

HIGH AND LOW FREQUENCY OSCILLATIONS IN DRUG EPIDEMICS

Jonathan P. Caulkins¹, Alessandra Gragnani², Gustav Feichtinger³, Gernot Tragler⁴

¹ *H. John Heinz III School of Public Policy Management, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, U.S.A., email: caulkins@andrew.cmu.edu*

² *Department of Electronics and Information, Politecnico of Milan, Via Ponzio 34/5, 20133 Milan, Italy, email: gragnani@elet.polimi.it*

³ *Institute for Mathematical Methods in Economics, Vienna University of Technology, Argentinierstrasse 8/105-4, A-1040 Vienna, Austria, email: or@eos.tuwien.ac.at*

⁴ *Institute for Mathematical Methods in Economics, Vienna University of Technology, Argentinierstrasse 8/105-4, A-1040 Vienna, Austria, email: tragler@eos.tuwien.ac.at*

Abstract: We extend the two-dimensional model of drug use introduced in Behrens *et al.* [1999, 2000, 2002] by introducing two additional states that model in more detail newly initiated (“light”) users’ response to the drug experience. Those who dislike the drug quickly “quit” and briefly suppress initiation by others. Those who like the drug progress to ongoing (“moderate”) use, from which they may or may not escalate to “heavy” or dependent use. Initiation is spread contagiously by light and moderate users, but is moderated by the drug’s reputation, which is a function of the number of unhappy users (recent quitters + heavy users). The model reproduces recent prevalence data from the U.S. cocaine epidemic reasonably well, with one pronounced peak followed by decay toward a steady state. However, minor variation in parameter values yields both long-run periodicity with a period akin to the gap between the first U.S. cocaine epidemic (peak ~1910) and the current one (peak ~1980), as well as short-run periodicity akin to that observed in data on youthful use for a variety of substances. The combination of short- and long-run periodicity is reminiscent of the elliptical burstors described by Rubin and Terman [2002]. The existence of such complex behavior including cycles, quasi periodic solutions, and chaos is proven by means of bifurcation analysis.

Running title: High and low frequency oscillations in drug epidemics

1. Introduction

Drug use varies over time in ways that have long been described in terms of “epidemics” [Brill and Hirose, 1969]. This metaphor is appropriate even though there is not literally a pathogen, as with HIV or cholera, because most initiation occurs through contact with current users, not at the urging of drug sellers [Kaplan, 1983].

Drug epidemics can be recurrent. There were cocaine and heroin epidemics in the U.S. from roughly 1880 to 1920 that essentially died out [Musto, 1987; Spillane, 2000]. Both drugs were used at very low levels for the next 50 years, until the current epidemics began in the late 1960s and early 1970s [Courtwright, 2001]. More such patterns have been documented for alcohol, for which the historical record is longer [Rorabaugh, 1979].

Historians and drug policy analysts explained this pattern qualitatively in terms of contagious spread by users who are happy with the drug combined with a delayed negative feedback from heavy users whose visible sufferings suppresses initiation [Musto, 1987; Kleiman, 1992]. The delay stems from the typical individual-level patterns of drug use which progresses through stages of use. Addiction and other negative consequences are associated with the later stages. Everingham and colleagues [Everingham *et al.*, 1994; Rydell & Everingham, 1994; Everingham *et al.*, 1995] pioneered mathematical models that distinguish between “light” and “heavy” users. These were extended by Behrens *et al.* [1999, 2000, 2002] to include positive and negative feedback on initiation from light and heavy use.

The empirical evidence suggests, however, that in addition to these “long-wave” fluctuations with periods on the order of 75 years, there are also shorter-term oscillations. Surveys of households and youth in the U.S. show that initiation of marijuana and other drugs rose to a peak in 1979, fell to a trough in 1991 or 1992, and has rebounded since [Johnson *et al.*, 1996; SAMHSA, annual]. Likewise, Marotta [1992] and Preti *et al.* [2002] collectively

report data on Italian overdose deaths from 1979 – 2000, a series that peaked in 1991, declined, peaked again in 1996, and fell again.

Data on self-reported heroin use in Australia suggest a possible explanation [Kaya *et al.*, 2004]. Initiation and quitting varied over time, with some hint of recurrent peaks roughly every three years, against a backdrop of an overall upward trend until the onset of the heroin drought [Dietze & Fitzgerald, 2002; Weatherburn *et al.*, 2003]. Curiously, quitting in a calendar year was strongly positively correlated with the number initiating in that year. Exogenous factors (such as price variation) could make use more appealing in some years than others, but they should cause high initiation to be correlated with low quitting, not the opposite. Furthermore, about half of survey respondents quit in the same year they first tried. This may be because many people do not enjoy their early experiences with heroin [Kaplan, 1983].

These data are consistent with the following simple story. People who are considering trying an illicit drug do not know what to expect. Some like it. Some do not. Both groups tell their friends what they discovered. Those who like it continue to use regularly and continue to tell their friends about the drug's virtues. Some of them eventually become addicted "heavy" users who serve as negative advertisements for the drug's dangers, but most persist in "moderate" use for some time. Those who did not like the drug quit rapidly and report to associates that the drug is not worth trying, but only for a time. A year after they tried and disliked the drug, they have had time to tell all their friends what they thought. They have no new information about the drug, and so stop contributing to potential initiates' decisions.

This paper extends Behrens *et al.*'s models to include this story by elaborating on these early reactions to drug initiation, creating a structure that yields both low and high frequency variation in drug use and a wealth of interesting mathematical results in the process.

2. Model Formulation

Behrens *et al.* [1999] describe a Markov model of cocaine demand in the U.S. that differentiates between light (L) and heavy (H) users, with parameters a , b , and g representing the annual rate at which light users quit, light users escalate to heavy use, and heavy users quit, respectively. The model incorporated feedback from the current prevalence, or level, of use to initiation into new use. Some initiation was “spontaneous” (e.g., because of immigration) but most occurred through interactions with current users.

Behrens *et al.* [1999] moderated the rate at which current users recruited new initiates by the negative “reputation” of the drug. The reputation was modeled as a negative exponential function of the relative number of current heavy (H) and light (L) users. For simplicity, the population of “susceptibles” was not modeled explicitly because, except for marijuana, illicit drugs are used by only a modest fraction of the population so literally exhausting the pool of potential initiates is not usually the constraining factor on epidemic spread. This LH model had interesting dynamics and generated policy insights such as highlighting the benefit of varying the mix of interventions dynamically over the course of a drug epidemic [Behrens *et al.*, 1999; 2000].

It was subsequently hypothesized that ever-heavy users (E) and/or a memory of accumulated years (Y) of heavy use might contribute to a drug’s negative reputation, so the LH model was extended to an LHE [Behrens *et al.*, 2002] and LHY [Behrens *et al.*, in submission] framework. The LHE and LHY models offer some advantages over the LH framework. Nevertheless, we take the LH model as our point of departure because we need to add two more states at the “front end” of the model (shortly after initiation) and we prefer to work with a four state rather than a five state model.

To be specific, in the present model, new initiates continue to enter the light state (L), but they quickly split into those who like the drug and progress to persistent moderate use (M)

and those who do not like it and quit (Q). In keeping with prior models, transitions between states are all governed by simple constant per capita flow rates. In particular, the state Q captures the flow rate a out of the L state and decays at some rate μ because the quitters' negative experiences become old news fairly quickly. We define Q as people who, on average, have quit within the last year, so the exit rate $\mu = 1$. As in Behrens *et al.* [1999], the parameters b and g will continue to denote the rates of escalation from L to the "next" use state and leaving from that next state, which is now moderate use (M) not heavy use. Furthermore, departures from that M state include both escalation to heavy use (at rate h) and quitting (at rate $g - h$). Exit from the heavy use state occurs at rate δ , which is numerically akin to the flow parameter g 's value in the LH model.

