Should drug policy aim at the fragmentation of cartels? Breaking down a peaceful equilibrium*

Juan Camilo Castillo†

April 9, 2013

Abstract

Violence increases when governments achieve their objective of fragmenting drug-trafficking organizations (DTOs). This unintended consequence of successful policy has been observed during the last decade in Colombia and Mexico, to name two recent examples. In this work I provide a theoretical framework to understand this behavior. Drawing elements from industrial organization, I model DTOs as firms that collude by not attacking each other in order to obtain larger profits from the drug trade. Profit-maximizing DTOs always collude, which means that previous works analyzing a static Nash equilibrium miss an important part of DTOs' behavior. I show that a peaceful equilibrium arises if there are only a few DTOs that care enough about the future. Policies resulting either in a larger number of DTOs or in more impatient leaders increase war between DTOs without any supply reduction. On the other hand, policies that focus on drug seizures, without directly attacking cartel leaders, are much more desirable since they can simultaneously reduce supply and decrease violence between DTOs.

1 Introduction

The illegal drug trade has shown an exceptional capability to transform itself according to the conditions it must face. Governments have tried to attack it with various methods, only resulting in traffickers finding new ways to operate, both locally and globally.

*This paper is a draft of my thesis for the Master's Degree in Economics (PEG) at Universidad de los Andes. I am especially grateful to my advisor, Daniel Mejía (Economics Department, Universidad de los Andes, e-mail: dmejia@uniandes.edu.co). I would also like to thank Leopoldo Fergusson, Ana María Ibáñez, and my fellow students at the Masters Thesis Seminar for their valuable comments.

†Economics Department, Universidad de los Andes, e-mail: jc.castillo126@uniandes.edu.co

Whenever authorities are able to eliminate one major trafficking route, a new one arises to replace it, as when the U.S. government shut down the Caribbean route from Andean nations in the mid-1980s, displacing traffic to the new route through Mexico and Central America that today still supplies the majority of cocaine to North America. And whenever one form of organization is effectively suppressed, new types of cartels arise, such as when the Colombian armed forces defeated the Medellín and Cali cartels, both fond of media exposure and political influence. The void of power these two cartels left behind spawned a new generation of low-profile cartel leaders that forged alliances with previously existing armed groups like the guerrillas and the paramilitaries.

Among the various mutations drug trafficking organizations (DTOs) have gone through, a perplexing one has happened when local governments turn to strategies that end up decentralizing control of the drug trade. The idea behind such strategies, based on the premises of the war on drugs led by the U.S. government since Richard Nixon's declaration in 1971, is that the illegal drug trade cannot be eliminated for good if large, powerful DTOs persist. The outcome has been, however, quite different from what governments wanted: drug trade continues, led by former bosses' lieutenants, with an important increase in violence as an unintended consequence. The surge in violence is driven by a large increase in the number of drug traffickers killed by other drug traffickers. A chaotic state with a void of power arises, in which DTOs increase the intensity of fights between each other, and in which members within each cartel fight to become the new leaders.

Some clear examples of this phenomenon come to mind. In the first half of the 2000s, a few Mexican cartels increased their dominance of the illegal drug trade, smuggling cocaine, marijuana, methamphetamine and heroin into the U.S. Their huge profits gave them, through briberies and threats, control of many aspects of society in the regions that were most important for the drug trade. Surprisingly, the levels of violence were low in comparison with other Latin American nations, despite DTOs having gained an amount of power that Mexicans were not willing to accept. This resulted in president Felipe Calderón being elected in 2006 after a campaign centered on the promise of frontal war against DTOs. Keeping his word, he started his term with large-scale military actions that continued until the end of his sexennium. Whereas the amount of cocaine crossing the U.S. border did not change significantly, violence started to increase year after year, to the point that the homicide rate in 2010 was more that twice the homicide rate in 2006.

Another example could be seen in some regions in Colombia. During the early 1990s some paramilitary groups had formed in order to defend populations from leftist guerrillas like the FARC[1] and the ELN[2]. These groups started forging alliances with drug traffickers, which ended up in paramilitaries being fused with former DTOs. Their peak of power came after they united under the single leadership of Carlos Castaño in the late 1990s under the name of the AUC[3], which came to exert a great amount of power in some regions of the country. Even though they still led gruesome battles against the FARC and ELN, the regions under their full control experienced a degree of peacefulness seen in few regions in the country at the time. The AUC ended around 2005, when they agreed to demobilize in a treaty with president Álvaro Uribe's government. Just as in Mexico, the subsequent void of power did not lead to a decrease in the amount of drug produced. Instead, multiple small bands emerged to fill up the position of control formerly held by the AUC. These groups, called bacrims (short for criminal bands in Spanish) have led multiple fights over the control of routes, breaking the previous state of calm.

Multiple observers have described this kind of behavior(Guerrero (2011); Castillo et al. (2013); Camacho (2009, 2011)). However, it has not been described satisfactorily from the economic theory of conflicts. The purpose of this work is thus to provide an explanation for the behavior that follows policies such as deliberate fragmentation of cartels and direct attacks on their leaders, and why this leads to an important increase in violence without reducing supply. With this in mind, I build a theoretical framework to understand those instances. Fragmentation can happen through two mechanisms: attacking leaders, after which cartels often split into a number of smaller groups, or by negotiating with a dominant group and convincing it to demobilize, leaving a void of power that is subsequently filled by a number of smaller DTOs. I explain why DTOs operate under two different states. In the first one, a few organizations control the whole drug trade in a region under relative calm. In the second one, there is a large number of DTOs, each one controlling a small amount of the drug trade, and they wage war among themselves in order to increase the share of the drug that each one of them holds.

In the first state DTOs collude, following a tacit treaty requiring them not to attack each other, while each one controls a given fraction of the drug trade. They are thus able to receive larger profits, since they do not have to spend resources in the conflict, and they have no losses from dead personnel and destruction caused by engaging others. However, just as in the theory of collusion in industrial organization, every individual

[1]Fuerzas Armadas Revolucionarias de Colombia
[2]Ejército de Liberación Nacional
[3]Autodefensas Unidas de Colombia

DTO will have incentives to betray others by attacking them in order to seize a larger portion of the drug trade, taking advantage that other DTOs are not expecting a betrayal. Thus, a peaceful equilibrium can only be sustained if DTOs face each other in repeated interactions, and if they care enough about the future that they prefer the collusive equilibrium to hold for a long period. If this is not the case, DTOs attack each other. This is the second state, in which war erupts, and DTOs fight for the control of the drug trade. The number of cartels present is essential to determine whether they end up in a peaceful equilibrium. Drawing again from the theory of industrial organization, peace is much easier to sustain if the number of cartels is low.

I also discuss the conditions that must be met in order for the peaceful equilibrium to be sustained, and if there is war, I show what determines the level of violence. For instance, the international price of drugs or the efficiency of governments in controlling cartels can be important in determining whether DTOs decide to fight. Furthermore, with the theoretical basis provided by the model, I show what policies can trigger the breakdown of the peaceful equilibrium. For example, I show that as the Colombian government succeeded in signing a demobilization treaty with the AUC, they drastically increased the number of DTOs present, breaking down the peaceful equilibrium and increasing violence. I also show that the Mexican government's strategy of beheading DTOs led to bosses being more short-sighted, and to an increase in the number of independent DTOs, which induced brutal wars between them.

Finally, I focus on two alternative approaches that have been followed by governments. The first one is centered on enforcement, by encouraging government forces to seize as much drugs as they can. The alternative approach is to attack the structures of drug cartels frontally, as Calderón did in Mexico. I analyze how effective each policy is, in terms of supply reduction, and comparing the level of violence between DTOs attained under each approach. Previous works had mostly focused on supply reduction (see section 2). Thus, mine is a more global assessment, since I evaluate policies in terms of their effect both on consumer nations (supply reduction) and on trafficking nations (violence between DTOs). Additionally, many analyses had compared the efficiency of government actions at different stages of the production chain of drugs until they reach final consumers, but no study had compared enforcement activities with direct attacks against cartel bosses. The results from this comparison will therefore be valuable material for governments willing to reassess the widely believed notion that bosses should be engaged frontally. The main conclusion is that the most effective policy against illegal drugs, if they are to remain illegal, is enforcement focused on seizures, both in terms of its impact on trafficking and consumer nations.

2 Literature Review

I will now summarize some of the relevant contributions to the economical theory of crime and conflict. A conflict is a situation in which a number of actors engage in a zero-sum game with the aim of obtaining some prize by investing some effort or resources[4]. Crime therefore involves conflict between the criminal and the victim. If the crime is stealing, the prize is the object that is being stolen, and if the crime is murder, the prize is the life of the victim. There is also a conflict between the criminal and governmental forces, in which the prize is whether the criminal succeeds in perpetuating the crime.

The idea of modeling conflicts economically[5] arose around the 1950s, with some early works such as Haavelmo (1954). Initially, the approach was simply to consider the idea of appropriating other's belongings as an imperfection of the markets, in which people devote resources that would otherwise be used as factors of production. Such early works also consider appropriation as a contradiction to well-defined property rights, which is an essential element of most economic models.

