Why Present-Oriented Societies Undergo Cycles of Drug Epidemics [+]

Doris A. Behrens
Vienna University of Technology, Department of Operations Research and Systems Theory; and University of Klagenfurt, Department of Economics

Jonathan P. Caulkins
Carnegie Mellon University, H. John Heinz III School of Public Policy and Management; and RAND Drug Policy Research Center

Gernot Tragler, and Gustav Feichtinger [*]
Vienna University of Technology
Department of Operations Research and Systems Theory

Abstract: Musto (1987) hypothesizes that cycles of drug use arise when the current generation of youth no longer remembers the adverse experiences of their forebears. This paper underlines this hypothesis through the results of an optimal control model of drug epidemics that incorporates an endogenous initiation function, models the reputation of a drug as being determined by memories of past use, and finds the optimal drug treatment strategy. Interestingly, unless the societal discount rate is quite low, if memories of past users decay moderately quickly, the optimal strategy is cyclic. Hence, not only do we find that "those who forget the past are condemned to repeat it", but also that "for those who forget the past and over-value the present it may even be optimal to have their future recreate the past". These findings are underlined by numerical analysis based on empirical data from the current U.S. cocaine epidemic.

JEL: K42

Keywords: Illicit drugs, chain of drug epidemics, optimal control, Hopf bifurcation, limit cycle

---

[+] This research was partly financed by the Austrian Science Foundation (FWF) under contract No. P11711-OEK ("Dynamic Law Enforcement"), the National Consortium on Violence Research, and by the US National Science Foundation under Grant No. SBR-9357936. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

[*] Corresponding author: Gustav Feichtinger; Department of Operations Research and Systems Theory, Vienna University of Technology, Argentinierstrasse 8/1192, A – 1040 Vienna, Austria; Tel.: +43 – 1 – 58801 – 11927; Fax: +43 – 1 – 58801 – 11999; E-mail address: or@e119ws1.tuwien.ac.at.

1 Introduction

In this paper we use an autonomous three-state, single control dynamic model to explore how best to control epidemics of drug use when initiation is both contagious (transmitted from existing users) and moderated by the memory of past experiences of drug use, as suggested by Musto (1987). We reconcile the competing tendencies of users to both promote and deter initiation by distinguishing two types of users, as in Everingham and Rydell (1994). In particular, “light” users are envisioned as creating and suffering few adverse consequences and, hence, helping to spread drug use. In contrast, “heavy users” are envisioned as suffering substantially more harm and, thus, the memory of all such people who have suffered harm is balanced against the number of currently happy (or apparently happy) light users to determine the drug’s reputation. When there are relatively many negative memories, the rate of initiation into drug use is dampened.

This analysis differs from similar but simpler models analyzed in the past (Behrens et al., 1999) because reputation is a function not merely of the current number of users but also of the memory of past heavy users. This extension has obvious appeal. All knowledge of the negative experiences of a heavy user is unlikely to disappear the moment the individual exits the population, particularly if the exit is by physical death from drug use (as opposed to ceasing use or moving out of the area).

Furthermore, this extension yields important insights. Of a practical vein, with this model there is never a time when removing heavy drug users is counter-productive. With past models, since removing a heavy user immediately erased all memory of that individual, it sometimes appeared preferable to allow a person to suffer rather than to help them recover. The benefits of helping them directly were over-ridden by the cost of not being able to make an example of their suffering. Such hard-hearted policies may not be needed to the extent that past users can be remembered.

In a more philosophical vein, we find a fascinating interaction between a society’s present-orientation, its ability to remember the past, and the occurrence of cycles in the future. We re-discover the old adage that “those who forget the past are condemned to repeat it”. More surprisingly, we find that it can actually be desirable to relive past epidemics – at least for myopic decision makers. Or to put it in simple terms, “*for those who forget the past and over-value the present it may be optimal to have their future recreate the past*”.

From a mathematical perspective, what is most interesting is that the qualitative nature of the system behavior can be plotted explicitly in terms of these two key parameters (memory and far-sightedness). This plot shows that it can be optimal to have both the control and the number of heavy or dependent users cycle, and it shows the combinations of parameter values for which this cycling is optimal. All these findings are underlined by numerical analysis based on empirical data from the current U.S. cocaine epidemic.

2 Model Formulation

Everingham and Rydell (1994) introduced and parameterized a discrete-time, Markov model of cocaine demand in the US that differentiated between light ($L(t)$) and heavy ($H(t)$) users. The model was used to describe past trends and project a baseline counterfactual for purposes of policy analysis (Rydell and Everingham, 1994), but initiation was scripted, not modeled explicitly.

Behrens et al. (1999) converted this model into continuous time and extended it to include an endogenous model of initiation in which some initiation was "spontaneous" (e.g., because of in migration) but most occurred through interactions with current light users. The rate at which light users recruited new initiates was moderated by the negative "reputation" of the drug. The reputation was modeled as a negative exponential function of the relative number of current heavy and light users. Omitting an explicit dependence on time, since all variables depend on the current stage of the epidemic, this can be written:

$$\dot{L} = I(L,H) - (a+b)L, \qquad L(0) = L_0$$
$$\dot{H} = bL - gH, \qquad H(0) = H_0 \tag{1a-b}$$
$$I(L,H) = \tau + s\,L\,exp\left[-q\tfrac{H}{L}\right] \tag{1c}$$

where

L = number of light users,
H = number of heavy users,
I(L,H) = initiation into light use,
a = rate of quitting from light use,
b = rate of escalation from light to heavy use,
g = rate of quitting from heavy use,
τ = rate of initiation into drug use without contact with existing users,
s = average rate at which light users attract non-users when drug's reputation is benign,
q = constant measuring the deterrent effect of heavy drug use on initiation.

This model had interesting dynamics and generated a variety of policy insights (Behrens et al., 1999), but it has one odd and undesirable characteristic. According to this model, removing a heavy user (either naturally or through treatment) immediately and entirely erases all of that individual's contribution to the negative reputation of the drug. In reality, memories of drug problems can persist.