The only non-linearity is reputation's effect on initiation. As in Behrens *et al.* [1999], initiation is a constant (τ) plus a constant (s) times the number of "content" users (now L + M, not just L) times a negative exponential reputation term. The reputation term here is like that in the LH model except that it is not only dependent on heavy users (H) but also on people who recently tried but disliked the drug (Q) that give the drug a bad reputation. The model's structure is illustrated in Fig. 1.

The two types of disaffected users (H and Q) need not be equally forceful voices. We know of no empirical work that bears on this issue, but note that dependent users are typically socially removed from potential initiates; they are older, poorer, and socially marginalized. In contrast, people who have recently tried the drug for the first time are likely to be socially close in age and other demographic respects to potential initiates. Hence, we would expect that Q might be weighted more heavily in this reputation term than H. Note: this does not mean that one person becoming addicted suppresses initiation less than does one person trying the drug and not liking it because addiction careers last many years. (I.e., δ , the flow rate out

of state H is much smaller than μ , the flow rate out of state Q.) Hence, the duration of the suppressive effects of someone becoming addicted is much longer, even if the effect per unit time is less pronounced.

We make one other change to the reputation term in light of Winkler *et al.* [2004]. With the original LH reputation function, in stages of the epidemic with few light users, the negative exponential term becomes overwhelmingly powerful, shutting off almost all initiation. There has in fact been a substantial decline in U.S. cocaine initiation (both in total and per light user) as the drug gained a negative reputation (from 2+ million per year around 1980 to an average of some 600,000 in the 1990s), but it has not approached zero [Caulkins *et al.*, 2004]. Hence, we add a constant (K) to the number of light users in the reputation term, in effect placing a floor under this term.

This discussion suggests investigating the following descriptive four-state model of light (L), moderate (M), and heavy (H) use, supplemented by the population of recent initiates who disliked the drug and quit (Q), which we refer to as the LMHQ model for obvious reasons. (Note from now on we omit explicit dependence of the state variables on time for sake of simplicity.)

$$\begin{aligned}
 \dot{L} &= I(L, M, H, Q) - (a + b)L \\
 \dot{M} &= bL - gM \\
 \dot{H} &= hM - \delta H \\
 \dot{Q} &= aL - \mu Q
 \end{aligned}
 \tag{1}$$

where the endogenous initiation into light use is specified by the function

$$I(L, M, H, Q) = \tau + s(L + M) \exp\left(-q \frac{H + \alpha Q}{L + K}\right),
 \tag{2}$$

3. Numerical Analyses

3.1 Simulating the current U.S. cocaine epidemic

To visualize the model's behavior we have to specify parameter values. We do this initially for the cocaine epidemic currently observed in the United States, because of its magnitude and because data on that epidemic are relatively abundant. The derivation of these parameter values is described in Appendix A.1 and the values themselves are summarized in Table 1.

To simulate the current epidemic, we initialize the model with 1977 conditions. In particular, denominating prevalence in millions of users, we set $H(1977) = 0.413$ [Caulkins *et al.*, 2004] and $L(1977) + M(1977) = 3.533$ [Everingham & Rydell, 1994]. Everingham and Rydell do not distinguish between the L and M states as we do, but household survey data suggest a split of roughly $L(1977) = 1$ and $M(1977) = 2.533$.¹ We have no historical data that bears directly on $Q(1977)$, so we set $Q(1977) = 0.351$, which is in line with the 1979 household survey result that roughly 30% of people who reported initiating in the past year said they had not used in the past 6 months. Fig. 2 shows the model simulation corresponding to the base case parameter values and these initial conditions.

With these parameter values and initial conditions, the model reproduces many of the key features of the recent U.S. cocaine epidemic. In particular, the sequence of peaks is correct, with initiation peaking first, then the number of "content" (L + M) users, then total past-year prevalence, and finally the number of heavy users. Furthermore, the prevalence peaks match the historical data quite well. For example, the modeled peak in past-year users who are not heavy users (i.e., L + M users) was 11.7 million at $t = 1979.16$, whereas the actual peak was

¹ The earliest household survey whose data are available for analysis was in 1979, not 1977 (analyzed on line at <http://www.icpsr.umich.edu/SAMHDA/das.html>). According to that survey, 24.8% of past-year cocaine users in 1979 had been using for a year or less. We round that up to $1/3.533 = 28.3\%$ because (1) the outflow rate from the L state is $0.5 + 0.42 = 0.92$, implying average dwell times of $1/0.92 = 1.09$ years, not just one year and (2) presumably the proportion of recent initiates was slightly higher two years earlier, in 1977 vs. 1979. These adjustments are rough, so we focus on $L(0) = 1$ just to take a round number.

probably at about that level in 1980 or 1981.² Likewise the modeled peak in heavy users was 2.73 million at $t = 1988.85$ vs. the best actual estimate of 2.6 million in 1989 [Caulkins *et al.*, 2004].

Modeled initiation peaks at roughly the right time ($t = 1978.14$ vs. 1980 in the historical data), but at a too high level and then modeled initiation drops off too quickly, falling to near zero levels, whereas historical initiation trends fell more slowly and never fell below 0.5 million per year.

Mathematically the reason for the disparity is that the negative reputation feedback from heavy users (H) seems too strong. This is similar to what Winkler *et al.* [2004] found. It may be due to the exponential form of the negative reputational feedback. Alternately, the problem may be the implicit assumption that all heavy users generate a feedback to reputation and initiation.

Certainly some heavy users behave in ways that are consistent with Musto's [1987] and Kleiman's [1992] story of serving as negative advertisements that deter initiation, but other heavy use may be essentially invisible. That is most clearly the case for middle-class heavy users who remain functional [Waldorf *et al.*, 1991], but even the plight of some highly dysfunctional users may not be visible to youth contemplating drug initiation precisely because heavily dependent users are often socially marginalized.

In terms of the model structure in Fig. 1, this suggests possibly adding a fifth state to distinguish between heavy users whose presence does influence the reputation of the drug (current H state) and heavy users who are invisible to potential initiates. Incorporating such a model extension does not require any new analysis because (1) the new, fifth state has no effect on any of the other stocks and flows and (2) the numbers of both types of heavy users

² The household survey detected 9.1 million past-year users who were not heavy in 1979 and 10.9 million in 1982. There were no surveys in 1980 or 1981, but levels were increasing before 1979 and decreasing after 1982, so the actual peak may well have been a bit larger than either of these figures and occurred in 1980 or 1981.

would always be directly proportional to each other, with the proportionality constant determined by the proportion escalating into visible heavy use vs. invisible heavy use. For example, if one believed that half of heavy users were visible and half were invisible, one would simply divide the flow rate h in half and remember that for purposes such as planning demand for drug treatment, the total number of heavy users was exactly twice the visible number of heavy users tracked in the model's H state.³ Hence, in the next section we explore implications of reducing the value of parameter h .

3.2 Effect of reducing parameter h

Reducing parameter h from its base case value (0.07), while holding the other parameters constant, models the possibility that some heavy users might be invisible to prospective initiates. We consider the effects of this change both on what would have been observed to date (roughly the first 25 years after the 1977 initial conditions described above) and also on longer-run behavior.

Starting with the initial conditions defined above and looking only at the system's transient behavior in the first roughly 25 years, the first change one observes as h shrinks is that there is an extra "bounce" in the trajectories of initiation, recent initiates (L), and recent quitters (Q). Rather than just increasing to a peak and then decreasing, they display damped oscillation with a relatively high frequency (period on the order of about 10 years, $h = 0.017$). With further reductions in h , the level of moderate use (M) also displays this damped oscillation ($h = 0.014$).

³ This equivalence works because we parameterized flow g to be the total exit from the M state, with h being a subset of that flow, rather than estimating g as specifically the flow from M to non-use. So changing h does not change our estimate of parameter g . Naturally one could imagine more complicated models with flows back and forth between the visible and invisible forms of heavy use, but we have no data to inform estimates concerning such flows.