It would take some time before the first attempts to model crime explicitly, which take into account damage brought to society and the possible intervention by governments in order to curb crime. The first such works came in the 1960s, such as Becker (1968), who considers profit maximizing criminals, as well as a government which must spend resources in enforcement and punishment in order to reduce crime. Many other works have come since then, with a special flourishing of the literature in the last 20 years. One of the unifying concepts is the *conflict technology*, or *contest-success function*, which describes the outcome of a conflict as a function of the resources spent on it by opposing sides. Many works related to crime and appropriation have used such functions, like Grossman and Kim (1995), Grossman (1999), and Noh (2002). Hirshleifer (1989) analyzed the properties of conflict-success functions, which are widely used in this work, and the implications of assuming some specific forms for them.

Within the economic theory of conflict, some works have focused on illegal drug markets, their effect on society, and actions that governments can take against them. Becker et al. (2006) make one of the main contributions. They suggest that the government has basically two means to reduce consumption of drugs: enforcement activities, which involve catching and punishing drug dealers, and setting high taxes on drugs (which inevitably implies legalizing them). Their work leads to two bold conclusions.

[4]Note that this definition of conflict is different from the much more specific idea of an armed conflict, broadly used in political science, that involves a number of groups fighting for political control of some territory or nation.

[5]Garfinkel and Skaperdas (2007) provide a good survey of literature on the economics of conflict.

First, they argue that the effectiveness of enforcement activities depends very strongly on the elasticity of demand for drugs: if demand is inelastic, as empirical evidence suggests it is, enforcement turns out to be very ineffective, meaning that very large expenditures only cause a small reduction in the deleterious effect drugs have on society. This happens because decreasing the amount of drugs sold actually increases the market size, measured as the price times the quantity. Second, they say that, from an economic point of view, it is *always* better to control illegal drug markets by taxing them heavily instead of prohibiting them. The reasoning behind this is that any level of enforcement can be replicated with high enough taxes. Although the government will have to invest resources in order to control the black markets that arise, they will be smaller than the original illegal market, and therefore less costly to control. The taxation solution will additionally benefit society because of the collection of taxes that can be significantly large.

Although Becker et al. (2006) analyze in detail the market for drugs, their analysis is best suited to the retail market. Other works have focused on dealing with the whole chain of production, which starts with crops that are used to produce drugs, then goes through drug-trafficking markets, and finally ends in the main consumers in the U.S. and Europe. Some of these works have focused on cocaine-producing Andean nations, particularly Colombia, and drug-trafficking countries like Mexico and other Central American nations. Grossman and Mejía (2008) build a model in order to compare the relative efficiency of governmental intervention with two different strategies against the cocaine production chain: first, control of land where coca plant is being cultivated, and second, eradication of coca plants and interdiction of the produce of coca crops. A more complete model is analyzed in Mejía and Restrepo (2008), which includes conflict over the control of arable land and conflict over the control of routes. The idea is to evaluate Plan Colombia, led by the U.S. government in order to subsidize the war against drug producers and traffickers in Colombia. They conclude that resources would have been better spent if more efforts from Plan Colombia had targeted the conflict against drug traffickers, and not against drug producers. Chumacero (2008) builds a general equilibrium model that includes the production, trafficking and consumption stages of the illegal trade, showing that stronger penalties on producers or traffickers can actually improve their situation. Mejía and Restrepo (2011) analyze the combination of efforts to fight illegal drugs in producer and consumer countries, with the finding that they are complementary.

The previous works analyze the drug trade in terms of the amount of drugs supplied and how policies can decrease it, but they say little about the high levels of violence caused in producer and trafficking nations. I therefore take a broader view, by also looking at how different policies increase or reduce bloodshed in trafficking nations. These works also follow Becker et al. (2006) and assume that demand for drugs is inelastic, which implies that the size of the drug market increases if governments succeed in decreasing supply. Competing DTOs then fight for higher stakes, resulting in higher levels of violence. This looks like a possible explanation for the Colombian and Mexican cases I mentioned, in which successful policies result in more violence. However, this contradicts the fact that no substantial decrease in supply was seen in either case. Thus, there is no adequate theory to explain these cases. Additionally, existing models have failed to consider multiple-period interaction between opposing sides beyond being leaders and followers *à la* Stackelberg. Thus, I attempt to fill this void by modeling DTOs as agents interacting repeatedly, which opens the possibility of governments succeeding in killing leaders and fragmenting DTOs, while achieving no noticeable decrease in supply and inducing an increase in violence.

The present work relies on industrial organization, by modeling drug traffickers in a region as an industry with barriers to entry, in which firms (DTOs) may choose, if it is in their own interest, to collude instead of engaging in free competition, i.e., war. The notion of a market with collusion is based on standard references such as Tirole (1988) and Motta (2004). The novelty is the use of such widespread models from industrial organization to the specific case of the drug trade. This goes in line with some other recent works that have sought to explain DTOs as operating in a complex industry that shares many characteristics with traditional industries, such as Baccara and Bar-Isaac (2008) or Bardey et al. (2013).

3 The trafficking industry

Consider drugs being trafficked from a producer market into a consumer market through some territory. There are n DTOs that participate in drug trafficking through the intermediate territory. I consider DTOs to be pure drug traffickers, whose sole purpose is to maximize their profit, without any craving for reputation or political control.[6] Due to the presence of significant barriers to entry, the number of DTOs is constant. The main barrier to entry is the power held by DTOs: if any small actor tries to enter the market, incumbent DTOs will attack them in order to suppress them. I will denote the

[6]This is certainly a strong assumption, as many DTOs have clearly shown, such as the Colombian cartels or guerrillas. However, my purpose is to model their trafficking behavior, and modeling their political behavior would make the task much more complicated, obscuring the main results.

set of indices for the DTOs by I. The DTO labeled by the index $i \in I$ uses two factors in order to provide drugs in the consumer market: an amount x_i of drugs bought in the producer market at a price p_p, and routes through the intermediate territory, R_i. The DTOs in the region fight over the control of a continuum of routes with mass normalized to one, i.e., $\sum_{i \in I} R_i = 1$.

The government in the territory is a non-strategic actor in this model, meaning that it acts first, choosing the amount of resources it spends in enforcement, e. The DTOs then interact strategically, and they choose their course of action based on the level of enforcement chosen by the government. The amount of drugs bought in the producer market, the amount of routes, and the level of enforcement are put together in a production technology that results in an amount $q_i(x_i, R_i, e)$ of drugs reaching the consumer market. The production function is increasing in the amount of drugs bought in the producing market and in the amount of routes held, and it is decreasing in the amount spent by the government in enforcement activities ($\frac{\partial q}{\partial x_i} > 0$, $\frac{\partial q}{\partial R_i} > 0$, and $\frac{\partial q}{\partial e} < 0$). Additionally, the marginal productivity of both factors of production is decreasing ($\frac{\partial^2 q}{\partial x_i^2} < 0$ and $\frac{\partial^2 q}{\partial R_i^2} < 0$), and the production function is concave in (x_i, R_i)[7], which will be necessary for the maximization problem of the DTO.

Any increase in enforcement by the government decreases the marginal productivity of both factors of production. This is a reasonable assumption: if routes are better watched a lower fraction of the drugs bought at the producer market will reach the final market, displacing the whole production function down by some proportionality factor, which results in a decrease in both marginal productivities. Formally, this can be seen as $\frac{\partial^2 q_i}{\partial e \partial x_i} = \frac{\partial}{\partial e}\left(\frac{\partial q_i}{\partial x_i}\right) < 0$ and $\frac{\partial^2 q_i}{\partial e \partial R_i} = \frac{\partial}{\partial e}\left(\frac{\partial q_i}{\partial R_i}\right) < 0$.

I will also assume that the production technology has constant returns to scale. This means that the function q_i is homogeneous of degree one in (x_i, R_i). Alternatively, this can be analyzed by means of a function $w_i(x_i, R_i, e)$ such that $q_i = w_i x_i$. In this case, w is the fraction of the drugs bought in the producer market that in the end reaches the consumer market. As the proportion $r_i = \frac{R_i}{x_i}$ decreases, the routes used by the DTO become saturated, and the effectiveness of the government in preventing drugs from reaching their destination increases ($\frac{\partial w_i}{\partial r_i} > 0, \frac{\partial w_i}{\partial e} < 0$). The only way for the DTO to maintain the fraction of drugs that reach the consumer market as they take a larger

[7]Although this seems to be a rather arbitrary assumption, only having the purpose of enabling the existence of an equilibrium, it is actually a consequence of the other conditions imposed on the function. 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. We will also require the function to have constant returns to scale. Quasiconcavity and constant returns to scale imply concavity.

amount of drugs is by transporting them through a larger number of routes. Thus, it is reasonable to assume that w_i depends only on r_i, without depending directly on either x_i or R_i. This is the same as saying that it is homogeneous of degree zero in (x_i, R_i), which implies that q_i is homogeneous of degree one. The function w_i is such that the marginal productivity of r_i is decreasing, i.e., $\frac{\partial^2 w_i}{\partial r_i^2}$. Otherwise, as the routes become less saturated the fraction of the drugs reaching the destination would increase without bound, at some point surpassing one. This fact has as an interesting consequence that $\frac{\partial^2 q_i}{\partial x_i \partial R_i} > 0$, meaning that with this technology routes and drugs are complementary production factors[8].