One remedy to this problem would be to introduce a third state that represents the number of people who had ever been heavy users and let this new state replace the current number of heavy users in the reputation function. This possibility was investigated by Behrens (1998), but is not really satisfactory. Whereas the original model was unrealistic in assuming there could be no memory of past users, this alternative is unrealistic in assuming that there is perfect and infinite memory.

Here we propose an intermediate approach, that the drug's reputation is governed by the relative number of light users and a decaying memory of people who have ever been heavy users (denoted $E(t)$). The flow into this state is the same as the flow into the number of current heavy users (bL). The outflow is a simple exponential decay governed by a memory parameter (δ). When this parameter is small, memories of heavy users are long-lived; when δ is large, the dampening effect of heavy users' problems on initiation dissipates rapidly.[1]

(Current) heavy users generate the greatest social costs, so for a policy control we focus on drug "treatment" that removes heavy users. "Treatment" is in quotes because any intervention that removes heavy users is germane, whether that intervention is traditional treatment, so-called coerced abstinence, or any of various partnerships between traditional treatment and the criminal justice system (e.g., so-called "drug courts").

As in Behrens et al. (forthcoming) and Tragler et al. (1997) we model treatment as having diminishing returns, in part for the usual reasons and in part because of "cream skimming" in which those who are easiest to treat are treated first. In particular, the exit rate due to treatment (β) as a function of the level of treatment spending (u) is modeled as a monotonically increasing function which is concave in the expenditures on treatment per heavy user ($H>0$), i.e. $\beta'>0,\ \beta''<0$:

$$\beta(H,u) = c\left(u/H\right)^{d} = c\,u^{d}H^{-d} \tag{2}$$

c = constant measuring the average efficiency of treatment
d = constant measuring the marginal efficiency of treatment

The quantity or weight of drugs consumed has advantages as a general purpose scalar measure of the size of a drug problem (Rydell et al., 1996). Hence, we minimize the discounted sum of the control costs (u) plus consumption (Q) times the social cost per unit consumed (κ). Consumption is itself a weighted sum of the number of light and heavy users, with weights determined by the per capita consumption rates. I.e., $Q = \mu_L L + \mu_H H$, where μ_L and μ_H are the rates of consumption for light and heavy users, respectively.

Hence, we consider a nonlinear, autonomous, infinite-horizon optimal control problem with a single control u (treatment) and three states, L, H, and E (light, heavy, and memory of ever-heavy users), where the net present value, discounting at rate $r>0$, of total costs has to be minimized. Initiation ($I(L,E)$) is specified by function (1c) with E replacing H, and treatment by function (2). That is, we seek values of u that maximize the integral

[1] There are at least two other ways to achieve similar "intermediate measures of deterrence". One uses decaying heavy-user-years as this measure (see Caulkins et al., 1999). Another makes the reputation of the drug a function of both light and heavy users, which creates analytical difficulties, but does not give more insights using numerical simulation.

$$J = -\int_0^{\infty} e^{-rt} \left(\kappa Q + u \right) dt \tag{3}$$

subject to
$$\begin{aligned} \dot{L} &= I(L,E) - (a+b)L, & L(0) &= L_0 \\ \dot{H} &= bL - (g + \beta(H,u))H, & H(0) &= H_0 \\ \dot{E} &= bL - \delta E, & E(0) &= H(0) = E_0 \end{aligned} \tag{4a-c}$$

where the integral always converges due to the specific structure of the problem (see Michel, 1982). Note, that we basically investigate a non-autonomous, one state/one control optimal control problem with respect to the state dynamics, $\dot{H} = G(t) - (g + \beta(H,u))H$, where the inflow into H, $G(t)$, is determined by the L-E-system. We solve this problem by a transformation into an autonomous optimal control problem with two more states, i.e. by adding the second system which determines the input function $G(t)$. This rearrangement also allows analytical stability analysis with respect to all parameter values. The following flow-diagram helps to visualize the optimal control Model (3, 4).

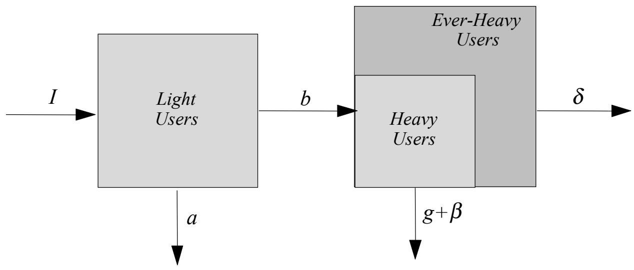

Figure 1. Flow diagram of optimal control Model (3, 4)

3 Model Solution

The derivation of the *necessary conditions* that constitute the maximum principle proceeds in the usual manner (Léonard and Long, 1992; Feichtinger and Hartl, 1986), where H denotes the Hamiltonian function, $\pi_0 \geq 0$ a nonnegative constant multiplier associated with the integrand of the objective function, and π_1, π_2 and π_3 the co-state variables; all in current values.

$$\mathsf{H} = -\pi_0(\kappa Q + u) + \pi_1\left(I - (a+b)L + \pi_2(bL - gH - \beta H) + \pi_3(bL - \delta E)\right) \tag{5}$$

Only in pathological cases will π_0 equal zero (Leonard and Long, 1992). In particular, if we assume $\pi_0 = 0$ in Equation (5), the Hamiltonian maximizing condition, $u^* = arg\,max\,\mathsf{H}$, yields $u^* = \infty$, because H is monotonically increasing in u and there is no upper bound on u. That, however, causes the objective function (Equation (3)) to be minimized which is a contradiction to our purpose of minimizing the total social costs. From that we conclude that $\pi_0 \neq 0$, so we may set $\pi_0 = 1$.