Damped oscillation is not uncommon in time series for drug prevalence across a range of drugs and cultures. Simple “overshoot” in which the peak is followed by some ebbing is the norm. Subsequent “undershoot” and more generally damped oscillation is not always observed, but it is far from uncommon [Caulkins, 2005].

With further reductions in h ($h = 0.008$) oscillations begin to manifest in heavy use, and the oscillations for almost all states are no longer purely decreasing in amplitude over time (damped).

Looking at the long-run system dynamics, reducing h generates more and more complex behavior. After decreasing h from 0.07 to 0.0093, the behavior becomes cyclic with a relatively high frequency (period of about 7 years), as shown in Fig. 3a. A further reduction in h produces quasi periodic behavior characterized not only by high-frequency oscillations but also by low-frequency ones (Figs. 3b and 3c). If one drew an envelope around these high-frequency oscillations one would see what looks like one big cycle of growth and then decay (low-frequency, about 100 years) comprised of many of these higher-frequency oscillations. Moreover, with a smaller h the magnitude of oscillations becomes different, growing for the first half dozen or so times before shrinking and then growing again (compare Figs. 3b and 3c). Sometimes the shrinkage is gradual, but other times it is abrupt, with the highest peaks coming just before periods with very little drug use.

Further reductions in h primarily has the effect of increasing the length of this long cycle, but then with further reductions ($h = 0.006$) one sees that the behavior is no longer quasi periodic and finally becomes chaotic, as shown in Fig. 3d and in peak-to-peak plots (not shown).

The pattern is as follows. Initially we get a feedback loop dominated by the L and Q states. When initiation starts, that increases L which in turn generates a positive feedback fostering further initiation (contagious spread from L users). Before very long, however, the

increase in L leads to an increase in Q and, hence, the negative feedback from the Q state to initiation, bringing initiation back down. However, over the course of this cycle, some users escalate to state M, where they tend to stay longer than do either L or Q users. So the L and Q states empty themselves, readying them for another round of oscillation in the LQ space, but this time the amplitude will be a little larger because of the additional positive feedback from M users to initiation.

This dynamic generates escalating cycles in LQ space and an accumulation of M state users. That accumulation does not go on indefinitely, however, because some of those M users escalate to the H state, which tends to dampen the oscillation in the LQ space. Furthermore, the structure of short term positive feedback loop with a lagged negative feedback loop that characterizes the LQ dynamic is completely paralleled by the MH dynamic. Ignoring, for the moment, the short-run oscillations generated by the LQ dynamic, the M state builds up generating a positive feedback on initiation, but with some delay, some users in the M state will escalate to the H state. Those escalated users dampen initiation for many years to come since exit from dependent use is much slower than is exit from any other state. Eventually, however, the H state is depleted and the system is ready for another (large-time scale) cycle. In short, the trajectory of use is one of “cycles of cycles”.

When h is large (e.g., at its base case value), the feedback from H to initiation develops so quickly that it shuts down the other cycles. When h is smaller, there is time for these other dynamics to play out before the negative feedback from H comes in and powerfully drives the system to steady state.

3.3 Bivariate sensitivity with respect to escalation rates

The previous section showed that the model behavior is quite sensitive to the rate of escalation (h) from moderate use (M) to heavy use (H). There is also sensitivity with respect to the other

rate of escalation (b), that from light use (L) to moderate use (M). Figure 4 shows a bifurcation diagram [Strogatz, 1994; Kuznetsov, 1995] describing the model's behavior in the (h,b) parameter space. All other parameters are kept constant at the base case values previously mentioned. Details of the bifurcation analysis are reported in Appendix A.2. The parameter space (h,b) is partitioned into different regions characterized by different asymptotic modes of behavior, as denoted in Fig. 4. Stationary, cyclic, quasi periodic and chaotic regimes are possible.

For base case parameter values ($h = 0.07$, $b = 0.4$) the system is solidly in the stationary region. As h is decreased, the system becomes first cyclic, then quasi-periodic and finally chaotic. Reducing the other escalation parameter (b) from its base case value ($b = 0.4$) with h at its lower level (0.007) reverses this sequence, taking the system from chaotic to quasi-periodic, cyclic, and finally stationary regime. So it would be incorrect to infer that reducing or increasing escalation necessarily destabilizes the system's behavior. Rather, at least for these two parameters and in this region of parameter space, a more accurate generalization is that parameter changes that bolster the number of moderate users (M), undercut stability.

Reducing parameter b has an interesting substantive interpretation. Some drugs are intrinsically more appealing to first-time users than are others. As a gross generalization, first-time users often like stimulants (such as cocaine) whereas central nervous system depressants (including heroin) are more likely to be an acquired taste. Holding other parameters constant, specifically parameter a , reducing b reduces the proportion ($b/(a+b)$) of novice users who like the drug enough to progress to regular, moderate use (from L to M). It is thus perhaps not so surprising that reducing the escalation parameter b might improve stability.

Reducing parameter h also has another interesting interpretation. To be precise, users who escalate at rate h from M to state H are not all visible heavy users, but rather all visible heavy users who contribute to the negative reputation of a drug. If there were a visible heavy user

who were perfectly happy with their compulsive use, they would not necessarily serve as much of a deterrent to initiation. There may not be very many dependent users of illicit drugs who are truly happy about their compulsive use, but not all are equally unhappy. Indeed, one major school of thought argues that a principal tenet of drug policy ought to be helping those unfortunate enough to have developed drug dependence so that they may be, in a word, less unhappy with their situation. This “harm reduction” philosophy is advocated in countries such as Australia, England, and the Netherlands. Figure 4 suggests that if a consequence of harm reduction were an effective reduction in the rate (h) at which moderate users (M) escalate to the sort of heavy use that is both visible to prospective initiates and associated with significant harms to the user (H), that might destabilize the behavior of the system, possibly leading in the long-run to recurrent waves of use. Or, equivalently, that it would require a larger fraction of first-time users to dislike the drug (smaller b) in order to avoid such instability.

Furthermore, merely having the drug appeal to a minority of initiates is not by itself sufficient to yield simple behavior if the escalation rate h is small enough. This is illustrated in Fig. 5 where there are regions with multiple attractors, so that depending upon initial conditions the system may tend toward an asymptotic behavior or toward another. Therefore, only very low values of both escalation rates and very particular initial conditions of the state variables may generate stationary behavior. All other cases correspond to more complex dynamics.

3.4 Effects of drug treatment and prevention

The previous section mentioned the possibility that “harm reduction” might destabilize the system. Harm reduction is a controversial approach, particularly in the United States. Two interventions that are almost universally valued are drug prevention for treatment. Prevention, specifically “primary” prevention, seeks to keep non-users from ever trying the drug [Caulkins

et al., 1999, 2002]. Drug treatment can be seen as an intervention that increases the outflow from heavy use [Rydell *et al.*, 1996; Behrens *et al.*, 2000; Tragler *et al.* 2001].

In the context of this model we can model drug treatment as increasing the outflow rate (δ) from heavy use. Likewise we can model prevention as reducing the “contagiousness” parameter s ; reducing s reduces the number of people who are “recruited” into use per current contented user per unit time. Figure 6 shows a bifurcation diagram in (s, δ) parameter space, with all other parameters at their base case values except h , which is set at its reduced value of 0.007 to reflect the idea that not all heavy users are visible to prospective initiates.

As for Figs. 4 and 5, all details on the bifurcation analysis are reported in Appendix A.2. Also in this case, a variation in the parameters s and δ may generate stationary, cyclic, quasi periodic and chaotic behavior.

The base case values ($s = 6$, $\delta = 0.062$) with $h = 0.007$ fall within the chaotic region. As might be expected, effective drug prevention that reduces contagious spread of drug initiation (shrinks s) can improve stability, eventually bringing the system into the stationary region.