In order to obtain routes, the DTO invests an amount g_i in the conflict for routes, which includes the salaries of gunmen, the cost of guns, losses associated with dead gunmen, etc. At the end of the conflict, the amount of routes held by the DTO is a function $R_i(g_i, g_{-i})$ that depends positively on its expenditure and negatively on the total amount $g_{-i} = \sum_{j \neq i} g_j$ spent by all other DTOs, that is, $\frac{\partial R_i}{\partial g_i} > 0$ and $\frac{\partial R_i}{\partial g_{-i}} < 0$. The marginal productivity of the expenditure in the conflict is decreasing ($\frac{\partial^2 R_i}{\partial g_i^2} < 0$), meaning that holding a large fraction of routes requires spending a lot of resources. As g_{-i} increases, any additional investment by DTO i is less in comparison with the size of the conflict, reducing the marginal productivity of g_i. Thus, $\frac{\partial}{\partial g^N}\left(\frac{\partial \hat{R}_i}{\partial g_i}\right) = \frac{\partial^2 \hat{R}_i}{\partial g^N \partial g_i} < 0$. I will finally assume that $\frac{\partial^2 R_i}{\partial g_i^2} < \frac{\partial^2 R_i}{\partial g_i \partial g_{-i}}$. In order to see why it is a reasonable assumption, consider $\frac{\partial R_i}{\partial g_i}$, the marginal productivity of g_i on the conflict for routes. This marginal productivity decreases both with an increase in g_i, expenditure by the same DTO, and in g_{-i}, expenditure by the other DTOs. However, if both increases are equal, the increase in g_{-i} would be spread across all other DTOs, so it is reasonable to assume that it has a milder effect on the marginal productivity. This can be stated mathematically as $\frac{\partial^2 R_i}{\partial g_i^2} < \frac{\partial^2 R_i}{\partial g_i \partial g_{-i}}$.

In the end, the DTO sells an amount q_i of drugs in the consumer market at a price p_c. Thus, the profit it obtains is given by:

$$\pi_i = p_c q_i(x_i, R_i(g_i, g_{-i}), e) - g_i - p_p x_i \tag{1}$$

[8]In order to see this, first note that $\frac{\partial^2 q_i}{\partial x_i \partial R_i} = x \frac{\partial^2 w_i}{\partial x_i \partial R_i} + \frac{\partial w_i}{\partial R_i}$. By using the chain rule, $\frac{\partial w_i}{\partial R_i} = \frac{\partial w_i}{\partial r_i} \frac{\partial r_i}{\partial R_i}$ and $\frac{\partial^2 w_i}{\partial x_i \partial R_i} = \frac{\partial^2 w_i}{\partial r_i^2} \frac{\partial r_i}{\partial R_i} \frac{\partial r_i}{\partial x_i} + \frac{\partial w_i}{\partial r_i} \frac{\partial^2 r_i}{\partial x_i \partial R_i}$. The derivatives of r_i can be readily calculated: $\frac{\partial r_i}{\partial R_i} = \frac{1}{x_i}$, $\frac{\partial r_i}{\partial x_i} = -\frac{R}{x_i^2}$, and $\frac{\partial^2 r_i}{\partial x_i \partial R_i} = -\frac{1}{x_i^2}$. Substituting everything in the initial expression for the cross derivative of q_i yields $\frac{\partial^2 q_i}{\partial x_i \partial R_i} = -\frac{R_i}{x_i^2} \frac{\partial^2 w_i}{\partial r_i^2}$, which is positive due to the decreasing marginal productivity of r_i.

Profit maximizing behavior will be soon described, but let us first define the relevant variables that allow the comparison of different equilibria. Let capital letters without a subscript denote aggregate quantities: $G = \sum_{i \in I} g_i$, $X = \sum_{i \in I} x_i$, $R = \sum_{i \in I} R_i$, and $Q = \sum_{i \in I} q_i$. The total amount of drug supplied to the consumer region is therefore Q, the sum of the quantity supplied by each DTO. Thus, it is in the best interest of the consumer region to reduce Q. On the other hand, I will use aggregate expenditure in the conflict as a proxy for the level of violence in the trafficking region. This means that the trafficking region would like G to be as low as possible.

3.1 The individual DTO's problem

Consider an individual DTO that responds to other DTO's behavior. This will be the basis for competitive behavior. In this case, the DTO observes others' behavior (i.e., their expenditure in the conflict g_{-i}), and, based on that, maximizes its profits. Thus, the problem it faces is

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

The first-order conditions for this problem are

$$p_c \frac{\partial q_i}{\partial x_i} = p_p \tag{3a}$$

$$p_c \frac{\partial q_i}{\partial R_i} \frac{\partial R_i}{\partial g_i} = 1 \tag{3b}$$

Both conditions are easy to interpret: (3a) states that the marginal productivity of the drugs bought in the producer market must equal their marginal cost (the price in the producer market), and (3b) states that the marginal productivity of the investment in the conflict must equal its marginal cost (one). The concavity of both q_i and R_i ensures that this solution is indeed a maximum[9], and the properties set upon the first and second derivatives of the functions mean that this problem leads to an interior solution. The equilibrium quantities give the best-response functions for q_i and g_i in terms of the quantities g_{-i} chosen by all other DTOs.

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

3.2 One-period Nash equilibrium

DTOs do not care about the future if they interact for a single period. That means that they maximize their instantaneous profit given the expenditure by all other DTOs. The solution is a Nash equilibrium, in which the response functions given by the first-order conditions (3) for all DTOs are fulfilled simultaneously. Considering all DTOs to be equal, the symmetry of the problem means that $g_i = g_j = g^N, \quad x_i = x_j = x^N \quad \forall i, j \in I$ (the superscript refers to the solution being the one-period Nash equilibrium). The comparative statics on the equilibrium amounts can be found from the total differential of both first-order conditions (see appendix A). The following proposition summarizes the most relevant results:

Proposition 1. *Under a symmetric competitive equilibrium, the comparative statics on each DTO's expenditure in the conflict and the amount of drugs taken to the consumer market by each DTO are as follows:*

- $\frac{\partial g^N}{\partial e} < 0, \frac{\partial q^N}{\partial e} < 0$*: A greater expenditure by the government in enforcement reduces individual expenditure in the conflict and the amount of drugs taken to the final market by each DTO.*
- $\frac{\partial g^N}{\partial n} < 0, \frac{\partial q^N}{\partial n} < 0$*: A greater number of DTOs in the territory under study decreases individual expenditure in the conflict and the amount of drugs taken to the final market by each DTO.*

Proof. Looking at the signs of all single derivatives in the expressions on the right-hand side of equations (27a)-(28b) and (31) (in appendix A), it is easy to check these results. □

The mechanisms behind proposition 1 are easy to see. Enforcement reduces the marginal productivity of both factors of production, which reduces the input to each factor of production. This means that each DTO invests less in the conflict. Less factors of production, combined with decreased productivity, mean that each DTO takes less drugs to the consumer market. On the other hand, a greater number of cartels decreases the fraction of routes each one of them holds, which decreases the marginal productivity of drugs bought in the producer market. Therefore, each DTO decides to buy less drugs, in turn decreasing the marginal productivity of routes. This results in each DTO investing less resources in the conflict in order to obtain routes.

These results are interesting, but the most relevant results are the comparative statics on the aggregate quantities Q and G. In order to analyze the aggregate amounts in equilibrium, I will first show that the aggregate productive behavior of DTOs can be isolated from the conflict between DTOs. By the productive behavior I mean the amount of drugs they buy in the initial market and the amount they sell in the consumer market.

In any symmetric equilibrium each DTO holds a fraction $\frac{1}{n}$ of the routes. The function q_i depends on x_i, R_i, and e, and the amount of routes R_i is now fixed, so the amount of drugs that reach the final market now depends only on the amount bought and the level of enforcement. Since q_i is homogeneous of degree one in (x_i, R_i), this amount is $Q(X, R, e)|_{R=1} = nq_i(x_i, R_i, e) = q_i(nx_i, nR_i, e) = q_i(X, 1, e)$. Homogeneity of degree one means that the derivatives of q_i are homogeneous of degree zero, so $\frac{\partial q_i(x_i,R_i,e)}{\partial x_i} = \frac{\partial q_i(X,1,e)}{\partial x_i} = \left.\frac{\partial Q(X,R,e)}{\partial X}\right|_{R=1}$. Substitution in (3a) yields

$$p_c \frac{\partial Q}{\partial X} = p_p \tag{4}$$

This result has a straightforward interpretation: *the aggregate productive behavior is the same as if there were a single DTO*. Thus, the number of cartels has no effect on the total amount of drugs being taken to the final consumer market. The effect of enforcement is also easy to determine intuitively. Since the marginal productivity of drugs bought in the initial market decreases, this must be compensated by a decrease in X (which increases the marginal productivity) for (4) to hold. Since X decreases and e increases, Q clearly decreases.

