from appendix B.

determines its optimal share of routes $R^d = R(g^d, (n-1)\bar{g})$. The deviator's profit is then

$$\pi^d(\bar{g}) = \pi^A R^d - g^d \tag{20}$$

The punishment strategy used, either Nash reversion or optimal punishment, determines the lowest level $\bar{g}$ that sustains a collusive equilibrium. Since optimal punishment maximizes incentives against betraying, it allows $\bar{g}$ to be as low as possible. Our aim will be to analyze what happens in that case. We will also look at what happens with Nash reversion, but only because it will be useful in proving some results regarding optimal punishment.

The total profits from deviating are therefore $\pi^d + \frac{\beta}{1-\beta}\pi^p$, while the total profits from complying are $\frac{1}{1-\beta}\pi^a$, so the new incentive constraint is:

$$\pi^a(\bar{g}) \geq (1-\beta)\pi^d(\bar{g}) + \beta\pi^p \tag{21}$$

If there exists some deterrent expenditure such that (21) is satisfied, an equilibrium with $\bar{g}$ can be sustained.

With these elements in mind, we can now prove Proposition 5:

Proof. Setting $\bar{g} = g^N$, $\pi^a(g^N) = \pi^N$, since all cartels spend the amount corresponding to the SGNE. A deviator's one-period optimal response is g^N, so $\pi^d(g_N) = \pi^N$. Thus, with Nash reversion the IC is satisfied with equality at $\bar{g} = g^N$.

From (18), $\frac{\partial \pi^a}{\partial \bar{g}} = -1$. From the envelope theorem on (19), $\frac{\partial \pi^d}{\partial \bar{g}} = \pi^A(n-1)\frac{\partial R}{\partial g_{-i}}$. From the betrayer's first order condition, $\pi^A \frac{\partial R}{\partial g_i} = 1$, so $\frac{\partial \pi^d}{\partial \bar{g}} = \pi^A \left(\frac{\partial R}{\partial g_i} + (n-1)\frac{\partial R}{\partial g_{-i}}\right) - 1$. For $\bar{g} = g^N$ the term in parentheses is $\frac{\partial R(\bar{g},(n-1)\bar{g})}{\partial \bar{g}}$, which is zero because all cartels increase their expenditure by the same amount. Thus, $\left.\frac{\partial \pi^d}{\partial \bar{g}}\right|_{\bar{g}=g^N} = -1$. This means that at $\bar{g} = g^N$ the derivative on the left hand side of the IC is -1, which is lower than the derivative on the right hand side, $-(1-\beta)$. So the IC holds strictly for some $\bar{g} < g^N$ with Nash reversion.[40]

With optimal punishment, the profit from betraying is lower than with Nash reversion. Thus, any level of $\bar{g}$ that satisfies the IC with Nash reversion also satisfies it with optimal punishment. □

A.7 Proposition 6: Comparative statics of violence in a collusive equilibrium

The implicit function theorem on (17) yields the following derivatives:

$$\frac{\partial g^a}{\partial \beta} = -\frac{\pi^d}{\frac{\partial \pi^a}{\partial \bar{g}} - (1-\beta)\frac{\partial \pi^d}{\partial \bar{g}}} \tag{22}$$

$$\frac{\partial g^a}{\partial e} = \left[-\frac{\partial \pi^a}{\partial e} + (1-\beta)\frac{\partial \pi^d}{\partial e}\right]\left[\frac{\partial \pi^a}{\partial \bar{g}} - (1-\beta)\frac{\partial \pi^d}{\partial \bar{g}}\right]^{-1} \tag{23}$$

The term in the denominators is positive: it is the difference between the derivatives of the profit from colluding and the profit from betraying. We are analyzing the collusive equilibrium, at which the profit from colluding is less negatively sloped (see figure 1). The profit from betrayal is positive, so $\frac{\partial g^a}{\partial \beta} < 0$. Multiplying it by n yields $\frac{\partial G}{\partial \beta} < 0$.

The sign of (23) is the sign of its numerator. Substituting the IC means that comparing both terms in the numerator is the same as comparing $\frac{\partial \pi^a}{\partial e}/\pi^a$ and $\frac{\partial \pi^d}{\partial e}/\pi^d$. This provides the main intuition for the channel through which enforcement may increase or decrease violence: whether it has a stronger impact on the profits of a complier or a deviator. Following this process is possible, but we will take a shortcut that will also allow us to determine the effect of the number of cartels and enforcement.