What is perhaps more surprising is that either increasing or decreasing the rate of outflow from heavy use (δ) can improve stability. So, increasing treatment from current conditions might help stability, but in the long run so might reducing the proportion of heavy users who are treated! Moreover, for a substance with just the right degree of contagiousness ($s = 5.5$) varying treatment availability so that the average rate of outflow from heavy use moves from 0.05 to 0.2 (i.e., average duration of a career of heavy use varies from 5 to 20 years) moves the system through the following set of regimes: stationary, cyclic, quasi-periodic, chaotic, quasi-periodic, cyclic, chaotic, and back to cyclic again.

A general conclusion from this observation is that the long-run behavior of drug systems may not only be complex, but also fairly sensitive to the parameters of drug control and not

always in intuitive ways. Indeed, to give one more counter-intuitive result, interventions with former users may yield structurally similar system behavior as do interventions with prospective users. That is, the bifurcation diagram in (μ, δ) space is structurally very similar to the diagram in (s, δ) space (Fig. 6). Reducing μ , the rate at which people who disliked the drug stop telling their friends about the bad experience, has very similar effects as does using prevention to reduce s , the rate at which current contented users recruit new initiates.

4. Conclusions

The analysis performed in this paper yields several important conclusions. First, a modest change to an established model of drug use epidemics can yield high as well as low frequency endogenous oscillation in patterns of use. Current trends in drug use (up or down) figure prominently in debates about the effectiveness of various drug policies and political administrations [Reuter & Caulkins, 1995]. If drug use is increasing, policy is assumed to have failed; if drug use is decreasing, policy is assumed to be working. The possibility that there can be multiple superimposed underlying oscillatory patterns greatly complicates such simplistic “score carding” and suggests the need for further research into the underlying dynamics of endogenous variation in drug use.

Second, the model proposed here can generate overall (long-term) cycles of use that take the form of repeated peaks and troughs, with the height of successive peaks growing over time. For some parameter values, the succession of short-term peaks after the overall peak decays slowly. For others, the highest peak is followed by an abrupt crash down to minimal levels of use. These patterns suggest that extrapolating recent drug use trends is dangerous [Caulkins *et al.*, 2003]. Declines from a peak may be merely temporary respite, to be followed shortly by a new peak higher than the previous one. Alternately, even a particularly severe peak might dissipate very rapidly into sustained very low levels of use. This complicates

setting of reasonable drug control goals (a perennial policy conundrum) and retrospectively constructing appropriate counterfactuals for purposes of policy evaluation.

Third, even when a succession of peaks is followed by a crash to very low levels of use that persists for many years, the model has the disquieting ability to generate endogenously a dramatic resurgence of use. Thus, even if the current drug epidemic ebbs and we enjoy a long period of very low rates of use, that does not guarantee drug epidemics will not return with equal ferocity. Furthermore, in this model, the resurgence is not driven by exogenous factors (such as technical innovation in drug production) or an accumulation of a pool of susceptibles, since neither exogenous factors nor susceptibles are included in this model. It is an endogenous function of the epidemic dynamics.

References

- Behrens, D.A., Caulkins, J.P., Tragler, G., Haunschmied, J.L. & Feichtinger, G. [1999] “A dynamic model of drug initiation: Implications for treatment and drug control,” *Mathematical Biosciences* **159**, 1–20.
- Behrens, D.A., Caulkins, J.P., Tragler, G. & Feichtinger, G. [2000] “Optimal control of drug epidemics: Prevent and treat – But not at the same time?,” *Management Science* **46**(3), 333–347.
- Behrens, D.A., Caulkins, J.P., Tragler, G. & Feichtinger, G. [2002] “Why present-oriented societies undergo cycles of drug epidemics,” *Journal of Economic Dynamics and Control* **26**(6), 919–936.
- Behrens, D.A., Caulkins, J.P., Tragler, G. & Feichtinger, G. [in submission] “Memory, contagion, and capture rates: Characterizing the types of addictive behavior that are prone to repeated epidemics”.
- Brill, H & Hirose, T. [1969] “The rise and fall of a methamphetamine epidemic: Japan 1945-55,” *Seminars in Psychiatry* **1**(2), 179-194.
- Caulkins, J.P. [2005] “Models pertaining to how drug policy should vary over the course of an epidemic cycle,” in *Substance Use: Individual Behavior, Social Interactions, Markets, and Politics, Advances in Health Economics and Health Services*, eds. Lindgren, B. & Grossman, M. (Elsevier), Research Volume 16, pp.407-439.
- Caulkins, J.P., Rydell, C.P., Everingham, S.S., Chiesa, J. & Bushway, S. [1999] *An Ounce of Prevention, a Pound of Uncertainty: The Cost-Effectiveness of School-Based Drug Prevention Program*, MR-923-RWJ, RAND: Santa Monica, CA, 1999.

- Caulkins, J.P., Pacula, R., Paddock, S. & Chiesa, J. [2002] *School-Based Drug Prevention: What Kind of Drug Use Does it Prevent?*, MR-1459-RWJ. RAND: Santa Monica, CA.
- Caulkins, J.P., Iguchi, M., Reuter, P.H. & Chiesa, J. [2003] *Drug Use and Drug Policy Futures: Insights from a Colloquium*, RAND: Santa Monica, CA.
- Caulkins, J.P., Behrens, D.A., Knoll, C., Tragler, G. & Zuba, D. [2004] “Modeling dynamic trajectories of initiation and demand: The case of the US cocaine epidemic,” *Health Care Management Science* **7**(4), 319-329.
- Courtwright, D. [2001] *Forces of Habit: Drugs and the Making of the Modern World* (Harvard University Press, Cambridge, MA).
- Dietze, P. & Fitzgerald, J. [2002] “Interpreting changes in heroin supply in Melbourne: Droughts, gluts or cycles?,” *Drug and Alcohol Review* **21**, 295-303.
- Everingham, S.S. & Rydell, C.P. [1994] *Modeling the Demand for Cocaine*, MR-332-ONDCP/A/DPRC, RAND: Santa Monica, CA.
- Everingham, S.S., Rydell, C.P. & Caulkins, J.P. [1995] “Cocaine consumption in the U.S.: Estimating past trends and future scenarios,” *Socio-Economic Planning Sciences* **29**(4), 305–314.
- Johnson, R.A., Gerstein, D.R., Ghadialy, P., Choy, W. & Gfroerer, J. [1996] *Trends in the Incidence of Drug Use in the United States, 1919–1992*. Substance Abuse and Mental Health Services Administration, Office of Applied Studies, U.S. Department of Health and Human Services.
- Kaplan, J. [1983]. *The Hardest Drug: Heroin and Public Policy* (University of Chicago Press, Chicago).

- Kaya, Y., Tugai, Y., Filar, J., Agrawal, M., Ali, R. & Gowing, L. [2004] “Heroin users in Australia: Population trends,” *Drug and Alcohol Review* **23**(1), 107-116.
- Khibnik, A.I., Kuznetsov, Yu.A., Levitin, V.V. & Nikolaev, E.V. [1993] “Continuation techniques and interactive software for bifurcation analysis of ODEs and iterated maps,” *Physica D* **62**, 360-371.
- Kleiman, M.A.R. [1992] *Against Excess: Drug Policy for Results* (Basic Books, New York, NY).
- Kuznetsov, Yu. A. [1995] *Elements of Applied Bifurcation Theory* (Springer-Verlag Applied Mathematical Sciences, New York).
- Marotta, M. [1992] “Drug Abuse and illicit trafficking in Italy: Trends and countermeasures 1979-1990,” *Bulletin on Narcotics* **1**, 15-22.
- Musto, D.F. [1987] *The American Disease* (Yale University Press, New Haven, CT).
- Preti, A., Miotto, P. & DeCoppi, M. [2002] “Deaths by Unintended Illicit Drug Overdose in Italy, 1984-2000,” *Drug and Alcohol Dependence* **66**, 275-282.
- Reuter, P. & Caulkins, J.P. [1995] “Redefining the goals of drug policy: Report of a working group,” *American Journal of Public Health* **85**(8), 1059-1063.
- Rorabaugh, W.J. [1979] *The Alcoholic Republic: An American tradition* (Oxford University Press, Oxford).
- Rubin, J.E. & Terman, D. [2002] “Geometric singular perturbation analysis of neuronal dynamics,” in *Handbook of Dynamical Systems, vol. 2: Toward Applications*, eds. Fiedler, B. (Elsevier, Amsterdam).
- Rydell, C.P. & Everingham, S.S. [1994] *Controlling cocaine. Supply versus demand programs*, MR-331-ONDCP/A/DPRC, RAND: Santa Monica, CA.