The effect on the level of violence is more complicated to determine, so we leave it to appendix A, which also proves the comparative statics on Q^N formally. The following proposition summarizes the results:

Proposition 2. *Under a symmetric competitive equilibrium, the comparative statics on the level of violence in the region and the total amount of drugs taken to the consumer market are as follows:*

- $\frac{\partial G^N}{\partial e} < 0, \frac{\partial Q^N}{\partial e} < 0$: *A greater expenditure by the government in enforcement reduces the level of violence and the total amount of drugs reaching the final market.*
- $\frac{\partial Q^N}{\partial n} = 0$: *The number of DTOs has no effect on the total amount of drugs reaching the final market.*
- $\frac{\partial G^N}{\partial n} > 0$: *An increase in the number of cartels increases the level of violence.*

Proof. Appendix A demonstrates these results. □

So far, the most important implication of this model is that the aggregate productive behavior does not depend on the conflict. This is a consequence of the fact that the conflict is a zero-sum game, which in the end results in all DTOs holding the same fraction of routes. In terms of public policy, this means that actions taken by the government to affect the conflict between cartels will affect violence, but they will have no effect on the amount of drugs reaching final consumer markets. Therefore, all actions taken by the government should be aimed against productive behavior (i.e., enforcement) if they are to reduce the supply of drugs in consumer markets.

The level of violence, on the contrary, depends on many other factors. On the one hand, it decreases with enforcement since it reduces the marginal productivity of routes, which in turn means that DTOs decide to invest less in the conflict. Following Becker et al. (2006), this result holds since I am assuming exogenous prices, but it should change if DTOs held an important share of the final consumer market, whose demand is inelastic. In that case, decreasing the amount of drugs would actually increases the size of the prize, thus increasing violence. On the other hand, violence increases as the number of DTOs increases. This depends on the assumption that the resources spent in the conflict have significantly diminishing returns to scale. Thus, if there are more cartels, they spend individually less in the conflict, which induces an increase in their marginal productivity. The response is an increase in their individual expenditure that partially offsets the decrease due to fragmentation, resulting in an increased overall level of violence.

3.2.1 A particular functional form

In order to analyze this problem with a concrete example, I will make assumptions about the particular functional form of q_i and R_i. This will allow me to show that the comparative statics from proposition 2 hold, and I will be able to show some numerical results. Let the amount of routes that each DTO controls at the end of the conflict be given by the following contest-success function:

$$R_i(g_i, g_{-i}) = \frac{g_i}{\sum_{i \in I} g_i} \tag{5}$$

This means that the unit mass of routes is distributed among the DTOs in proportion to the amount invested in the conflict by each one of them. An implicit assumption is that all cartels are equally efficient in the conflict. It is easy to check that this type of function fulfills both conditions imposed before, i.e., it is increasing in g_i, and it is concave[10]. Furthermore, this type of contest-success function has been widely used in

[10]More specifically, both conditions are fulfilled as long as $\exists j \in I, j \neq i$ such that $g_j \neq 0$

the literature, for instance in Grossman and Kim (1995), Grossman (1999), Noh (2002), Grossman and Mejía (2008).

Once the DTO controls an amount R_i of routes and attempts to transport an amount x_i of drugs into the consumer market, it engages the government forces in a second conflict. After that conflict, the fraction of the drugs that reach the consumer market is given by

$$w(r_i, e) = \frac{r_i}{r_i + \varphi e} \tag{6}$$

which again has the widely-used form described for the contest-success function from (5). Since w is to be homogeneous of degree zero, it depends only on $r_i = \dfrac{R_i}{x_i}$, a measure of the availability of routes per unit of drugs being trafficked. The parameter φ measures the relative efficiency of the government interdiction activities compared with the efficiency of the DTO.

This function can be multiplied by x_i to give the following drug production function:

$$q_i(x_i, R_i, e) = x_i \frac{R_i}{R_i + \varphi x_i e} \tag{7}$$

It is straightforward to check that this function is homogeneous of degree one, increasing both in x_i and R_i, and concave in (x_i, R_i), the conditions needed for the existence of a maximum.

With these functional forms, the first-order conditions (3a) and (3b) are

$$p_c \frac{g_i^2}{(g_i + \varphi x_i e G)^2} = p_p \tag{8a}$$

$$p_c \frac{\varphi q_i^2 e g_{-i}}{(g_i + \varphi x_i e G)^2} = 1 \tag{8b}$$

where $G = \sum_{i \in I} g_i$. I can now solve for both g^N and x^N in the symmetric case, as well as for all other variables in the model[11]:

$$x^N = \frac{\gamma - 1}{\varphi e n} \qquad g^N = \frac{p_p(n-1)(\gamma-1)^2}{\varphi e n^2} \tag{9a}$$

$$R^N = \frac{1}{n} \qquad w^N = \frac{1}{\gamma} \qquad q^N = \frac{\gamma - 1}{\gamma}\frac{1}{\varphi e n} \tag{9b}$$

$$\pi^N = \frac{\gamma - 1}{\phi e n}\left[\frac{p_c}{\gamma} - ps\frac{\gamma n - \gamma + 1}{n}\right] \tag{9c}$$

$$Q^N = \frac{\gamma - 1}{\gamma}\frac{1}{\varphi e} \qquad G^N = \frac{p_p(n-1)(\gamma-1)^2}{\varphi e n} \tag{9d}$$

where $\gamma = \sqrt{\frac{p_c}{p_p}}$. It is straightforward to check that all results from proposition 2 hold.

3.3 Repetitive interaction and collusion

I will now consider the same situation described before, but being repeated for multiple periods. The total profits obtained by a DTO is the discounted sum of the profit obtained in each of the periods, namely

$$\Pi_i = \sum_{t=0}^{\infty} \beta^t \pi_{i,t} \tag{10}$$

where $\pi_{i,t}$ is the profit obtained by DTO i in period t, and β is the discount factor. Note that I will distinguish between the one-period profit, denoted by π, and the present value of all profits, denoted by Π. Repeated interactions make many more strategies available to any DTO, by responding to the actions taken by other DTOs in previous periods. The baseline strategy is simply repeating the non-cooperative Nash equilibrium from the one-period model perpetually, which results in each DTO obtaining a profit $\Pi^N = \dfrac{\pi^N}{1-\beta}$.

The discount factor depends on two different elements: the psychological discount factor, which I will call δ, and the probability p that the current leader of the DTO will still be the leader in the next period. The discount factor will then be $\beta = \delta p$. The probability depends on the government's actions, since policies aimed at capturing or killing leaders decrease the probability that they will be standing during the next

[11] There is, of course, a second solution, where $g_i = 0$. However, this is an unstable solution, since any cartel can appropriate the full benefit in municipality i by investing an infinitesimal amount of resources in the conflict. This is a consequence of the fact that $R_i(g_i)$ is concave *as long as* $g_{i,k} \neq 0$ *for some* $k \neq j$. The solution that we find here is the only Nash equilibrium. The solution with $g_i = 0$ is the collusive equilibrium that we will explore in the next section.

period. This means that DTO leaders are selfish, since they do not value the future of their organization after they are captured or killed.

I will now explore the various strategies that are now available to DTOs. It is not hard to imagine that, if the interactions are repeated, they would benefit if they all reduced the expenditure in the conflict by the same amount. In that case, they would all end up controlling an amount $\frac{1}{n}$ of land, the same as before, with a reduced expenditure, thus obtaining a larger profit. This was not possible in the one-period case because a lower level of expenditure by all other DTOs meant an increase in the marginal utility of expenditure for every DTO (since $\frac{\partial^2 q}{\partial g \partial g_{-i}} < 0$), implying an incentive to deviate and increase expenditure. However, as is usual in repeated games, fear of some kind of retaliation can provide the incentive to stay at an amount of expenditure different from the one-period Nash equilibrium.

More specifically, assume that all DTOs collude by choosing a reserve level of expenditure g_r. The income by each DTO in this cooperative case will then be

$$\pi^c(g_r) = \max_{x_i} p_c q_i(x_i, 1/n, e) - g_r - p_p x_i \tag{11}$$

The first-order condition from this maximization problem is the same as in the non-cooperative case, (3a), which with the particular functional forms that we are using (equations (5) and (6)) is

$$p_c \frac{g_r^2}{(g_r + \varphi x_i e n g_r)^2} = p_p \tag{12}$$

The expenditure g_r can be cancelled in the numerator and denominator, which makes sense since all DTOs are equal and spend equal amounts in the conflict, resulting in each one holding the same fraction of the routes. After isolating x_i, the optimal amount is

$$x^c = \frac{\gamma - 1}{\varphi e n} \tag{13}$$

which is equal to x^N, the amount under a cooperative equilibrium. This would seem surprising. However, this is a consequence of the fact that the aggregate productive behavior is independent of the conflict (equation (4)), and this does not change if the conflict takes the form of a collusive equilibrium or a one-period Nash equilibrium. Therefore, under a cooperative equilibrium the amount of drugs reaching the final consumer market is the same as under the one-period Nash equilibrium.