Note that solving the deviator's optimality problem yields optimal de-

[40] This is a particular case of a general theorem in Mas-Colell et al. (1995), chapter 12 appendix A, that states that any SGNE can be improved by using Nash reversion strategies.

viation $g^d = -(n-1)\bar{g} + \sqrt{\pi^A(n-1)\bar{g}}$. Substituting in the profit function for the deviator and then in the IC yields a quadratic equation with solution $\bar{g} = \pi^A \frac{-n(\beta-1)^2+n^2(\beta-1)^2+\beta-2\sqrt{n(n-1)(\beta-1)^2\beta}}{n(n+\beta-n\beta)^2}$. It is straightforward to check that $n\bar{g}$ is increasing in n for $\beta < \frac{n-1}{n}$, which proves that $\frac{\partial G}{\partial n} > 0$. This expression is also increasing in π^A, which means that the effect of enforcement goes in the same direction as the effect of enforcement on aggregate profits π^A. From equation (23), this also means that enforcement has a greater effect on the profits of deviating than on the profits of complying. Finally, this expression is decreasing in β which provides an alternative proof for $\frac{\partial G}{\partial \beta} < 0$.

Appendix B Second order conditions for the cartel

The cartel's problem is $\max_{(g_i,x_i)} \pi_i = p_c q(x_i, R_i(g_i, g_{-i}), e) - g_i - p_p x_i$, with first order conditions (4) and

$$p_c \frac{\partial q}{\partial R_i} \frac{\partial R_i}{\partial g_i} = 1 \tag{24}$$

The second-order conditions are $\frac{\partial^2 \pi_i}{\partial x_i^2} = p_c \frac{\partial^2 q_i}{\partial x_i^2} < 0$, $\quad \frac{\partial^2 \pi_i}{\partial g_i^2} = p_c \left[\frac{\partial q_i}{\partial R_i} \frac{\partial^2 R_i}{\partial g_i^2} + \frac{\partial^2 q_i}{\partial R_i^2} \left(\frac{\partial R_i}{\partial g_i} \right)^2 \right] < 0$, and $\frac{\partial^2 \pi_i}{\partial x_i^2} \frac{\partial^2 \pi_i}{\partial g_i^2} - \left(\frac{\partial^2 \pi_i}{\partial x_i \partial g_i} \right)^2 = p_c^2 \left[\frac{\partial^2 q_i}{\partial x_i^2} \frac{\partial q_i}{\partial R_i} \frac{\partial^2 R_i}{\partial g_i^2} + \left(\frac{\partial R_i}{\partial g_i} \right)^2 \left(\frac{\partial^2 q_i}{\partial x_i^2} \frac{\partial^2 q_i}{\partial R_i^2} - \left(\frac{\partial^2 q_i}{\partial x_i \partial R_i} \right)^2 \right) \right] > 0$. Strict concavity of R_i and concavity of q_i ensure that all three conditions are satisfied.

It still remains to show that (14) and (24) are equivalent. Homogeneity of degree one means that derivatives are homogeneous of degree zero, so $\frac{\partial q}{\partial R_i}$ is the same if it is evaluated at (x_i, R_i, e) or $(X, 1, e)$. Euler's theorem means that $Q = X \frac{\partial q(X,1,e)}{\partial X} + \frac{\partial q(X,1,e)}{\partial R}$, and from (5), $p_c \frac{\partial q}{\partial R_i} = p_c Q - p_p X = \pi^A$, so both first order conditions are indeed equivalent.

Appendix C Elasticity threshold for more general functions

Figure 4 plots the function $\hat{\epsilon}_c = -\left(1 + \frac{\sqrt{\gamma}}{2(1-\sqrt{\gamma})}\right)$. As the gap between consumer and producer prices widens, and γ goes down to zero, costs become a smaller share of revenues and the threshold gets closer to -1. For high γ, the threshold goes well below -1, but empirical evidence shows that each step in the production chain of drugs, from producers to consumers, implies a large increase in prices (**?**), and the results for high values of γ are therefore not very relevant.