- Rydell, C.P., Caulkins, J.P. & Everingham, S.S. [1996] “Enforcement or Treatment: Modeling the Relative Efficacy of Alternatives for Controlling Cocaine,” *Operations Research*, **44**(5), 687-695.
- Spillane, J. [2000] *Cocaine: From Medical Marvel to Modern Menace in the United States, 1884-1920* [Johns Hopkins University Press].
- Strogatz, S. H. [1994]. *Nonlinear dynamics and chaos* (Addison Wesley, Reading MA).
- Substance Abuse and Mental Health Services Administration (SAMHSA). “*National Household Survey on Drug Abuse: Population Estimates [year]*” Washington, D.C.: U.S. Department of Health and Human Services, annual.
- Tragler, G., Caulkins, J.P. & Feichtinger, G. [2001] “Optimal dynamic allocation of treatment and enforcement in illicit drug control,” *Operations Research* **49**(3), 352-362.
- Waldorf, D., Reinerman, C. & Murphy, S. [1991]. *Cocaine changes: The experience of using and quitting* (Temple University, Philadelphia).
- Weatherburn, D., Jones, C., Freeman, K. & Makai, T. [2003] “Supply control and harm reduction: lessons from the Australian heroin ‘drought’,” *Addiction* **98**, 83-91.
- Winkler, D., Caulkins, J.P., Behrens, D. & Tragler, G. [2004] “Estimating the relative efficiency of various forms of prevention at different stages of a drug epidemic,” *Socio-Economic Planning Sciences* **38**(1), 43-56.

Appendices

A.1 Specification of parameter values for the U.S. cocaine epidemic

The National Household Survey on Drug Abuse (NHSDA) offers the only nationally representative data on general patterns of drug use by people of all ages [SAMHSA, annual]. From the 1998 survey (most recent year for which variables were available) we cross-tabulated recency of cocaine use (variable COCREC) with lifetime reported days of cocaine use (COCTOT). For those (about 17 million) who report having used cocaine but not in the past 3 years (and, hence, are likely to have completed their use), 26.4% reported using on only one or two days and 30.4% reported using on 3 – 11 days. The next category of lifetime use (12-100 times) exceeds that which the state Q is meant to represent, so we assume that 56.8% of all initiates enter state Q (i.e., $a/(a+b) = 0.568$).

From the 1991, 1995, and 1999 NHSDA's we computed the probability of using but not heavily as a function of time since initiation over a 20-year horizon. The curve drops sharply at first, then more slowly, implying that flow rate $a > g$, as expected. The best fitting parameter values a , b , and g in the sense of minimizing the sum of the squared error subject to the constraint that $a/(a+b) = 0.568$ were $a = 0.515931$, $b = 0.392398$, and $g = 0.145912$. To avoid giving a false sense of precision, we round these to $a = 0.52$, $b = 0.4$, and $g = 0.15$.

The NHSDA does not generally ask about lifetime prevalence of heavy use, but the 1992, 1993, and 1994 NHSDA surveys did ask about lifetime prevalence of having ever used daily for two or more weeks. That is a different definition of reaching the H state than in Everingham and Rydell [1994] or Behrens *et al.* [1999], but not an irrelevant concept. For people who initiated after 1975 (i.e., in the modern era) and had initiated at least six years earlier (giving them time to escalate if they were ever going to) about 19% of all initiates

reported having ever used daily for 2+ weeks. This suggests that $(h/g) * b/(a+b) = 0.19$, or $h = 0.065217$, which we round to $h = 0.07$.

We interpret Q as people who tried the drug but also quit within the last year, so $\mu = 1$. Exit from heavy use means the same thing in this model as in the LH model, so we use Behrens *et al.*'s [1999] value of $\delta = 0.062$. Likewise we borrow Behrens *et al.*'s value for $\tau = 0.05$ and set the deterrence constant $q = 3.5$ based on Knoll and Zuba [2000].

Without much basis we choose $\alpha = 2$, implying that recent initiates are twice as close, socially, to potential initiates as are dependent users. Likewise, we arbitrarily set $K = 1$, implying that the floor on the effective number of light users in the reputation term is one million. In this model both heavy users and those who tried recently but quit help suppress use, so the contagiousness parameter s needs to be increased in order to still reproduce in rough form the historical epidemic. We increase it relative to Behrens *et al.*'s [1999] value by about a factor of 10 (from 0.61 to 6). This parameterization of the initiation function is admittedly ad hoc, but that is largely unavoidable because the U.S. household surveys (unlike the 1998 Australian survey that Kaya *et al.* [2004] used) have not asked respondents who have used the drug but not in the last 12 months, what was their last year of use. Hence, we cannot reconstruct the historical time series for $Q(t)$ to use in fitting this function. Adding such questions to future surveys is currently under deliberation (Donald Goldstone, personal communications).

A.2 Bifurcation analysis

The relationships among parameters characterizing the bifurcations of model (1,2) cannot be found analytically because of the complexity of the model. Nevertheless, by means of LOCBIF, a program implementing a powerful technique for bifurcation analysis [Khibnik *et al.*, 1993], it is possible to numerically detect almost all bifurcations characterizing system

(1,2) involving equilibria, cycles and tori and to display them in any two-dimensional parameter space. By contrast, the analysis of chaos has been performed numerically by means of a simulation program. In this paper, bifurcation curves involving only attractor, namely long-run dynamics of the system, are shown in the parameter spaces (h,b) and (s,δ) .

Figure 4 is the bifurcation diagram of the attractors of model (1,2) in the parameter space (h,b) . All other parameters are kept constant at their base case values specified in the text. Figure 5 zooms into the lower-left corner of Fig. 4.

In these diagrams there are several bifurcation curves, identified by symbols which have the following meaning [Strogatz, 1994; Kuznetsov, 1995]:

H Hopf bifurcation

NS Neimarck-Sacker bifurcation

f_∞ Flip (or period doubling) bifurcation (for cycles of infinite period)

T Tangent bifurcation of cycles

T_d Torus destruction

BS Blue-Sky bifurcation

All these curves have been produced by LOCBIF, except the torus destruction T_d which has been obtained by simulation. The bifurcation curves are the boundary of different regions characterized by different asymptotic modes of behavior. Stationary, cyclic, quasi periodic and chaotic regimes are possible.

As for Fig. 4, for high values of h the system only has stationary behavior, i.e. it tends toward a stable equilibrium. Decreasing h , crossing the Hopf bifurcation curve (H) would imply a transition to a cyclic regime; the stable equilibrium becomes unstable and a stable limit cycle arises. Decreasing h again and crossing the Neimark-Sacker bifurcation curve (NS), the stable cycle becomes unstable (a saddle cycle) and is surrounded by a stable torus

(quasi periodic behavior). As h is further decreased, curve T_d is crossed; the torus collides with a saddle cycle, disappears and is substituted by a chaotic attractor (torus destruction). Notice that curve T_d is only a sketch of the boundary of the chaotic region because it actually has a fractal structure.