What determines the reserve level of expenditure in the collusive case? Substituting the optimum (13) in the function to be maximized yields the profit obtained by the DTOs as a function of g_r:

$$\pi^c(g_r) = p_p \frac{(\gamma - 1)^2}{\varphi e n} - g_r \tag{14}$$

As one would expect, this amount is greater than π^N, and the DTOs would be better off, as long as $g_r < g^N$. If the DTOs were somehow able to cooperate and maintain this level of profits indefinitely, their total profit would be $\Pi^c(g_r) = \frac{\pi^c(g_r)}{1-\beta}$.

Let us look at whether any particular DTO has any incentive to deviate from this strategy. The expenditure by other DTOs (g_r) being relatively low, DTO i would probably want to increase its expenditure in order to take a large fraction of the routes. In the next period all other DTOs would retaliate by returning to the default non-cooperative equilibrium, after which no single DTO could deviate to their advantage. This is the standard punishment strategy used in repetitive games in game theory, and it is used in industrial organization as a punishment for firms that deviate from collusion in an oligopoly (see Motta (2004)). Since the benefit from deviating from the cooperative strategy only lasts for one period, the traitor DTO would want to take as much profit as it can for that single period, i.e., the optimal expenditure in the conflict and drugs bought from initial markets given that all other DTOs spend g_r. Thus, in the case of treason, the traitor i can obtain a profit that is given by

$$\pi^t(g_r) = \max_{x_i, g_i} \left[p_c q_i(x_i, R_i(g_i, (n-1)g_r), e) - g_i - p_p x_i \right] \tag{15}$$

The first order conditions are the same as that for the one-period equilibrium except for the expenditure by other DTOs, which is now $g_{-1} = n g_r$. With the specific functional forms from section 3.2.1, they can be written as

$$\frac{p_c \varphi x_i^2 e_i (n-1) g_r}{(g_i(1 + \varphi e x_i) + \varphi e x_i (n-1) g_r)^2} = 1 \tag{16a}$$

$$\frac{p_c g_i^2}{(g_i(1 + \varphi e x_i) + \varphi e x_i (n-1) g_r)^2} = p_p \tag{16b}$$

By dividing the first equation by the second one, it can be seen that $g_i^2 = p_p \varphi e (n-1) g_r x_i^2$, and after substituting this in equation (16a) in order to leave it only in terms of x_i, a quadratic equation can be found, which leads to the following solution:

$$x^t(g_r) = \frac{\gamma - 1}{\varphi e} - \sqrt{\frac{(n-1)g_r}{p_p \varphi e}} \qquad g^t(g_r) = (\gamma - 1)\sqrt{\frac{p_p (n-1) g_r}{\varphi e}} - (n-1) g_r \tag{17}$$

$$\pi^t(g_r) = (n-1) \left[(\gamma - 1)\sqrt{\frac{p_p}{\varphi e (n-1)}} - \sqrt{g_r} \right]^2 \tag{18}$$

The profits of the traitor would then be $\Pi^t = \pi^t + \beta\pi^N + \beta^2\pi^N + \ldots$. This means that no DTO would have an incentive to deviate from the cooperative equilibrium if $\Pi^c(g_r) \geq \Pi^t(g_r)$, namely, if

$$\pi^c(g_r) + \frac{\beta}{1-\beta}\pi^c(g_r) \geq \pi^t(g_r) + \frac{\beta}{1-\beta}\pi^N \tag{19}$$

This means that if there exists some reserve level of expenditure in the conflict such that the *incentive constraint (IC)* (19) is fulfilled, the cooperative Nash equilibrium can be supported. However, it is easy to see that such level *always* exists. By setting $g_r = g^N$, $\pi^c(g_r = g^N)$ becomes π^N, since all DTOs will be cooperating with the conflict expenditure corresponding to the one-period Nash equilibrium. Betraying becomes useless since the one-period optimal response to cooperation with $g_r = g^N$ is a level of investment g^N, so $\pi^t(g_r = g^N)$ becomes π^N as well, and the IC is fulfilled with equality.

The question thus becomes if there exists a level of expenditure $g_r < g^N$ that fulfills the IC. In that case, the DTOs will spend $\bar{g_r}$, the minimum amount that ensures that the IC is fulfilled, which is defined by

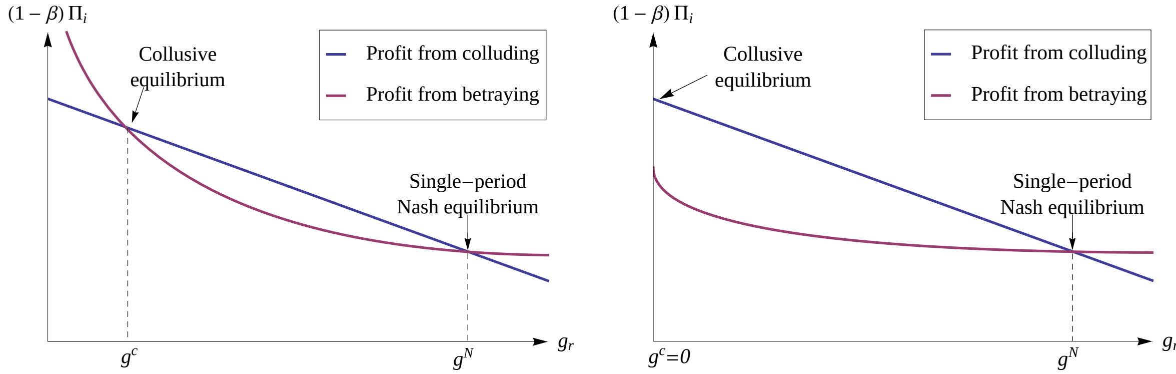
$$\pi^c(\bar{g_r}) + \frac{\beta}{1-\beta}\pi^c(\bar{g_r}) = \pi^t(\bar{g_r}) + \frac{\beta}{1-\beta}\pi^N \tag{20}$$

I will call $\bar{g_r}$ the *dissuasive expenditure*, since it is the minimum amount that DTOs must spend in the conflict in order to dissuade others from betraying.

This analysis can be understood more easily by looking at it graphically. Figure 1 shows how two different cases can arise. If the discount factor is low, meaning that the DTOs do not value the future too much, $\Pi^t(g_r)$ crosses $\Pi^c(g_r)$ twice, at $g_r = g^N$, and at a second nonzero level, which is the level of investment in the conflict by each DTO, since it is the minimum value for which the IC is fulfilled. However, if the discount factor is high, it is possible that there is no such second crossing, meaning that even with zero investment in the conflict other DTOs would prefer not to betray, since returning to the one-period Nash equilibrium would mean a harm greater than the potential benefit from betraying.[12]

[12]It would seem that a third possibility exists. If the derivative of the right side of the IC at the one-period Nash equilibrium were lower than the derivative of the left side, g^N would be the lowest level for which the IC is fulfilled, i.e., $\bar{g_r} = g^N$. However, the derivative of the left side is greater than the derivative of the right side, regardless of the functional forms used. In order to see this, first note that $\frac{\partial \pi^c}{\partial g_r} = -1$, since under collusion $R_i = \frac{1}{n}$, so the earnings are constant, which means that the only change in the profits is the increase in the expenditure in war. On the other hand, from the envelope theorem

$$\left.\frac{\partial \pi^t(g_r)}{\partial g_r}\right|_{g_r=g^N} = p_c \frac{\partial q_i}{\partial R_i}\left.\frac{\partial R_i}{\partial g_r}\right|_{g_r=g^N} - 1 \tag{21}$$



(a) $\beta < \frac{n}{n+1}$: The investment in the conflict is nonzero, since any lower level would make it profitable for cartels to betray.

(b) $\beta > \frac{n}{n+1}$: The future is important enough that even with zero investment in the conflict the collusive equilibrium can be supported.

Figure 1: Comparison of the profit from colluding and betraying. For direct comparison, the vertical axis measures the profit that would have to be earn perpetually to equal the profits on each case, i.e., the total profit Π_i multiplied by $1-\beta$. If DTOs do not value the future too much (if β is not too large, in subfigure 1(a)) the incentive to betray becomes large, so the cooperative equilibrium involves some violence. But if the future becomes important, an equilibrium with no violence is possible, as in figure 1(b).