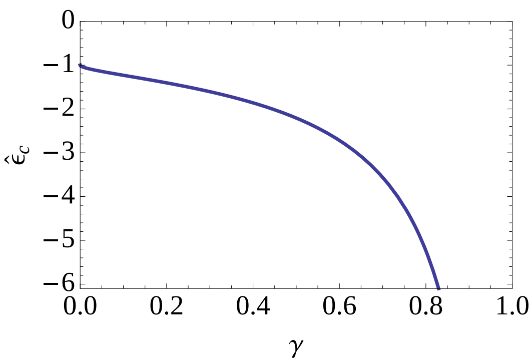


Figure 4: Variation of the threshold as a function of γ

Let's look now at more general functional forms. Suppose that w depends on the ratio of effective routes r to enforcement e. In order to allow for different efficiencies and increasing or decreasing returns to scale, we assume that w is a function of $\rho = \frac{r}{\varphi e^{\eta}}$: φ is a parameter that captures the relative efficiency of enforcement, and η is a parameter that captures whether the returns to scale of enforcement decrease faster than the returns to scale of effective routes. Thus, $w(r,e) = w(\rho)$. The conditions set on the derivatives of q in section 2 mean that $w' > 0$ and $w'' < 0$. This kind of function includes a variety of production technologies. For instance, if $w(\rho) = \rho^{1-\alpha}$ the production function is $q = e^{\eta(1-\alpha)} x^{\alpha} R^{1-\alpha}$, a Cobb-Douglas function, and the same CSF used in section 3.3 results if $w(\rho) = \frac{\rho}{1+\rho}$ with $\eta = 1$.

We will now show that such functions result in a correction that lowers the elasticity threshold $\hat{\epsilon}_c$. In terms of w, the threshold is $\hat{\epsilon}_c = -1 - \frac{\left(w - r\frac{\partial w}{\partial r}\right)^2}{r^2 \frac{\partial w}{\partial e}\frac{\partial^2 w}{\partial r^2}}\left(\frac{\frac{\partial w}{\partial e}}{w} - \frac{\frac{\partial w}{\partial e} - r\frac{\partial^2 w}{\partial r \partial e}}{w - r\frac{\partial w}{\partial r}}\right)$ The term in parentheses, which determines its sign, is now $\frac{rww''\rho_r\rho_e + rww'\rho_{re} - r(w')^2\rho_r\rho_e}{w(w - rw'\rho_r)}$. The denominator is positive, and by substituting the derivatives of ρ, its numerator is $-\rho ww'' - ww' + \rho(w')^2$, which is positive if

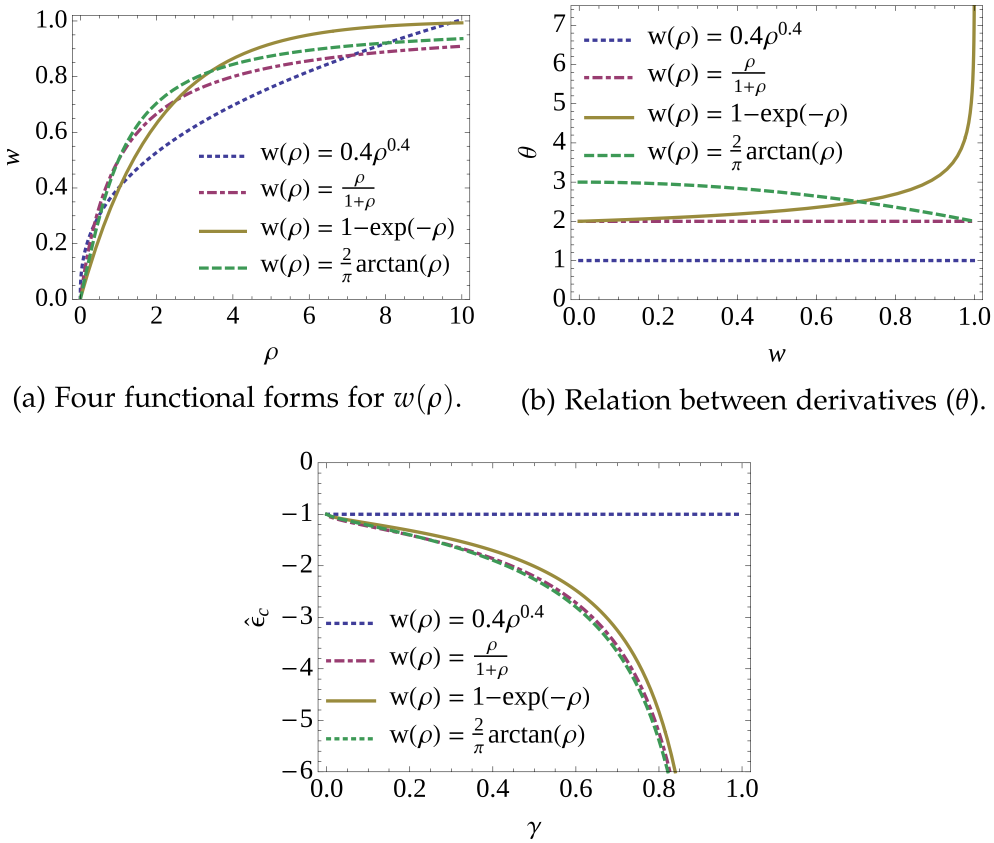
$$\theta = \frac{w''}{\frac{(w')^2}{w} - \frac{w'}{\rho}} > 1 \qquad (25)$$