Finally, changing parameters, curves denoted by BS and f_∞ can be crossed. When curve BS is crossed, a Blue-Sky bifurcation occurs, resulting in a tangent bifurcation of cycles. Crossing this curve, a stable and a saddle cycle disappear and an infinite number of saddle limit cycles which belong to a Smale-Williams solenoid chaotic attractor [Kuznetsov, 1995] is generated. Moreover, the above stable cycle may undergo an infinite sequence of flip bifurcations; crossing each flip curve a stable cycle of period T is substituted by a stable cycle of period $2T$. The sequence accumulates on curve f_∞ over which the behavior becomes chaotic.

As for Fig. 5, attractors involved in bifurcation curves denoted by H , NS and T_d follow the same path as in Fig. 4. Moreover, crossing curve T a stable and an unstable cycle appear. Thus, for example, starting from the lower right corner (stationary behavior) and decreasing h , so that curve T is crossed, would imply alternative attractors, the already existing equilibrium and the newly born cycle.

Figure 6 reports the bifurcation diagram of the attractors of model (1,2) in the parameter space (s, δ) . All other parameters are kept constant at their base case values specified in the text, except h which is lowered. Symbols identifying each curve have the above mentioned meaning, the only exceptions are curves denoted by f_1 and f_2 corresponding to the first and the second flip bifurcation of the Feigenbaum cascade. For low values of s and δ the system only has stationary behavior, i.e. it tends toward a stable equilibrium. Crossing the Hopf bifurcation curve (H) would imply a transition to a cyclic regime (the stable equilibrium

becomes unstable and a stable limit cycle arises). The cycle may undergo either a sequence of flip bifurcations or a Neimark-Sacker bifurcation. In the first case, curves denoted by f_1 and f_2 corresponding to the first and the second flip bifurcation of the Feigenbaum cascade are initially crossed (a stable cycle of period T is substituted by a stable cycle of period $2T$). The sequence of flips accumulates on curve f_∞ which corresponds to the boundary of the chaotic region. In the second case, crossing the Neimark-Sacker bifurcation curve (NS), the stable cycle becomes unstable (a saddle cycle) and is surrounded by a stable torus (quasi periodic behavior). When curve Td is crossed, the torus collides with a saddle cycle, disappears and is substituted by a chaotic attractor (torus destruction). As in Fig. 4, curve T_d is only sketched.

Table 1: Base case parameter values for current U.S. cocaine epidemic

Parameter	Value	Description
a	0.52	<i>annual rate at which light users quit</i>
b	0.4	<i>annual rate at which light users progress to moderate use</i>
g	0.15	<i>annual rate at which moderate users quit or escalate</i>
μ	1	<i>annual rate at which people exit the state of having tried and quit within the last year</i>
h	0.07	<i>annual rate at which moderate users escalate to heavy use</i>
δ	0.062	<i>annual rate at which heavy users quit</i>
τ	0.05	<i>millions of innovators per year</i>
s	6	<i>annual rate at which light users attract non-users</i>
q	3.5	<i>constant which measures the deterrent effect of reputation</i>
α	2	<i>constant measuring how socially proximate recent initiates vs. heavy users are to potential initiates</i>
K	1	<i>constant in reputation term that augments number of light users</i>

Figure Captions

- Figure 1. The LMHQ model's stocks (boxes), flows (solid arrows), and information flows (dashed arrows).
- Figure 2. Time evolution of initiation (*a*), content users (L+M) (*b*), and heavy users (*c*). Parameters are at base case values (Table 1); initial conditions are specified in the text.
- Figure 3. Trajectories in M-H state space and time evolution of light users showing cyclic (*a*), quasi periodic (*b* and *c*), and chaotic (*d*) behavior. Parameters are at their base case values (see Table 1), except $h = 0.0093$ (*a*), $h = 0.009$ (*b*), $h = 0.008$ (*c*), $h = 0.006$ (*d*).
- Figure 4. Bifurcation diagram in the (h, b) parameter space. All other parameters are at their base case values (see Table 1); for label interpretation see Appendix A.2. The left-lower corner of the diagram is expanded in detail in Fig. 5.
- Figure 5. Expanded bifurcation diagram in the (h, b) parameter space (see Fig. 4).
- Figure 6. Bifurcation diagram in the (s, δ) parameter space. All other parameters are at their base case values (see Table 1), except $h = 0.007$; for label interpretation see Appendix A.2.

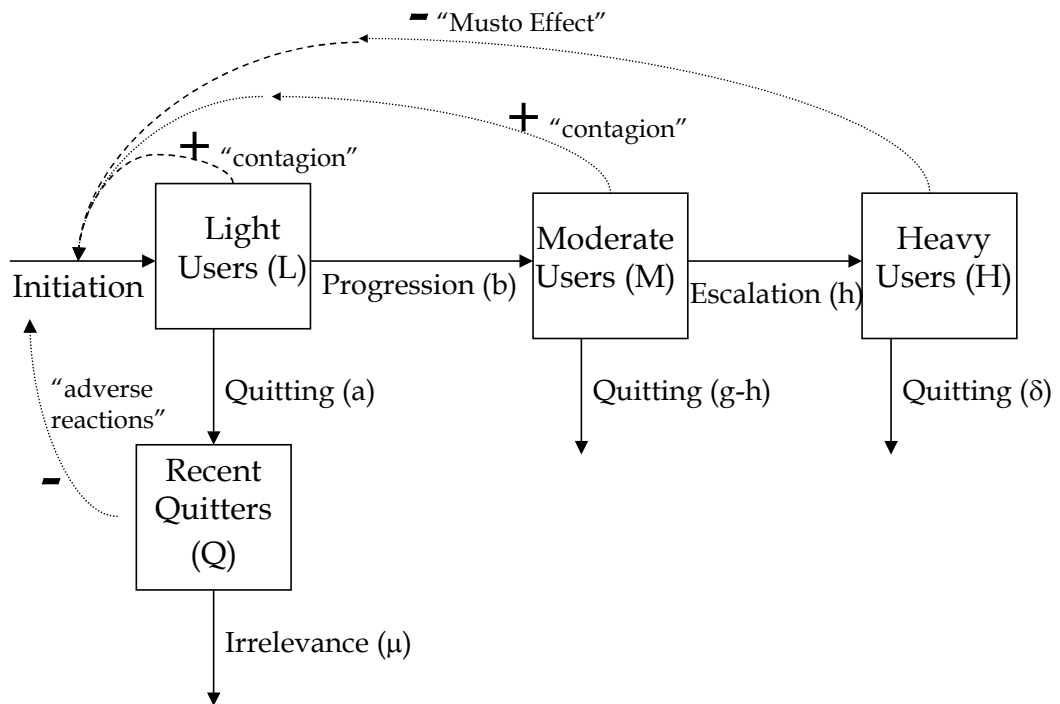


Figure 1

Caulkins, J.P.
 Gragnani, A.
 Feichtinger, G.
 Tragler, G.