The non-negativity constraint on treatment spending will never be binding because the marginal efficiency of the first dollar spent on treatment is infinite. Hence, there is no need to augment the Hamiltonian with a Lagrangean-type expression that incorporates the constraint (with a multiplier). Consequently, applying the maximum principle to Equation (5) (for $\pi_0 = 1$) yields necessary conditions for an *optimal interior control* u, i.e. $\mathsf{H}_u = 0$, where we omit the asterisks to simplify notation,

$$\mathsf{H}_u = -1 - \underbrace{\pi_2}_{<0} H \underbrace{\beta_u}_{>0} = 0 \qquad \Rightarrow \qquad u = u(H, \pi_2) = H(-cd\pi_2)^{1/(1-d)} > 0\,. \tag{6}$$

As the second derivative of the Hamiltonian with respect to *u* is always negative, $u > 0$ is uniquely determined at each instant of time. Furthermore,

Proposition 1*: For the optimal control Problem (3, 4) the necessary conditions are sufficient since the Hamiltonian is jointly concave in the states and the control, i.e., the Hessian matrix* $D^2\mathsf{H}(L, H, E, u)$ *is negative semidefinite.* (For a proof, see Appendix (A.1).)

Choosing treatment spending, *u*, to maximize the Hamiltonian, H, at each instant of time leads to a set of differential equations constituting the canonical system of the optimal control problem (3, 4) in the state/costate-space. Note that the rate at which heavy users quit abuse due to optimal treatment spending will be expressed by the term

$$B(\pi_2) := \beta(H, u^*) = c\,(-cd\pi_2)^{d/(1-d)}\,, \tag{7}$$

which depends on the value of an additional heavy user, π_2, along the optimal path but not on the number of heavy users, H, itself.

$$\dot{L} = I - (a + b)L \tag{4a}$$
$$\dot{H} = bL - (g + B)H \tag{4b}$$
$$\dot{E} = bL - \delta E \tag{4c}$$
$$\dot{\pi}_1 = \kappa\mu_L + \pi_1(r + a + b - I_L) - (\pi_2 + \pi_3)b \tag{4d}$$
$$\dot{\pi}_2 = \kappa\mu_H + \pi_2(r + g + (1 - d)B) \tag{4e}$$
$$\dot{\pi}_3 = -\pi_1 I_E + \pi_3(r + \delta) \tag{4f}$$

The optimal value of a costate variable at any instant of time, $\pi_i, i = 1, 2, 3$, can be interpreted as the imputed value or "shadow price" of an additional light, heavy or remembered ever-heavy user, respectively, at that time (Léonard and Long, 1992).[2] Furthermore, we can interpret $-\dot{\pi}_i, i = 1,2,3$, as the rates at which the value of an additional user depreciates. According to Léonard and Long (1992), the costates for the light and heavy users are negative along the optimal path since $\partial(-\kappa Q - u)/\partial L < 0$ and $\partial(-\kappa Q - u)/\partial H < 0$, i.e., $\pi_1, \pi_2 < 0$. That is, an additional light or heavy user is regarded as a cost. On the contrary, additional memory of ever-heavy users is regarded as a benefit ($\pi_3 > 0$) because it suppresses initiation and imposes no cost (defined by Equation (9f)). The limiting transversality conditions for the costates, given by Conditions (8a-c), hold since (nonnegative and bounded) states and (bounded) costates approach a stable state.

$$\lim_{t\to\infty} e^{-rt}\pi_1(t)L(t) = 0 \tag{8a}$$

$$\lim_{t\to\infty} e^{-rt}\pi_2(t)H(t) = 0 \tag{8b}$$

$$\lim_{t\to\infty} e^{-rt}\pi_3(t)E(t) = 0 \tag{8c}$$

The meaning of the rates of change in the value of an additional light, heavy or ever-heavy user, respectively, described by Equations (4d-f) require some explanation. For instance, the first term in the condition for the rate of change in the value of an additional heavy user (Equation (4e)) is the marginal effect of prevalence (in terms of heavy use) on the instantaneous total costs, while the second term reflects the product of the value of an additional heavy user and the marginal effect of the level of heavy users on its own growth rate. (Equations (4d) and (4f) can be interpreted analogously.)

4 The Equilibrium Strategy and Conditions for Cycling

Let the equilibrium reputation of the drug be denoted by $p = exp[-q\,b/\delta]$, and let the equilibrium rate at which heavy users quit due to optimal treatment measures be represented by Equation (7). Furthermore, we make use of the abbreviations $\Lambda := \partial H/\partial \pi_2(\hat{P})$ and $\chi := \hat{I}_L - a - b = sp(1 + q\,b/\delta) - a - b$.

Proposition 2*: If the rate at which light users persuade non-users to try the drug of concern (s) weighted by the equilibrium reputation of that drug (p) is strictly smaller than the sum of the rates at which light users flow out of their sources state (a+b), i.e. $\Omega := a + b - sp > 0$, the interior steady state, $\hat{P}$, is uniquely defined by the simultaneous solution of the following set of equations.*

[2] The meaning of the costate variables parallels that of the multipliers and dual variables encountered in static optimization. They will be expressed in "value units" per unit of prevalence at time *t*.

$$\hat{L} = \frac{\tau}{\Omega} > 0 \tag{9a}$$

$$\hat{H}(\hat{\pi}_2) = \frac{b\tau}{\left(g + c(-cd\hat{\pi}_2)^{d/1-d}\right)\Omega} > 0 \iff \hat{H}(\hat{\pi}_2) = \frac{b\tau(1-d)(-\hat{\pi}_2)}{(\kappa\mu_H + \hat{\pi}_2(r+gd))\Omega} > 0 \tag{9b}$$

$$\hat{E} = \frac{b\tau}{\delta\Omega} > 0 \tag{9c}$$

$$\hat{\pi}_1(\hat{\pi}_2) = -\frac{\delta(r+\delta)(b\hat{\pi}_2 - \kappa\mu_L)}{rspqb - \delta(r+\delta)(r+\Omega)} < 0 \tag{9d}$$

$$\hat{\pi}_2\left(r + g + c(1-d)(-cd\hat{\pi}_2)^{d/1-d}\right) - \kappa\mu_H = 0 \Rightarrow \hat{\pi}_2 < 0 \tag{9e}$$

$$\hat{\pi}_3(\hat{\pi}_2) = \frac{\delta qsp(b\hat{\pi}_2 - \kappa\mu_L)}{rspqb - \delta(r+\delta)(r+\Omega)} > 0 \tag{9f}$$

It is obvious from Equations (9a-c) that the number of users in steady state and, according to Equation (6), steady state treatment spending scale linearly in the number of innovators, τ, but τ has no effect on the qualitative nature of the solution.