The numerator is clearly negative, and the numerator is also negative since the conditions on w imply that $w > \rho w'$. If (25) is satisfied, the effect of enforcement on marginal productivity is greater than the effect on productivity, so the threshold is lower than -1. Condition (25) has the advantage that it is scale free: θ does not change by substituting $w(\rho)$ with $\hat{w}(\rho) = w(a\rho)$, where a is an arbitrary constant. It is also independent of η.

Setting $w_{CD} = \rho^{1-\alpha}$, a Cobb-Douglas technology, yields $\theta_{CD} = 1$. But as we argued in the main text, this is not a very reasonable form for w since it increases without bound. For it to be bounded above, given some value $w = w_{CD}$ and some value of $w' = w'_{CD}$, w'' should be less than for a Cobb-Douglas function ($w'' < w''_{CD}$) so that the function curves downward fast enough that it does not go past $w = 1$. This implies that $\theta > 1$. The relevance of θ being scale-free now becomes clear: the scale parameter a can be chosen so that $w = w_{CD}$ and $w' = w'_{CD}$, allowing comparison of θ and θ_{CD} only in terms of w'' and w''_{CD}.

Figure 5 illustrates our argument graphically with three functions that fulfill the conditions for $w(\rho)$:[41] $w = \frac{\rho}{1+\rho}$, $w = 1 - \exp(-\frac{1}{2}\rho)$, and $w = \frac{2}{\pi}\arctan\rho$. We also show $w = 0.4\rho^{0.4}$ for comparison. The particular values of the parameters were chosen so that the functions are relatively similar, although this does not change our conclusions. Figure 5a shows the general

[41] $w(0) = 0$, $w > 0$, $\lim_{\rho\to\infty} w(\rho) = 1$, $w' > 0$, and $w'' < 0$



(a) Four functional forms for $w(\rho)$.

(b) Relation between derivatives (θ).

(c) Elasticity threshold.

Figure 5: Comparison of different functional forms.

form of the functions. Figure 5b shows how θ behaves as a function of the value of w, and, in particular, that for all three functional forms $\theta > \theta_{CD}$. Finally, figure 5c shows the threshold that results for each functional form in terms of $\gamma = \frac{p_p}{p_c}$. Comparison with figure 4 shows that the conclusions from section 3.3 are not a peculiarity of the functional form that we chose for w.

Appendix D Variable prices in the producer market

In this section we relax the assumption that prices in the producer market are fixed. Since cartels are price takers, their individual behavior does not change in any way, and their maximization problem is the same, both in the SGNE and with repeated games. The comparative statics, however, must now take into account that changes in policy will have an effect in the producer market, thus changing p_p. This effect is described by the elasticity of supply ϵ_p.