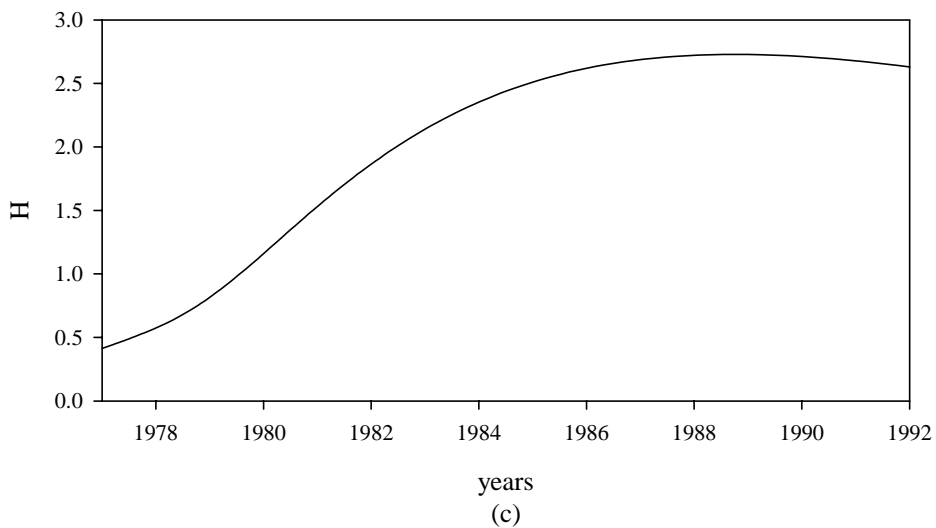
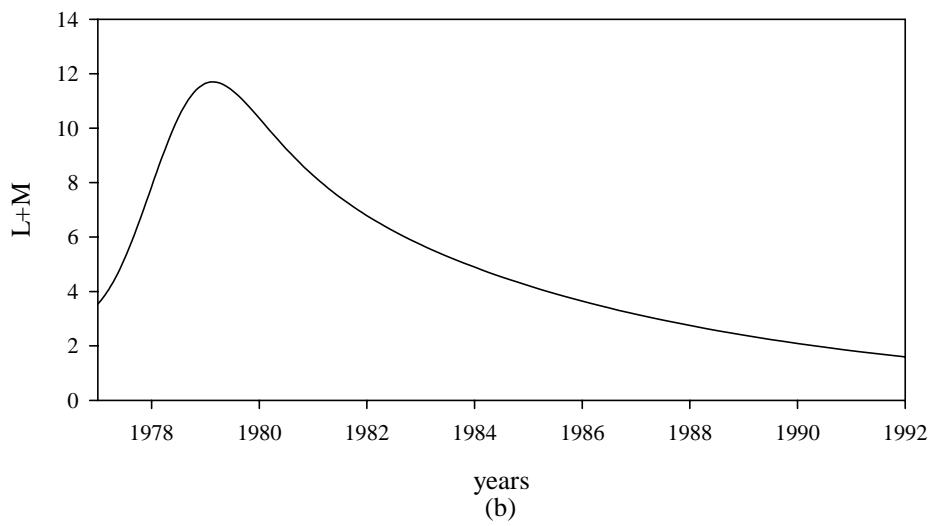
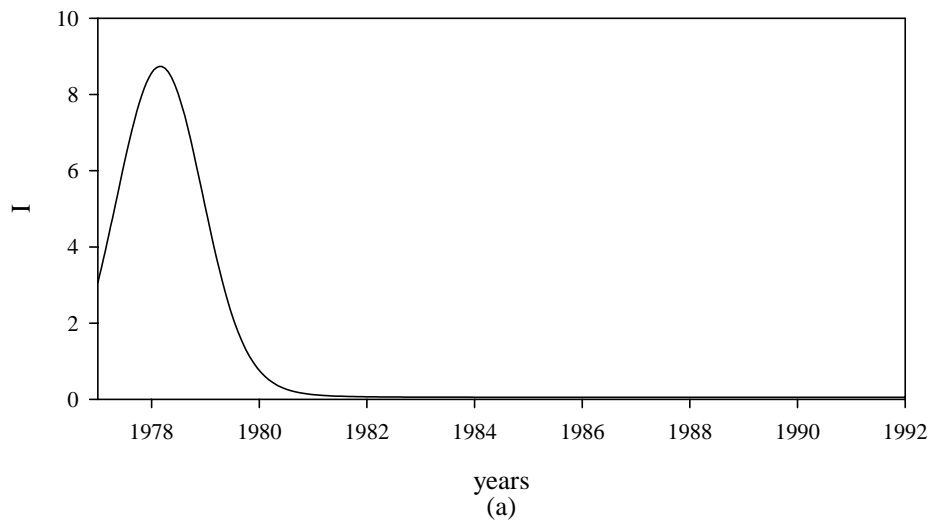