The equilibrium numbers of users as well as control spending decline with a growing flow rate out of light use, a, and a growing deterrent power of negative memories, q, while they increase with a growing rate at which light users recruit non-users, s. Furthermore, the equilibrium number of heavy users and, consequently, treatment spending decrease with an increasing rate at which addicts quit drug abuse, g, but this outflow rate does not affect any of the other equilibrium quantities. The equilibrium number of light users declines with a growing rate at which light users escalate to heavy use, b, because these users contribute to the negative memories instead of acting as dynamic proselytizers as they would have if they maintained in the state of light use. All these results are intuitively appealing. *Less obvious is that the equilibrium number of heavy users, remembered ever-heavy users, and treatment spending decline with a growing rate at which light users escalate to heavy use, b, if and only if* $\hat{\imath}_L > a$*, i.e. if the change in incidence associated with a unit increase in the prevalence of light users is strictly larger than the flow rate out of occasional use.* (See Appendix (A.2)).

Any increment in the constants governing the average or marginal efficiency of treatment spending, c and d, respectively, diminishes the number of heavy users by raising treatment spending, but leaves all other equilibrium quantities untouched. We observe the same effects on heavy use and treatment spending if the social costs per gram consumed, κ, are assumed to be larger than their base case estimations, while an increasing time preference rate, r, which is characteristic for less farsighted decision-makers, raises the equilibrium number of heavy users. In other words, the more one is interested in what happens in the future, the higher will be the rates of treatment spending, since treatment costs money today, but yields benefits throughout the time period the treated individual would have continued to use. (High discount rates discount the benefits of treatment but not the costs.) Equation (9f) reflects the fact that a present-oriented decision-maker regards the equilibrium value of additional memory of ever-heavy users less than a farsighted decision-maker does. Also, the smaller the effect of a given amount of memory, q, the smaller is the equilibrium value of an additional ever-heavy user. On the other hand, if non-users are better at "remembering" past negative consequences of drug abuse, i.e., if δ declines, then the equilibrium numbers of light and heavy users decline and, consequently, treatment spending does too.

We proceed with the analysis of the optimal control problem (Model (3, 4)) in the usual manner. Since $\hat{I}_E = -spq$, the Jacobian, $\hat{\Im}$, of System (4a-f) evaluated at the equilibrium point is given as follows,

$$\hat{\Im} = \left(\begin{array}{ccc|ccc} \chi & 0 & -spq & 0 & 0 & 0 \\ b & -\left(g+\hat{B}\right) & 0 & 0 & \Lambda & 0 \\ b & 0 & -\delta & 0 & 0 & 0 \\ \hline -\pi_1 \hat{I}_{LL} & 0 & -\pi_1 \hat{I}_{LE} & r-\chi & -b & -b \\ 0 & 0 & 0 & 0 & r+g+\hat{B} & 0 \\ -\pi_1 \hat{I}_{LE} & 0 & -\pi_1 \hat{I}_{EE} & spq & 0 & r+\delta \end{array} \right).$$

The eigenvalues of $\hat{\Im}$ can be calculated explicitly, where $\hat{\pi}_2$ is uniquely defined by Equation (9e):

$$\lambda_{1,2}^{3,4} = \frac{1}{2}\left\{ r \pm \left(r+\delta-\chi\right) \pm \sqrt{\left(\delta+\chi\right)^2 - 4spqb} \right\}$$

$$\left.\begin{array}{l} \lambda_5 = r + g + c\left(-cd\hat{\pi}_2\right)^{d/_{1-d}} \\ \lambda_6 = \ -g - c\left(-cd\hat{\pi}_2\right)^{d/_{1-d}} \end{array}\right\} \qquad \lambda_5, \lambda_6 \, real;\ \lambda_5 > 0,\ \lambda_6 < 0$$

According to Feichtinger et al. (1994), we are able to classify the local stability properties of the equilibrium (defined by Equations (9a-f)) as given in Table 1, according to the indicators K and D, defined by Dockner (1985):

$$D := \left(spqb - \delta\chi\right)\left(spqb + \left(r-\chi\right)\left(r+\delta\right)\right) \text{ and } K := \chi\left(r-\chi\right) - \delta\left(r+\delta\right) + 2spqb\,.$$

To assure that the initial conditions for the costates, $\left(\pi_{10}, \pi_{20}, \pi_{30}\right)$, are uniquely determined in dependence of any set of initial conditions along the optimal path, $\left(L_0, H_0, E_0\right)$, we have to specify conditions for the existence of a three-dimensional stable manifold in the six-dimensional phase-space. According to Table 1, the term D has to be positive to exclude instability. *I.e., if the sum of the flow rates out of light use, the forgetting rate, and the time preference rate,* $a+b+\delta+r$*, is larger than the change in incidence associated with a unit increase in the prevalence of light users,* $\hat{I}_L$*, then the trajectories will not diverge from the equilibrium state,* $\hat{P}$*, (defined by Equation (9a-f)).* Then we have to figure out the conditions for monotonous, cyclical or oscillating behavior of the trajectories.