D.1 Aggregate productive behavior

From proposition 2, the number of cartels has no effect on productive behavior, which means that it does not affect the amount of drugs bought from the producer region, and p_p. Thus, $\frac{\partial Q}{\partial n}$ stays the same. On the other hand, $\frac{\partial X}{\partial e}$ and $\frac{\partial X}{\partial e}$ do change. The analysis must now take into account that prices in producer markets are increasing in X, so marginal cost is increasing. The implicit function theorem yields the following expression, which replaces (8):

$$\frac{\partial X}{\partial e} = -\frac{\frac{\partial^2 q}{\partial X \partial e} + \frac{1}{Q\epsilon_c}\frac{\partial q}{\partial X}\frac{\partial q}{\partial e}}{\frac{1}{Q\epsilon_c}\left(\frac{\partial q}{\partial X}\right)^2 + \frac{\partial^2 q}{\partial X^2} - \frac{1}{X\epsilon_p}\frac{p_p}{p_c}} \tag{26}$$

The only change is a new term in the denominator, which does not change the sign, although the magnitude of the effect is less. From the chain rule, the new expression that replaces (9) is

$$\frac{\partial Q^e}{\partial e} = \frac{\frac{\partial^2 q}{\partial X^2}\frac{\partial q}{\partial e} - \frac{\partial q}{\partial X}\frac{\partial^2 q}{\partial X \partial e} - \frac{1}{X\epsilon_p}\frac{p_p}{p_c}\frac{\partial q}{\partial e}}{\frac{1}{Q\epsilon_c}\left(\frac{\partial q}{\partial X}\right)^2 + \frac{\partial^2 q}{\partial X^2} - \frac{1}{X\epsilon_p}\frac{p_p}{p_c}} \tag{27}$$

The sign of this expression does not change either. The comparative statics thus remains the same.

D.2 Threshold for the elasticity of demand

The effect of enforcement on violence depends on the effect it has on the aggregate productive profit. The new dependence of producer prices on quantities means that $\frac{\partial \pi^A}{\partial e} = \frac{\partial p_c}{\partial Q}\frac{\partial Q}{\partial e}Q + p_c\frac{\partial Q}{\partial e} - \frac{\partial p_p}{\partial X}\frac{\partial X}{\partial e}X - p_p\frac{\partial X}{\partial e}$. Rewriting $\frac{\partial p_c}{\partial Q}$ and $\frac{\partial p_p}{\partial X}$ in terms of elasticities leads to

$$\frac{\partial \pi^A}{\partial e} = p_c\left(1+\frac{1}{\epsilon_c}\right)\frac{\partial Q}{\partial e} - p_p\left(1+\frac{1}{\epsilon_p}\right)\frac{\partial X}{\partial e} \tag{28}$$

instead of (10). Substituting $\frac{\partial Q}{\partial e}$ and $\frac{\partial X}{\partial e}$ from (26) and (27) and isolating ϵ_c yields the following threshold for the elasticity of demand:

$$\hat{\epsilon}_c = -1 - \frac{\left(1+\frac{1}{\epsilon_p}\right)\left(\frac{\partial q}{\partial X}\right)^2}{\frac{\partial^2 q}{\partial X^2}\frac{\partial q}{\partial e} + \frac{1}{\epsilon_p}\frac{\partial q}{\partial X}\left(\frac{\partial^2 q}{\partial X \partial e} - \frac{1}{X}\frac{\partial q}{\partial e}\right)}\left(\frac{\frac{\partial q}{\partial e}}{Q} - \frac{\frac{\partial^2 q}{\partial X \partial e}}{\frac{\partial q}{\partial X}}\right) \tag{29}$$

Two new terms arise. First, the correction is smaller, since increasing marginal cost means that changes in X are smaller (the new term in the denominator)[42]. On the other hand, any change in X induces a larger change in costs, since p_p changes with X (see $\left(1+\frac{1}{\epsilon_p}\right)$ in the numerator). The sign of the correction is still determined by the sign of $\frac{\partial \log q}{\partial e} - \frac{\partial \log \frac{\partial q}{\partial X}}{\partial e}$.

Appendix E Response of violence to drug seizures

In this section we extend the analysis in Castillo et al. (2016) to show that, although violence in Mexico responded to cocaine supply changes from seizures in Colombia after Calderón's crackdown, the data shows no evidence of such response before December 2006. The results are consistent

[42]The sign of the correction could actually change if supply is very inelastic and $\frac{\partial^2 q}{\partial X \partial e} > \frac{1}{X}\frac{\partial q}{\partial e}$, but expanding this in terms of the derivatives of w shows that this would imply $\frac{\partial^2 w}{\partial e \partial r} > 0$.

with the idea that, in the absence of aggressive enforcement policies, repeated interaction creates incentives for traffickers to divide the business peacefully.