Let us now find the level of violence in the collusive case for the functional forms specified in section 3.2.1. By substituting the functions for the profits in each situation, we obtain a quadratic equation for $\sqrt{\bar{g}_r}$, which yields two solutions. The first one is simply $\bar{g}_r = \frac{p_p(n-1)(\gamma-1)^2}{\varphi e n^2} = g^N$, which should be expected from the fact that setting $g_r = g^N$ makes the profits in each of the three situations equal (π^N, π^c, and π^t). The second solution leads to $\bar{g}_r = \frac{p_p(n-1)(\gamma-1)^2}{\varphi e n^2}\left(\frac{n(1-\beta)-\beta}{1+(1-\beta)(n-1)}\right)^2$. This solution is valid for small β, and as β increases, we can easily see that $\bar{g}_r$ decreases, until $\bar{g}_r = 0$ for $\beta = \frac{n}{n+1}$[13]. From that point on, DTOs value the future so much that even though

But $\frac{\partial R_i}{\partial g_r} = 0$, so $\left.\frac{\partial \pi^t}{\partial g_r}\right|_{g_r=g^N} = -1$. This means that at $g_r = g^N$ the derivative of the left hand side of the IC is $-\frac{1}{1-\beta}$, whereas the derivative of the right hand side is -1. Thus, there is always some level of conflict $g_r < g^N$ for which the cartels can increase their profit if they collude, and the two cases illustrated in figure 1 covers all possible cases.

Note that this analysis does not depend on the contest-success functions having any particular functional form. Additionally, nothing precludes the profit from betraying from being concave, which would only mean that it would be easier for the peaceful equilibrium to exist. However, one would expect it to be convex at least for a very low level of g_r, since the initial reserve expenditure in the conflict has a very strong impact on whether it would be beneficial for DTOs to betray.

[13]The expression for $\bar{g}_r$ still has a meaning for $\beta > \frac{n}{n+1}$, but it is irrelevant for our problem: Since

betraying is free, since a minimal investment in the conflict allows them to take all routes, they do not do it because it will break up the cooperative equilibrium. The illegal drug market will then settle at the level of violence described by the following proposition:

Proposition 3. *In the multiple-period cooperative equilibrium, the investment in violence by each DTO is*

$$g^c = \begin{cases} p_p \dfrac{n-1}{n^2} \dfrac{(\gamma-1)^2}{\varphi e} \left(\dfrac{n(1-\beta)-\beta}{1+(1-\beta)(n-1)} \right)^2 & \text{if } \beta < \dfrac{n}{n+1} \\ 0 & \text{if } \beta \geq \dfrac{n}{n+1} \end{cases} \tag{22}$$

As a direct consequence, the total level of violence will be

$$G^c = \begin{cases} p_p \dfrac{n-1}{n} \dfrac{(\gamma-1)^2}{\varphi e} \left(\dfrac{n(1-\beta)-\beta}{1+(1-\beta)(n-1)} \right)^2 & \text{if } \beta < \dfrac{n}{n+1} \\ 0 & \text{if } \beta \geq \dfrac{n}{n+1} \end{cases} \tag{23}$$

The optimal levels of x_i, q_i, and Q_i are the same as in the one-period Nash equilibrium:

$$Q^c = \frac{\gamma-1}{\gamma} \frac{1}{\varphi e} \qquad x^c = \frac{\gamma-1}{\varphi e n} \qquad q^c = \frac{\gamma-1}{\gamma} \frac{1}{\varphi e n} \tag{24}$$

Proof. The level of g^c is the lower solution to (20). In order to find G^c one simply has to multiply g^c by n.

Since all DTOs invest the same amount in the conflict, all of them control a fraction $R_i = \frac{1}{n}$ of the routes, which is the same as in the one-period Nash equilibrium. Thus, the optimal level of x_i (and q_i) will also be the same, meaning that collusion will in no way change the amount of drugs being taken and reaching the final consumer market. □

The total level of violence is shown in figure 2. It can be clearly seen that in the lower right corner there is a region of low n/high β where there is no violence, and that decreasing β or increasing the number of cartels increases violence. Let us now find the comparative statics on the equilibrium amounts of violence and drugs reaching the final consumer market, this time algebraically. In order to do so, first note that the level of violence is equal to the non-cooperative Nash equilibrium multiplied by a factor

we are finding an expression for $\bar{g}_r$ after having solved an equation for $\sqrt{\bar{g}_r}$, our solution also includes a solution corresponding to the negative square root of $\bar{g}_r$. This, however, means changing the sign of the square root of $g_r$in our expression for π^t, which is a mathematical problem with no economic interpretation.

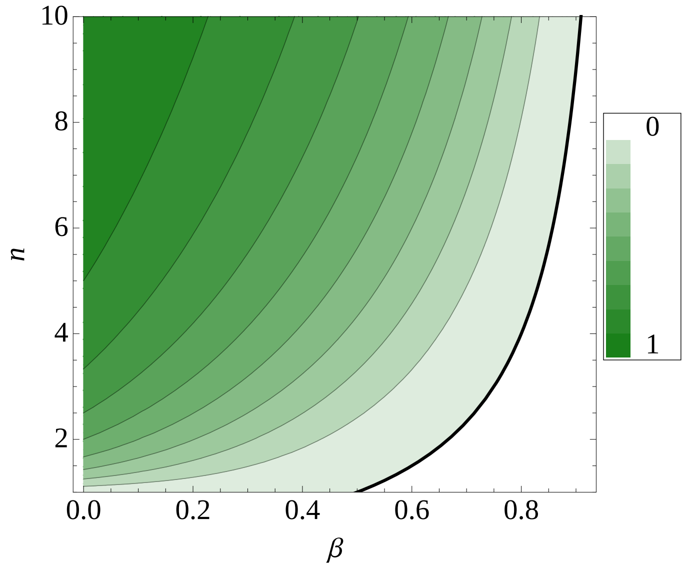


Figure 2: Level of violence as a function of the number of cartels and the discount factor. The level of violence is measured as a fraction of $p_p \frac{(\gamma-1)^2}{\varphi e}$, the level of violence in the one-period Nash equilibrium with $n \to \infty$. The region to the right of the thick black line represents the combinations of n and β for which an equilibrium without violence can be supported.

$\left(\frac{n(1-\beta)-\beta}{1+(1-\beta)(n-1)}\right)^2$, as long as $\beta < \frac{n}{n+1}$. The level of violence is otherwise zero. This can be summarized as $G^c = \theta(n,\beta)G^N$, where

$$\theta(n,\beta) = \begin{cases} \left(\frac{n(1-\beta)-\beta}{1+(1-\beta)(n-1)}\right)^2 & \text{if } \beta < \dfrac{n}{n+1} \\ 0 & \text{if } \beta \geq \dfrac{n}{n+1} \end{cases} \tag{25}$$

As long as there is some violence, i.e., $\beta < \frac{n}{n+1}$, this factor is closer to one the greater n is, and the smaller β is, which can be easily seen by finding its derivatives. These two facts, combined with the comparative statics on the one-period Nash equilibrium quantities, lead to the following comparative statics:

Proposition 4. *Under a symmetric multiple-period collusive equilibrium, the comparative statics on the total expenditure in the conflict and the total amount of drugs taken to the consumer market are as follows:*

1. *The amount of violence is nonzero if and only if* $\beta < \dfrac{n}{n+1}$ *(meaning that the number of DTOs is sufficiently high or the DTOs are sufficiently impatient).*

2. *If violence is nonzero:*

- $\frac{\partial G^c}{\partial e} < 0$: *More expenditure by the government in enforcement reduces the level of violence.*
- $\frac{\partial G^c}{\partial n} > 0$: *A greater number of DTOs increases the level of violence.*
- $\frac{\partial G^c}{\partial \beta} < 0$: *If DTOs are more impatient, the level of violence increases.*

3. *Regardless of violence being zero or nonzero:*

- $\frac{\partial Q^c}{\partial e} < 0$: *More expenditure by the government in enforcement reduces the amount of drugs reaching the final consumer market.*
- $\frac{\partial Q^c}{\partial n} = 0$: *The number of DTOs has no effect on the amount of drugs reaching the final consumer market.*
- $\frac{\partial Q^c}{\partial \beta} = 0$: *Whether DTOs are impatient or not has no effect on the amount of drugs reaching the final consumer market.*

The first outcome of the repetitive-interaction model is that if DTOs have perfect information, they will always collude, resulting in lower levels of violence than in the non-cooperative Nash equilibrium. This means that most of the analyses that had been developed until now missed an important part of the behavior of DTOs: They never engage with the level of violence predicted by one-period models if they value the future. Instead, it is to their benefit to invest less in the conflict, as long as all DTOs do so. The degree to which this new element that I am analyzing reduces the predicted level of violence depends on the discount factor and on the number of cartels: a greater number of cartels increases the level of violence, as well as a lower discount factor.

This model shows that under the right circumstances cartels can coexist without any violence. This depends mainly on two conditions: there must be a low number of cartels, and they must place a high value on their future earnings. Even though an individual DTO could betray and seize all the routes for one period at a low cost, meaning huge benefits in the short run, this would also mean reducing their profits in the long run. If they are sufficiently fearful of the future reduction in their profits, they will never want to betray, no matter how easy it is for them to take all the routes. On the other hand, if there are more than a few cartels, by seizing all routes they would obtain a great increase in profits, which would require a very high discount factor for a totally peaceful equilibrium to exist.