Table 1: Classification of equilibria according to K and D

Conditions	Equilibrium of the linearized canonical system
$K<0$ $0< D \leq \left(K/2\right)^2$	Saddlepoint stability due to real eigenvalues, where three are negative and three are positive; local *monotonicity*, i.e. monotonous approach of the equilibrium on a three-dimensional stable manifold.
$D > \left(K/2\right)^2$ $D > \left(K/2\right)^2 + r^2\, K/2$	Saddlepoint stability, but we observe transient *oscillations* on a 3D manifold due to two real eigenvalues (one positive, one negative) and two pairs of conjugate complex eigenvalues, one pair having negative real parts.
$D < 0$	*Instability* except for a 1D manifold associated with a single negative real eigenvalue, while all other eigenvalues have positive real parts. Unstable in all but one directions.
$K > 0, D > 0$ $D = \left(K/2\right)^2 + r^2\, K/2$	The system undergoes a *Hopf bifurcation* due to the occurrence of two purely imaginary eigenvalues, where the imaginary axis is crossed with non-zero velocity. Note that we are assured of the existence of a cycle, but not of its stability properties.
$D > \left(K/2\right)^2$ $D < \left(K/2\right)^2 + r^2\, K/2$	Locally *unstable spirals* due to two real eigenvalues (one positive, one negative) and two pairs of conjugate complex eigenvalues with positive real parts.

We are particularly interested in prerequisites for the possible occurrence of a chain of drug. This process might happen as follows. Initially there is little history of heavy use, so the drug's reputation is benign. The resulting contagious spread of initiation leads to near exponential growth in the number of users. Although negative experiences begin to accumulate, they are overwhelmed by the rapid increase in the number of light users. Eventually, however, accumulated experience with the drug dampens initiation. As long as the physical and psychological suffering, adverse health consequences, financial despair, etc. are continuously kept in the minds of non-users, the number of initiates will remain low and prevalence will diminish. The reduced prevalence also implies reduced flow into the pool of ever heavy users. Though this situation is desirable at first sight it gives rise to the process of forgetting negative experiences (made by older and maybe socially distant people) and the bad reputation of the drug decays. When the rumors of "that bad drug" have faded, the epidemic could start again. In the absence of proactive prevention programs, a generation unaware of the dangers of abuse might be very innovative with respect to experimental drug use and could repeat the mistakes of their forebears.

The following inequality summarizes the conditions for cyclical system behavior in terms of the time preference rate, $r > 0$, the "forgetting-rate", δ, the rate at which light users quit, a, the rate at which occasional users escalate to heavy use, b, the rate at which light users attract non-users, s, and the measure of the deterrent power of negative memories, q,

$$G(r,\delta,a,b,s,q) = 4spqb\,(r+\delta-\chi)^2 - [\chi(r-\chi) - \delta(r-\delta)]^2 > 0\,. \qquad (10)$$

Condition (10) has a rather interesting implication. One scenario which might favor the occurrence of a chain of drug epidemics is that decision-makers are not farsighted enough, which may formally be written as,

$$r \gtrless \frac{\delta^3-\chi^3+(4spqb+\delta\chi)(\delta-\chi)}{4spqb-(\delta+\chi)^2} \pm \sqrt{\left(\frac{\delta^3-\chi^3+(4spqb+\delta\chi)(\delta-\chi)}{4spqb-(\delta+\chi)^2}\right)^2 - 4spqb\cdot\left(\delta-\chi^2\right)+\left(\delta^2-\chi^2\right)^2}\,. \qquad (10^*)$$

5 System Behavior for the Modeled Cocaine Problem

To visualize the system behavior by depicting "stability regions", we have to specify some of the parameter values. We do this for the cocaine epidemic currently observed in the United States.

5.1 The Specification of the Parameters

Average annual consumption rates for light and heavy users under base case conditions (μ_L and μ_H) are taken from Everingham and Rydell (1994), i.e. $Q = 16.42L + 118.93H$. We discount at 4% per year to be consistent with other work in this area (e.g., Rydell and Everingham, 1994).

The flow rates a and b are computed as the time-continuous equivalents of the Everingham-Rydell estimates (1994, p.43). For example, they estimate that 15% of light users quit use each year and 2.4% escalate to heavy use, so we set $a = -ln[1-0.15] \cong 0.163$ and $b = -ln[1-0.024] \cong 0.024$, to three decimal places. Additionally, Everingham and Rydell (1994, p.42) gave ranges of parameters for which they found good fits to historical data reported by the National Institute on Drug Abuse (1991), which are $a \in (0.157, 0.168)$ and $b \in (0.02, 0.03)$.

According to Behrens et al. (1999), reasonably good fits to the historical data can be achieved for a wide range of parameters s and q, as long as larger values of s are associated with larger values of q. We choose the set $\{s = 0.61, q = 7.25\}$. This makes the constant q equal to the ratio of the average heavy user's consumption rate to that of the average light user. The corresponding size of parameter $s=0.61$ can be interpreted as indicating that approximately two light users would "persuade" one non-user per year to try cocaine, if there was no memory of abuse giving cocaine a bad reputation.

Rydell and Everingham (1994, p.38) report a societal cost estimate (based on Rice et al., 1990) of \$19.68 billion for cocaine in 1992. This cost is associated with 291 metric tons of consumption, yielding a value of κ=\$67.6 per gram. This places very low estimates on costs associated with crime (e.g. no pain or suffering costs), so we follow Behrens et al. (forthcoming) and take as our base estimate a figure that is κ=\$113 per gram.

To be consistent with Tragler et al. (1997) we assume the diminishing returns to treatment parameter d = 0.6. Rydell et al. (1996) report the number of cocaine treatments in 1992 was 548,000, base year spending was \$930 million per year, and 13% of those treated exited heavy use, with one-third of those exits being to a state of no use. Hence, the parameters of the function in Equation (2) must be chosen such that $\beta(\$\,930\, million\, /\, 548{,}000\, heavy\ users) = 0.045$, which implies $c = 5\times10^{-4}$.