Figure 2

**Caulkins, J.P.
Gragnani, A.
Feichtinger, G.
Tragler, G.**

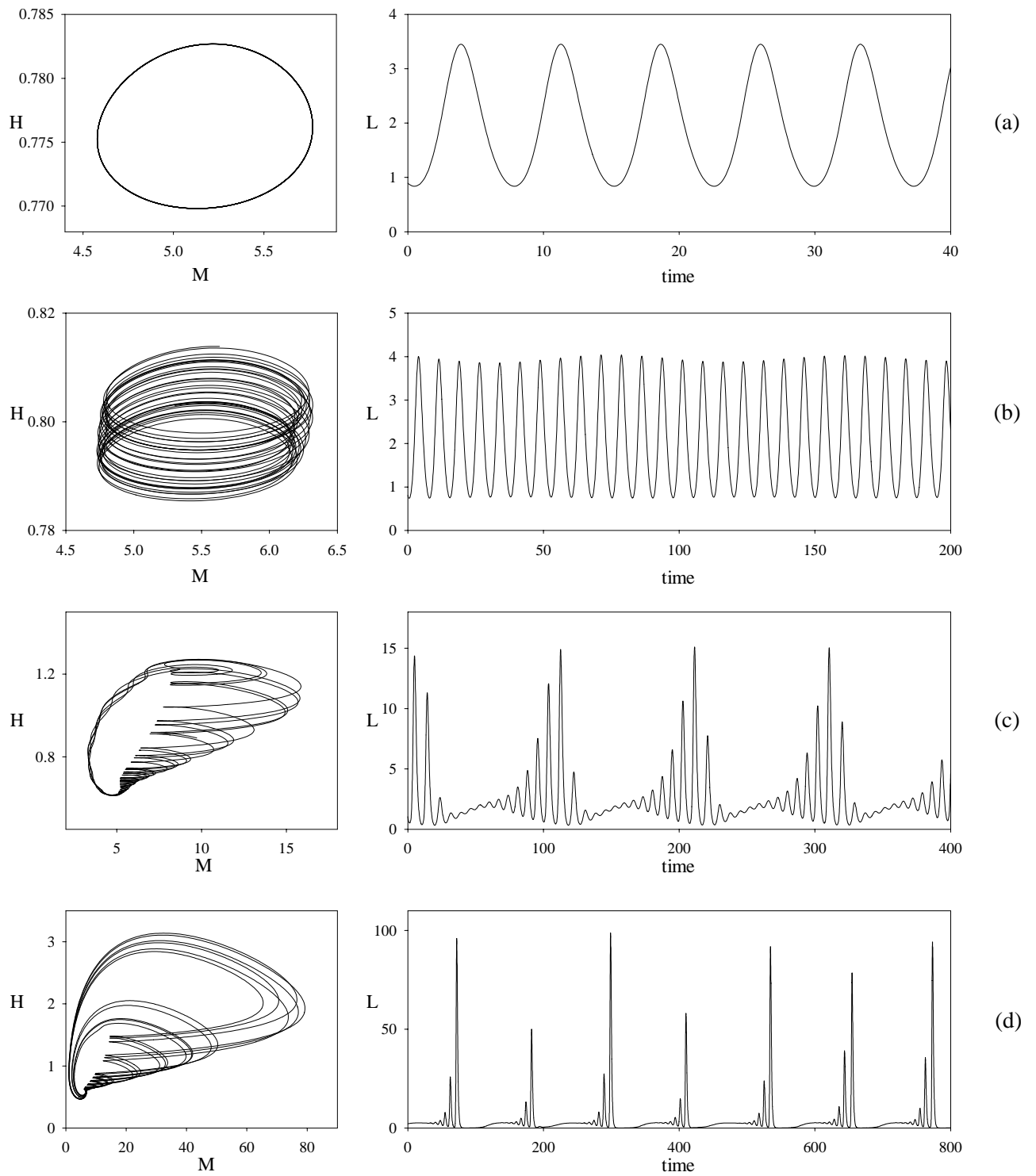


Figure 3

**Caulkins, J.P.
Gragnani, A.
Feichtinger, G.
Tragler, G.**

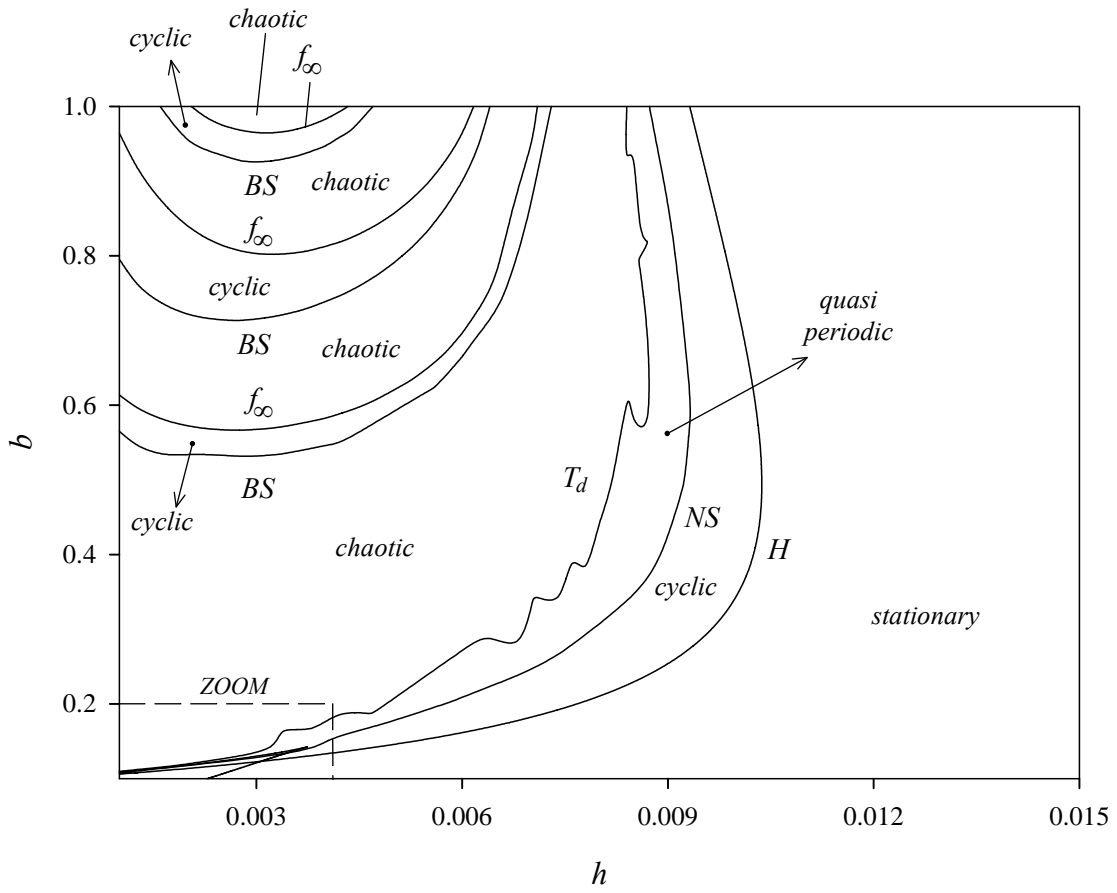


Figure 4

**Caulkins, J.P.
Gragnani, A.
Feichtinger, G.
Tragler, G.**

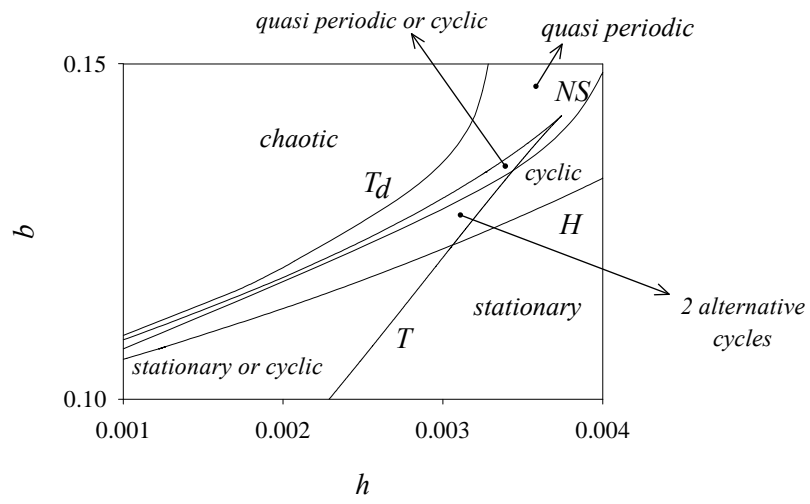


Figure 5

Caulkins, J.P.
Gragnani, A.
Feichtinger, G.
Tragler, G.

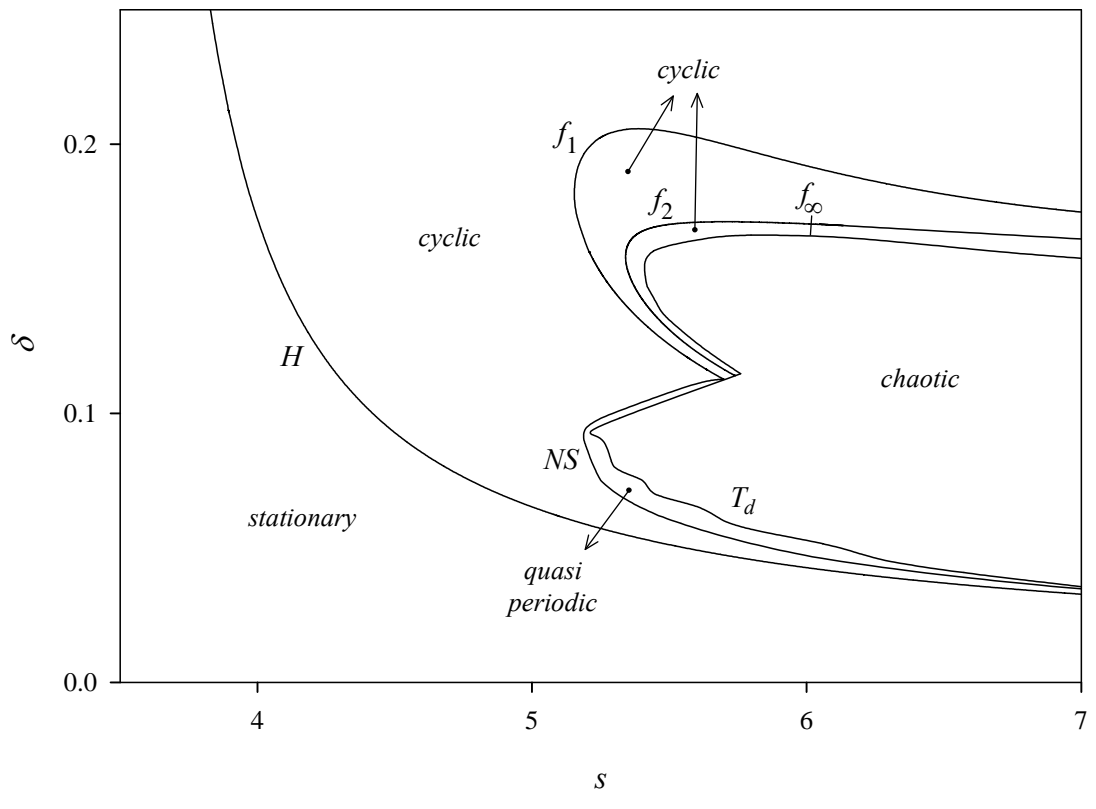


Figure 6

Caulkins, J.P.
Gragnani, A.
Feichtinger, G.
Tragler, G.