Using month-level time-series data on the Mexican homicide rate and cocaine seizures in Colombia, we estimate:

$$\begin{aligned} \ln h_t = & \alpha \ln S_t \times D_t + \beta \ln S_t \times (1 - D_t) + F_t(X_t; \gamma) \times D_t \\ & + F_t(X_t; \delta) \times (1 - D_t) + \epsilon_t \end{aligned} \quad (30)$$

where h_t is the homicide rate in Mexico in month t (as reported by the Mexican statistics agency, INEGI), S_t are cocaine seizures in Colombia (as reported by the Colombian Ministry of Defense), D_t is an indicator variable that takes a value of one from December 2006 onwards, and F_t includes (a) a cubic polynomial in t, (b) year dummies, and (c) time-varying controls X_t. If we were to assume that $E(\epsilon_t | S_t, D_t, F_t) = 0$, we could interpret α as the effect of cocaine seizures in Colombia on violence in Mexico prior to December 2006, and β as the effect of seizures in Colombia on violence in Mexico after December 2006. Castillo et al. (2016) explains in detail why this assumption might be justified; in brief, they contend that short-term (monthly) fluctuations in seizures are largely determined by chance. If this assumption were not justified, however, such that $E(\epsilon_t | S_t, D_t, F_t) \neq 0$, we would interpret α and β as the partial correlation between seizures in Colombia and violence in Mexico, conditional on F_t, before and after December 2006.

Table 1 shows our estimates of α and β. The first three columns show regressions for all of Mexico. The rest of the columns focus on municipalities in the first two quintiles of distance to the US, where the effect measured in Castillo et al. (2016) is strongest. In some specifications we control for the unemployment rate, by the Índice Global de la Actividad Económica, an early indicator of GDP that is computed every month, and for weather using dummies for the rainy and hurricane season in Mexico.

Table 1: Relationship between cocaine seizures in Colombia and violence in Mexico (Estimates of Equation 30)

	All of Mexico			Quintiles 1 and 2			First quintile		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Dependent variable: Log homicide rate.								
Before Dec. 2006 (α)	-0.015	-0.017*	-0.019**	-0.009	-0.010	-0.013	-0.000	-0.003	-0.019
	(0.012)	(0.010)	(0.010)	(0.015)	(0.013)	(0.012)	(0.019)	(0.019)	(0.015)
Dec. 2006 onward (β)	0.059**	0.056**	0.050*	0.082**	0.079**	0.073**	0.119***	0.120***	0.109***
	(0.027)	(0.025)	(0.026)	(0.033)	(0.032)	(0.034)	(0.033)	(0.034)	(0.037)
Dif ($\beta - \alpha$)	0.073**	0.074***	0.069**	0.091**	0.089***	0.086**	0.119***	0.122***	0.128***
	(0.030)	(0.027)	(0.028)	(0.036)	(0.034)	(0.036)	(0.038)	(0.039)	(0.040)
Observations	96	96	96	96	96	96	96	96	96
R-squared	0.957	0.964	0.967	0.957	0.963	0.965	0.964	0.965	0.968
Controls:									
Unemployment rate		✓	✓		✓	✓		✓	✓
Economic activity		✓	✓		✓	✓		✓	✓
Weather			✓			✓			✓

Note: Errors are robust to heteroskedasticity. Coefficients with *** are significant at the 1% level, with ** at the 5% level and with * at the 10% level.

The estimates for β simply replicate the findings in Castillo et al. (2016): After December 2006, cocaine seizures in Colombia caused an increase in homicide rates, especially in municipalities close to the U.S. border. The estimates for α reveal a different pattern before Calderón's term: no strong relationship between cocaine seizures in Colombia and violence in Mexico.

These results are consistent with the predictions of our theoretical model. Before Calderón began aggressively targeting kingpins, cartels operated in a peaceful equilibrium—and, therefore, violence did not respond to supply shocks. Calderón's crackdown broke the peace treaties, pushing Mexico into a new equilibrium in which violence responded to supply shocks. While the data themselves do not identify the timing of the switch from one equilibrium to the other (instead, we select December 2006 as a cutoff, since that is when Calderón took office), the contrast between the pre- and post- periods suggests that our model helps explain violence in Mexico during this period.