Many policies that can be implemented by governments have an important effect on the level of violence and on the amount of drugs reaching the final consumer market.

For instance, increasing the level of enforcement decreases both amounts. Taking into account what Becker et al. (2006) show, this is only the case if demand for drugs is elastic. Therefore, an important extension to this model is to consider what would happen if prices are not exogenous, but instead depend on the amount of drugs being bought at the producer market and being taken to the consumer market.

The type of policies implemented in Mexico and Colombia are ineffective in terms of reducing the amount of drugs, since the aggregate trafficking behavior of DTOs does not depend on the conflict that defines who holds the routes. But these policies do cause an unintended increase in violence. When governments succeed in fragmenting cartels, as happened in Colombia, a greater number of DTOs means that if they betray they can take a fraction of the share held by a greater number of cartels, thus increasing the profitability of betrayal. The strategy followed by Calderón's government instills a feeling of restlessness and impatience in cartel leaders, who consider betrayal more seriously, due to the one-time benefit they could obtain. Additionally, the actions followed by the Mexican government resulted in many cartels being split due to disagreements after the fall of previous leaders, which resulted in the emergence of new cartels. In both cases the actions taken by the government are the kind of policies that may lead to breaking down the peaceful equilibrium. Even if DTOs were already in an equilibrium with some level of war, DTOs would respond to these policies by increasing the investment in the conflict as a means to dissuade others from breaking the (relatively) peaceful collusive equilibrium, leading to an increase in the level of violence. This is a direct contradiction to the assumptions of the war on drugs promoted by the U.S. government during the last two decades.

As in models of collusion in industrial organization, a significant factor that is important for the cooperative equilibrium is the existence of proper information channels between the participants. This is the only way for them to know whether others are still cooperating, or if they are betraying. The participants need to know this for a believable threat of retribution to exist, which provides the incentive for others to stay in the collusive equilibrium. The traditional information channel in the case of an oligopoly are prices: once firms see a drop in prices, they assume that somebody betrayed, and the market goes back to the one-period Nash equilibrium. In the case of DTOs, the logical information channel is the level of violence. If a DTO observes an escalating conflict, it might well assume that some other DTO has betrayed them. However, if it is complicated for DTOs to observe the actual investment in the conflict by others (i.e., if they can only measure it with some degree of uncertainty), it is possible that they believe that others have betrayed when they actually received a wrong signal. Therefore, uncertainty makes it much harder for a collusive equilibrium to be held. This shows the inconvenience of another kind of action that could be used by governments: they could attack DTOs, while trying to deceive the victim by making him think that another DTO planned the attack. DTOs would the operate in an environment of uncertainty, in which they are never sure if others are colluding or not. Although this may seem at first to be beneficial for governments, this model shows that this kind of uncertainty results in an increase in violence, since it is a factor that makes it harder for the collusive equilibrium to be held. This is an important point that could be further analyzed.

4 Conclusions

In this work I extend the analysis of DTOs as single-period profit maximizers to a multiple-period approach. DTOs are modeled as firms that buy drugs at a producer region and attempt to take them to consumers through a trafficking region. In the process, they engage in two conflicts. First, they fight against other DTOs over who controls routes in the trafficking region. Then, they engage government forces who try to seize drugs on their way to consumers. If DTOs have perfect information, they will never be at the one-period Nash equilibrium that previous works analyze. Instead, they collude by decreasing the amount of resources spent in the conflict against other DTOs, which results in less bloodshed than was previously predicted. Even a peaceful equilibrium without any violence between DTOs can be sustained if there are only a few powerful cartels that are interested in maximizing the present value of their profits with a high enough discount factor.

Another important result is that productive behavior (the amount of drugs bought from upstream markets and the amount of drugs sold to consumer markets) remains unchanged if governments attack cartel leaders or if DTOs are more fragmented; this is a consequence of the fact that productive behavior is independent of the conflict. Thus, some of the traditional policies preached by the U.S.-led war on drugs do not accomplish their purpose of curbing supply. As an unintended consequence, such policies increase violence between DTOs: they harm trafficking regions while attaining no positive effect on consumer regions. However, governments do have the means to reduce supply: enforcement activities, focused on drug seizures, decrease the amount of drugs reaching final markets. Additionally, enforcement decreases the level of violence in trafficking regions. This means that both for consumer nations, who wish to decrease the supply of drugs, and for trafficking nations, who wish to decrease violence, enforcement activities focused on seizing drugs are preferable to policies aimed at fragmenting DTOs and killing cartel leaders.

References

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

Bardey, David, Daniel Mejía, and Andrés Zambrano, "The endogenous dynamics of crime structure: an entry and exit story," *MIMEO, Universidad de los Andes*, 2013.

Becker, Gary S., "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 1968.

_ **, Kevin M. Murphy, and Michael Grossman**, "The Market for Illegal Goods: The Case of Drugs," *Journal of Political Economy*, 2006, *114* (1).

Camacho, Álvaro, "Paranarcos y narcoparas: trayectorias delicuenciales y políticas," in Álvaro Camacho, ed., *A la sombra de la guerra. Ilegalidad y nuevos órdenes regionales en Colombia.*, Ediciones Uniandes - CESO, 2009.

_ , "Narcotrafico: mutaciones y política," in Alejandro Gaviria and Daniel Mejía, eds., *Políticas antidroga en Colombia: éxitos, fracasos y extravíos*, Ediciones Uniandes, 2011.

Castillo, Juan Camilo, Daniel Mejía, and Pascual Restrepo, "Illegal drug markets and violence in Mexico: the causes beyond Calderón," *MIMEO, Universidad de los Andes*, 2013.

Chumacero, Rómulo A., "Evo, Pablo, Tony, Diego and Sonny: General Equilibrium Analysis of the Market for Illegal Drugs," *World Bank Policy Research Working Paper No. 4565*, 2008.

Garfinkel, Michelle R. and Stergios Skaperdas, "Economics of Conflict: An Overview," in Todd Sandler and Keith Hartley, eds., *Handbook of Defense Economics - Defense in a Globalized World*, Vol. 2, Elsevier, 2007.

Grossman, Herschel I., "Kleptocracy and Revolutions," *Oxford Economic Papers*, 1999, *51*, 267–283.

— **and Daniel Mejía**, "The war against drug producers," *Economics of Governance*, 2008, *9* (1), 5–23.

— **and Minseong Kim**, "Swords or Plowshares? A Theory of the Security of Claims to Property," *Journal of Political Economy*, 1995.

Guerrero, Eduardo, "La raíz de la violencia," *Nexos*, 2011.

Haavelmo, Trygve, *A study in the theory of economic evolution*, North-Holland, Amsterdam, 1954.

Hirshleifer, Jack, "Conflict and rent-seeking success functions: Ratio vs. difference models of relative success," *Public Choice*, 1989, *63* (101-112).

Mejía, Daniel and Pascual Restrepo, "The War on Illegal Drug Production and Trafficking: An Economic Evaluation of Plan Colombia," *Documento CEDE 2008-19*, 2008.

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

Motta, Massimo, *Competition Policy: Theory and Practice*, Cambridge University Press, 2004.

Noh, Suk Jae, "Production, appropriation, and income transfer," *Economic Inquiry*, 2002, *40* (2), 279–287.

Tirole, Jean, *The Theory of Industrial Organization*, The MIT Press, 1988.

A Comparative statics for a one-period Nash equilibrium

In order to find the comparative statics on the one-period Nash equilibrium, I find the total differential of both first-order conditions (3a) and (3b):

$$p_c\frac{\partial^2 q_i}{\partial x_i^2}dx^N + p_c\frac{\partial^2 q_i}{\partial x_i \partial R_i}\frac{\partial R_i}{\partial g^N}dg^N = -p_c\frac{\partial^2 q_i}{\partial x_i \partial e}de - \frac{\partial q_i}{\partial x_i}dp_c + dp_p - p_c\frac{\partial^2 q_i}{\partial x_i R_i}\frac{\partial R_i}{\partial n}dn \quad (26a)$$

$$p_c \frac{\partial^2 q_i}{\partial x_i \partial R_i} \frac{\partial R_i}{\partial g_i} dx^N + p_c \left[\frac{\partial q_i}{\partial R_i} \frac{\partial^2 R_i}{\partial g_i \partial g^N} + \frac{\partial^2 q_i}{\partial R_i^2} \frac{\partial R_i}{\partial g_i} \frac{\partial R_i}{\partial g^N} \right] dg^N =$$
$$- p_c \frac{\partial^2 q_i}{\partial R_i \partial e} \frac{\partial R_i}{\partial g_i} de - \frac{\partial q_i}{\partial R_i} \frac{\partial R_i}{\partial g_i} dp_c - p_c \left[\frac{\partial^2 q_i}{\partial R_i^2} \frac{\partial R_i}{\partial g_i} \frac{\partial R_i}{\partial g_{-i}} + \frac{\partial q_i}{\partial R_i} \frac{\partial^2 R_i}{\partial g_i \partial n} \right] dn \qquad (26b)$$

In the previous equations the treatment of g^N is not the same as that of x^N. This is because a change in x^N only influences the first-order conditions through x_i, which means that differentiating with respect to x^N is equivalent to differentiating with respect to x_i. On the other hand, g^N influences the first-order conditions through both g_i and g_{-i}. This means that differentiating with respect to g^N, g_i, and g_{-i} are all different. In particular, note that if the expenditure in the conflict of all DTOs is the same, which is the case if the problem is symmetrical, the amount of land that each one of them holds is $\frac{1}{n}$, irrespective of the amount spent by each one of them. Therefore, $\frac{\partial R_i}{\partial g^N} = 0$, which greatly simplifies the expressions above.