Table 2: Base case parameter values for US cocaine epidemic

Parameter	Value	Description
a	0.163	annual rate at which light users quit
b	0.024	annual rate at which light users escalate to heavy use
g	0.062	annual rate at which heavy users quit
s	0.610	annual rate at which light users attract non-users
q	7.250	constant which measures the deterrent effect of heavy use
τ	$5\text{x}10^4$	number of innovators per year
μ_L	16.42	average annual consumption rate for light users (grams per year)
μ_H	118.93	average annual consumption rate for heavy users (grams per year)
c	$5\text{x}10^{-4}$	constant measuring the average efficiency of treatment
d	0.600	constants measuring the marginal efficiency of treatment
r	0.040	annual discount rate (time preference rate)
κ	113	social cost of cocaine consumption (dollars per gram)

5.2 The consequences of forgetting and the benefits of prevention

We do not have data to estimate the "forgetting rate", δ, so we conduct sensitivity analysis with respect to this parameter. Hence, in Figure 2 we see how the qualitative system behavior depends on two parameters: (1) the discount rate, r, which governs how present- vs. future-oriented the society is and (2) the parameter δ which governs how quickly ever-heavy users are "forgotten". The higher δ is, the sooner past heavy use is forgotten and the less negative is the drug's reputation.

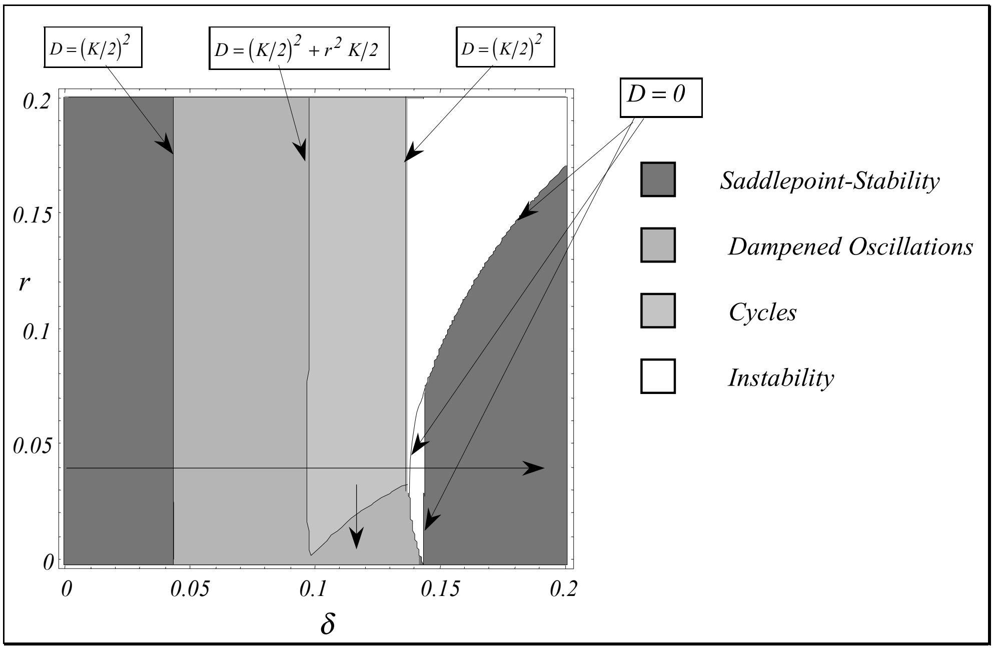

Figure 2. Stability regions for any combination of r and δ values for $a = 0.163, b = 0.024, s = 0.61$ and $q = 7.25$. (The horizontal arrow indicates the effect of a society forgetting the negative consequences of drug abuse faster as δ becomes larger, where the decision-maker discounts total societal cost with 4% per annum.)

- When δ is small ($0<\delta \leq 0.043$), we are in the saddlepoint region (where the trajectories approach the equilibrium monotonously on the 3D stable manifold).
- As δ increases beyond 0.043 (up to 0.094), we cross the region where the trajectories spiral into the steady state.
- The transition from oscillation to cyclic system behavior happens due to a (supercritical) Hopf-bifurcation at about $\delta = 0.094$.[3] Close to the Hopf-bifurcation point, $D=(K/2)^2+r^2K/2$, the cycles "grow" with the size of $\sqrt{\delta-\delta_{Hopf}}$, where $\delta_{Hopf} \approx 0.094$ denotes the bifurcation values of parameter δ. Hence, when δ is between 0.094 and about 0.138 ($(K/2)^2+r^2K/2>D>(K/2)^2$), there is cycling unless r is small. The vertical arrow shows the effect of becoming more far-sighted (r smaller) for a given level of memory ($\delta = 0.12$). When r is small, the system has oscillations, because control is altered in a way that eliminates cycling in the "*H*-system". (As obvious from Behrens et al., 1999, the "*L-E* system" must still be cycling because if treatment cannot affect the *L-E* system, surely the discount parameter cannot do so.) I.e. future orientation changes the pattern of treatment spending in a way that

[3] Since we are assured of the stability properties of the cycles and the precise features of the bifurcation point at *D=0* only by numerical computation, the terms *supercritical* and *transcritical* are put in parentheses.

eliminates cycling and makes oscillation towards a stable level of prevalence possible.

- The regime 0.138<δ<0.143 exhibits unstable system behavior.
  - We have double-eigenvalues, $\lambda_{1,2} = \lambda_{3,4} = \frac{1}{2} \cdot \left(r \pm \sqrt{r^2 - 2K} \right) \neq 0$, at $D = (K/2)^2$. Hence, for $0 < D < (K/2)^2$ the cycles break, and the trajectories spiral off to infinity.
  - If we further increase δ until $D = 0$, a (transcritical)[3] bifurcation occurs. (At $D = 0$ we have $\lambda_1 = r$, $\lambda_{2,4} = \frac{1}{2} \cdot \left(r \pm \sqrt{r^2 - 4K} \right)$, $\lambda_3 = 0$).
  - Inside the small region where $D < 0$ different types of instability (spiral or monotonous path towards infinity) may occur. (These types of behavior depend on whether $K > 0$ or $K < 0$ and $D > (K/2)^2 + r^2 (K/2)$ or vice versa, e.g. for $(r/2)^2 \cdot (4 - rK) < D < 0$ all eigenvalues become real (5 positive, one negative) which leads to monotonous diverging trajectories.)