Since n is discrete, it does not make a lot of sense to derive with respect to n. However, I can create a continuous function $\hat{R}$ such that if the investment of all cartels is equal, $\hat{R}(g^N, n) = R_i(g_i, g_{-i})$. As long as this function matches the value it should have for all integer values of n, finding that the derivative with respect to n of any given function is positive leads to the conclusion that that function is strictly increasing in n. The opposite can be concluded if the derivative of any function with respect to n is decreasing. From now on, it should be clear that what I mean by expressions such as $\frac{\partial^2 R_i}{\partial g_i \partial n}$ is actually $\frac{\partial^2 \hat{R}_i}{\partial g_i \partial n}$. Note that as the number of cartels increases, the total size of the conflict increases, so $\frac{\partial}{\partial n} \left(\frac{\partial \hat{R}_i}{\partial g_i} \right) = \frac{\partial^2 \hat{R}_i}{\partial n \partial g_i} < 0$.

Equations (26a) and (26b) can be combined into a unique matrix equation, in which a matrix multiplies the vector (dx_i, dg^N) on the left side, and another matrix multiplies the vector (de, dp_c, dp_p, dn) on the right hand side. This matrix equation describes how the response variables, g^N and x_i, respond to infinitesimal changes in the exogenous variables from the right hand side. The left-side vector can be isolated by multiplying both sides by the inverse of the left-side matrix. Then, by setting all right-side differentials but any single one equal to zero, the partial derivatives of the response variables can be found. The final result of this process yields the following derivatives:

$$\frac{\partial x^N}{\partial e} = -\frac{\partial^2 q_i}{\partial x_i \partial e}\left(\frac{\partial^2 q_i}{\partial x_i^2}\right)^{-1} \tag{27a}$$

$$\frac{\partial g^N}{\partial e} = \frac{1}{\Delta}\left[\frac{\partial^2 q_i}{\partial x_i \partial R_i}\frac{\partial^2 q_i}{\partial x_i \partial e}\frac{\partial R_i}{\partial g_i} - \frac{\partial^2 q_i}{\partial x_i^2}\frac{\partial^2 q_i}{\partial R_i \partial e}\frac{\partial R_i}{\partial g_i}\right] \tag{27b}$$

$$\frac{\partial x^N}{\partial n} = -\frac{\partial^2 q_i}{\partial x_i \partial R_i}\frac{\partial R_i}{\partial n}\left(\frac{\partial^2 q_i}{\partial x_i^2}\right)^{-1} \tag{27c}$$

$$\frac{\partial g^N}{\partial n} = \frac{1}{\Delta}\left[\left(\frac{\partial^2 q_i}{\partial x_i \partial R_i}\right)^2\frac{\partial R_i}{\partial g_i}\frac{\partial R_i}{\partial n} - \frac{\partial^2 q_i}{\partial x_i^2}\frac{\partial^2 q_i}{\partial R_i^2}\frac{\partial R_i}{\partial g_i}\frac{\partial R_i}{\partial n} - \frac{\partial^2 q_i}{\partial x_i^2}\frac{\partial q_i}{\partial R_i}\frac{\partial R_i^2}{\partial g_i \partial n}\right] \tag{27d}$$

where $\Delta = \frac{\partial^2 q_i}{\partial x_i^2}\frac{\partial q_i}{\partial R_i}\frac{\partial^2 R_i}{\partial g_i \partial g^N} > 0$. The sign of all the individual derivatives in the expressions above are clear, which allows me to find the sign of each expression. I am also interested in finding the derivatives of q_i, in order to see how the amount of drugs taken to the final market changes. In order to find its derivatives, I find the following derivatives by using the chain rule:

$$\frac{\partial q^N}{\partial e} = \frac{\partial q_i}{\partial e} + \frac{\partial x_i}{\partial e}\frac{\partial q_i}{\partial x_i} + \frac{\partial R_i}{\partial g^N}\frac{\partial g^N}{\partial e}\frac{\partial q_i}{\partial R_i} \tag{28a}$$

$$\frac{\partial q^N}{\partial n} = \frac{\partial x_i}{\partial n}\frac{\partial q_i}{\partial x_i} + \frac{\partial R_i}{\partial n}\frac{\partial q_i}{\partial R_i} \tag{28b}$$

Since $\frac{\partial R_i}{\partial g^N} = 0$, all terms involving it can be cancelled.

The signs of all these derivatives are clear from the signs of the individual derivatives, except for the sign of $\frac{\partial g^N}{\partial n}$, which is so far ambiguous. I use an alternate approach to find it. Since q_i is homogeneous of degree one in (x_i, R_i), the total amount of drugs that reach the consumer market is $Q(X, R, e)|_{R=1} = nq_i(x_i, R_i, e) = q_i(nx_i, nR_i, e) = q_i(X, 1, e)$. The homogeneity of degree one means that its derivatives are homogeneous of degree zero, so $\frac{\partial q_i(x_i,R_i,e)}{\partial R_i} = \frac{\partial q_i(X,1,e)}{\partial R_i} = \left.\frac{\partial Q(X,R,e)}{\partial R}\right|_{R=1}$. Clearly, this marginal productivity does not depend on n. The first order condition (3b) can now be written as

$$p_c\frac{\partial Q}{\partial R}\frac{\partial R_i}{\partial g_i} = 1 \tag{29}$$

Since neither p_c nor $\frac{\partial Q}{\partial R}$ depends on the number of cartels, $\frac{\partial R_i}{\partial g_i}$ cannot depend on it either, which means that its derivative with respect to n must be zero:

$$\frac{\partial^2 R_i}{\partial n \partial g_i} = \left[\frac{\partial^2 R_i}{\partial g_i^2} + (n-1)\frac{\partial^2 R_i}{\partial g_i \partial g_{-i}}\right]\frac{\partial g^N}{\partial n} + g^N\frac{\partial^2 R_i}{\partial g_i \partial g_{-i}} = 0 \tag{30}$$

and $\frac{\partial g^N}{\partial n}$ can now be isolated:

$$\frac{\partial g^N}{\partial n} = -g^N\frac{\partial^2 R_i}{\partial g_i \partial g_{-i}}\left[\frac{\partial^2 R_i}{\partial g_i^2} + (n-1)\frac{\partial^2 R_i}{\partial g_i \partial g_{-i}}\right]^{-1} \tag{31}$$

The sign of this expression can now be determined to be negative.

I now turn to finding the comparative statics on the aggregate quantities G and Q. Finding the derivatives with respect to e is trivial, since they are are the individual quantities multiplied by n, which is held fixed, so their signs are the same as those of the individual quantities. In order to find the derivative of G^N with respect to n, I use the fact that $\frac{\partial G^N}{\partial n} = g^N + n\frac{\partial g^N}{\partial n}$, which leads to

$$\frac{\partial G^N}{\partial n} = g^N \left[\frac{\partial^2 R_i}{\partial g_i^2} - \frac{\partial^2 R_i}{\partial g_i \partial g_{-i}}\right] \left[\frac{\partial^2 R_i}{\partial g_i^2} + (n-1)\frac{\partial^2 R_i}{\partial g_i \partial g_{-i}}\right]^{-1} \tag{32}$$

which is positive. Finally, to find the derivative of Q^N with respect to n, I will follow a procedure similar to the one used to obtain (29), but on the other first-order condition. The homogeneity of q_i means that $\frac{\partial q_i(x_i,R_i,e)}{\partial x_i} = \frac{\partial q_i(X,1,e)}{\partial x_i} = \left.\frac{\partial Q(X,R,e)}{\partial X}\right|_{R=1}$, and substitution in (3a) yields

$$p_c \frac{\partial Q}{\partial X} = p_p \tag{33}$$

In the last expression both prices are constant, meaning that any change in n must result in no change in $\frac{\partial Q}{\partial X}$. Q is a function of X, R, and e, but a change in n induces no change in $R = 1$ or e, which means that X must be held fixed. None of the variables Q depends on change, so it does not change either. Thus, $\frac{\partial Q}{\partial n} = 0$.