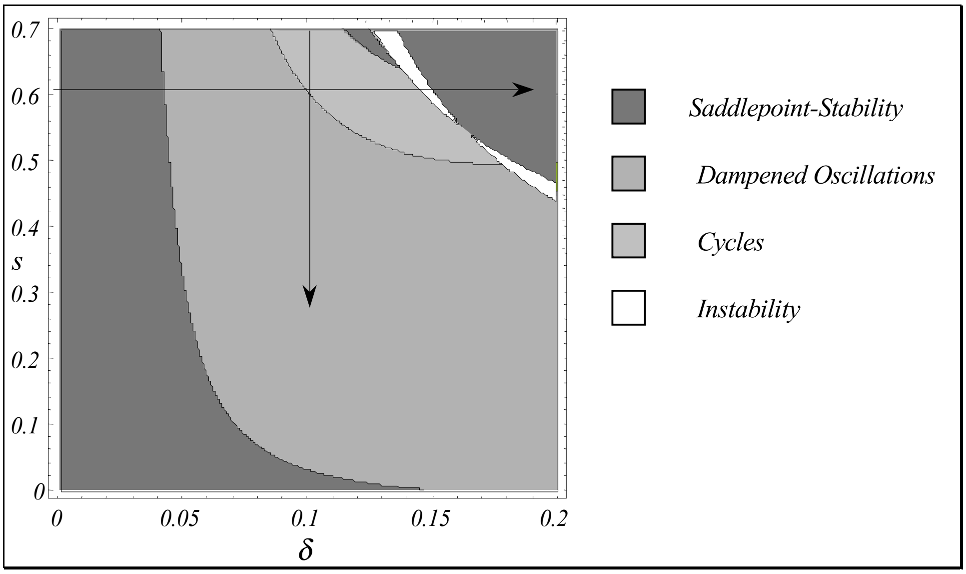

Figure 3. Stability regions for any combination of s and δ values for $a = 0.163,\ b = 0.024$, $r = 0.04$ and $q = 7.25$.

♦ At the second branch of the $D=0$ line, a (transcritical) bifurcations occurs again, indicating an exchange in stability behavior due to the fact that λ_3 becomes negative for $D>0$. (At $D=0$ we have $\lambda_1=r$, $\lambda_{2,4}=\frac{1}{2}\cdot\left(r\pm\sqrt{r^2-4K}\right)$, $\lambda_3=0$). Note that $\lambda_{2,4}$ are real since now $K<0$ and $D>(K/2)^2+r^2(K/2)$. Consequently, for $D>0$ we have three negative real (and three positive real) eigenvalues which ensures monotonous approach towards the equilibrium.

Needless to say that the size and the shape of the stability regions, depicted in Figure 2 (and 3), depend on the parameter values of *a, b, s,* and *q* (but not on the values of τ, *c, d, g,* κ, μ_L, and μ_H). While changes in the values of the flow rates, *a* and *b*, within the range described as plausible by Everingham and Rydell (1994) do not drastically change the basic pattern observed here[4], one sided deviations of the parameters *s* and *q* from their base case values do. These changes may be used to approximate the effect of prevention spending by simple sensitivity analysis. I.e., *for large values of the forgetting rate,* δ*, the principal policy insight is that prevention is necessary to avoid cycles*. The reason for thinking that preventive measures might increase the threshold value of δ at which cycles begin is that Behrens et al. (1999) found that cycles also emerged when the parameter *s* was large, and prevention is equivalent to reducing *s*.

The effect of prevention is shown by the vertical arrow in Figure 3. For the base-case set of parameter values, if δ was larger than 0.094 then the "*L-E* system" would cycle if there was no prevention. Decreasing *s*, without adapting *q*, removes the cycling system behavior and reduces prevalence, both in overall and in equilibrium terms. That is, moderating the contagious aspect of initiation reduces the likelihood of cycles and instability.

Increasing *q*, without adapting *s*, produces similar effects. The smaller the deterrent power of memories of drug abuse, the more likely society is to wind up with a chain of drug epidemics. Thus, weakening the feedback effect through past memories either by having memories fade faster (δ larger) or be less poignant (*q* smaller) increases the likelihood of cycling.

6 Conclusions

The model presented here yields several conclusions. First, it underlines Musto's (1987) hypothesis that cycles of drug use may arise when the current generation of youth no longer remembers the adverse experiences of their forebears. In addition, this scenario is favored by myopic decision makers. In other words, forgetting the past and not caring about the future, which is characteristic of present-oriented societies, raises the danger that drug epidemics will occur again and again, i.e. that the future recreates the past. Second, we find that treatment programs are useful for reducing social costs.

---

[4] Note that a *b*-value at the upper bound of the Everingham-Rydell interval increases the "oscillation region" and shifts the "instability region" down and to the left.

Also they are not necessarily counterproductive at the onset of an epidemic (cf. Behrens et al., forthcoming, 1999) if the deterrent effect of a drug is based not on the current number of heavy users but rather on some (imperfect) memory of all individuals who ever abused illicit drugs. Third, prevention programs that soften the contagious character of drug initiation would lower the threat to a society of repeating the past, as would programs that raise the deterrent effect spread by a single ever-heavy user.

Though future research may be done in several directions, we want to describe a rather promising one – especially from the mathematical point of view. Remember that - for values of the forgetting rate high enough - we observed "instability" for the model discussed in this paper. A "DNS-threshold" (also denoted as "Skiba point", see Skiba, 1978)[5] separates two basis of attraction. In that case initial states may lead to different long run equilibria. In our case it might happen that a substantial reduction of initiation, e.g. by prevention, might lead to an extinction of the drug epidemics. Research on that issue is under way.

[5] The occurrence of multiple equilibria which are separated by thresholds is quite common in economic models. Skiba (1978) discussed such critical values, but his proof is incomplete. A first proof of existence was given by Dechert and Nishimura (1983).

Appendix

A.1 Proof of Proposition 1

A sufficient condition for the Hamiltonian, H, to be (strictly) concave with respect to the states, L, H, and E, and the control, u, is that the Hessian of the Hamiltonian (Equation 5) is negative semidefinite (definite) (Feichtinger and Hartl, 1986). Therefore, we construct this matrix,

$$D^2\mathsf{H}(L,H,E,u)=\begin{pmatrix} \pi_1 I_{LL}<0 & 0 & \pi_1 I_{LE}>0 & 0 \\ 0 & -\pi_2(1-d)\beta_H<0 & 0 & -\pi_2(1-d)\beta_u>0 \\ \pi_1 I_{LE}>0 & 0 & \pi_1 I_{EE}<0 & 0 \\ 0 & -\pi_2(1-d)\beta_u>0 & 0 & -\pi_2\beta_{uu}H<0 \end{pmatrix},$$

and calculate its minors, i.e. the determinants of the sub-matrices. Note that the derivatives of the initiation function take the forms $I_{LL}=(E/L)^2 I_{EE}$ and $I_{LE}=-(E/L)I_{EE}$.

$$M_1=\underbrace{\pi_1}_{<0}\underbrace{I_{LL}}_{>0}\quad<0 \qquad M_2=\underbrace{-\underbrace{\pi_1\pi_2}_{>0}\underbrace{I_{LL}}_{>0}\underbrace{(1-d)}_{>0}\underbrace{\beta_H}_{<0}}_{<0}\quad>0$$

$$M_3=\underbrace{\pi_1^2\pi_2(1-d)\beta_H}_{>0}\underbrace{\left[I_{LE}^2-I_{LL}I_{EE}\right]}_{=0}\quad=0 \qquad M_4=\underbrace{\pi_1^2\pi_2^2(1-d)^2\beta_u^2}_{>0}\underbrace{\left[I_{LE}^2-I_{LL}I_{EE}\right]}_{=0}\quad=0$$

Therefore, the Hamiltonian of optimal control Problem 3, 4 is jointly concave (but not strictly concave) in states and control, which implies that u defined by Equation 6 determines optimal treatment spending. I.e. the necessary conditions are sufficient.

A.2 Equilibrium Behavior With Respect to b

The equations for the steady states of heavy and ever-heavy users as well as the equation for treatment spending contain the term

$$\vartheta(b)=b(a+b-s\,exp[-q\,b/\delta])^{-1},$$

treated as a function of the rate at which light users escalate to heavy use, b. The marginal change in ϑ with respect to b, determined by

$$\frac{\partial\vartheta}{\partial b}=\frac{a-s\,exp[-q\,b/\delta](1+q\,b/\delta)}{(a+b-s\,exp[-q\,b/\delta])^2}=\frac{a-\hat{I}_L}{(a+b-s\,exp[-q\,b/\delta])^2},$$

is larger than zero if and only if $\hat{I}_L$ is smaller than a. Therefore, the equilibrium number of heavy and ever-heavy users as well as equilibrium treatment spending decline with a growing rate at which light users escalate to heavy use, b, if and only if $\hat{I}_L>a$, i.e. if the change in incidence associated with a unit increase in the prevalence of light users, $\hat{I}_L$, is strictly larger than the flow rate out of occasional use, a.

References

Behrens, D.A., 1998, The US cocaine epidemic: An optimal control approach to heterogeneous consumption behavior (Ph.D. Thesis, Vienna University of Technology).

Behrens, D.A., J.P. Caulkins, G. Tragler and G. Feichtinger, forthcoming, Optimal control of drug epidemics: Prevent and treat - but not at the same time?, Management Science.

Behrens, D.A., J.P. Caulkins, G. Tragler, J.L. Haunschmied and G. Feichtinger, 1999, A dynamical model of drug initiation: Implications for treatment and drug control, Mathematical Biosciences 159, 1 – 20.

Caulkins, J.P., C.P. Rydell, S.S. Everingham, J. Chiesa and S. Bushway, 1999, An ounce of prevention, a pound of uncertainty: The cost-effectiveness of school-based drug prevention programs (RAND, Santa Monica, MR-923-RWJ).

Dechert, W.D. and K. Nishimura, 1983, A complete characterization of optimal growth paths in an aggregated model with a non-concave production function, Journal of Economic Theory 31, 332 – 354.

Dockner, E., 1985, Local stability analysis in optimal control problems with two state variables, in: G. Feichtinger, ed., Optimal control theory and economic analysis 2 (North Holland, Amsterdam).

Everingham, S.S. and C.P. Rydell, 1994, Modeling the demand for cocaine (RAND, Santa Monica, MR-332-ONDCP/A/DPRC).

Feichtinger, G. and R.F. Hartl, 1986, Optimale Kontrolle Ökonomischer Prozesse - Anwendungen des Maximumprinzips in den Wirtschaftswissenschaften (deGruyter, Berlin).

Feichtinger, G., A. Novak and F. Wirl, 1994, Limit cycles in intertemporal adjustment models, Journal of Economic Dynamics and Control 18, 353 – 380.

Léonard, D. and N.V. Long, 1992, Optimal control theory and static optimization in economics (Cambridge University Press, Cambridge).

Michel, P., 1982, On the transversality condition in infinite horizon optimal control problems, Econometrica 50, 975 – 985.

Musto, D.F., 1987, The American disease (Yale University Press, New Haven).

National Institute on Drug Abuse, 1991, Overview of the 1991 household survey of drug abuse, Press release, December.

Rice, D.P., S. Kelman, L.S. Miller and S. Dunmeyer, 1990, The economic costs of alcohol and drug abuse and mental illness: 1985 (Institute for Health and Aging / University of California, San Francisco).

Rydell, C.P., J.P. Caulkins and S.S. Everingham, 1996, Enforcement or treatment? Modeling the relative efficacy of alternatives for controlling cocaine, Operations Research 44(6), 687 – 695.

Rydell, C.P. and S.S. Everingham, 1994, Controlling cocaine. Supply versus demand programs (RAND, Santa Monica, MR-331-ONDCP/A/DPRC).

Skiba, A.K., 1978, Optimal growth with a convex-concave production function, Econometrica 46, 527 – 539.

Tragler, G., J.P. Caulkins and G. Feichtinger, 1997, The impact of enforcement and treatment on illicit drug consumption (Working paper, Vienna University of Technology); submitted to Operations Research.