December 30, 2011.

Disinhibition and Immiserization in a Model of Susceptible-Infected-Susceptible (SIS) Diseases

by

Mark Gersovitz*
Department of Economics
The Johns Hopkins University

Abstract

Infectious diseases induce externalities in private choices about prevention and therapy. An improvement in either the technology of prevention or therapy may lead private agents to decrease their preventive or therapeutic efforts, a phenomenon termed disinhibition by epidemiologists. If governments cannot or do not adopt interventions to internalize these externalities, a technological improvement may even lead to disinhibition so extreme that the infection rate rises. In turn, a rise in the infection rate is a necessary but not sufficient condition for immiserization, the paradoxical fall in welfare consequent on a technological improvement. These issues are investigated in a model in which susceptibles may become infected and infecteds may recover to be again susceptible. Conditions are provided for when the infection rate can rise depending on whether the improvement is to the technology of prevention or therapy and at whom prevention or therapy is targeted, as well as other parameters of the model. Results are provided for a general formulation.

0. Introduction

There is a strong intuition that infectious diseases exhibit externalities, and recent models have formalized this notion (Goldman and Lightwood, 2002, Gersovitz, 2003 and Gersovitz and Hammer, 2004). In principle, governments can internalize externalities through a conventional strategy of taxes and subsidies, as these authors show. But in practice, much preventive and therapeutic behaviour is unobserved and therefore cannot be taxed or subsidized, often because it is intimate as in the case of sexually transmitted diseases. This situation raises problems of the second best, in particular, when there are improvements in prevention or therapy.

Epidemiologists' worry that after such an improvement, people may slacken efforts to prevent or recover from infection, a phenomenon they term disinhibition. Disinhibition could even be so strong that infection rates may rise. Epidemiologists do not, however, focus on a comprehensive welfare measure of such changes inclusive of their impacts on the costs of prevention and therapy. Moreover, epidemiologists do not distinguish between situations in which the first-best has been implemented and it has not, an important distinction from a welfare perspective. Economists worry in general that when externalities have not been internalized, seemingly desirable innovations might lead to a fall in well-being, immiserization.

A particular application of the principle of immiserization is that innovations in prevention and therapy might produce immiserization, but under what conditions? Answering this question is the main goal of this paper which builds on Gersovitz and Hammer (2004). It requires a dynamic model because the infection process is inherently dynamic but such models pose challenges of tractability. Although a number of relatively general results are presented, the paper also uses a

special functional form for a critical component of the model, the relation between preventive or therapeutic inputs and the dynamics of the model. This functional form has intuitive properties and allows for the derivation of some analytical results when the general results on disinhibition/immiserization are ambiguous. The special functional form may also be useful in further closed-form examination of the economics of infection control for instance in the analysis of multiple optimal steady states (Gersovitz, 2012) as well as in simulation.

1. The Social Planner and the First-Best

1.1 The Dynamic Constraints:

The starting point for this study of optimal policy toward infectious diseases is the susceptible-infected-susceptible (SIS) model of mathematical epidemiology. It provides an accounting for people in society, who belong to two mutually exclusive groups and are either susceptible to infection or infected and infectious, that is capable of infecting the susceptibles. Once susceptibles are infected, they recover probabilistically to be again susceptible.

The total number of people (N) is therefore the sum of the number who are: (1) susceptible (S) and (2) infected and infectious (I):

$$(1) \qquad N = S + I \ .$$

The proportions of these groups in the population are denoted by s and i, with $s + i = 1$. The birth rate of the population is ϵ and deaths do not occur, so that the change in the population is:

$$(2) \qquad \dot{N} = \epsilon N \ .$$

The number of susceptibles changes according to:

(3) $\dot{S} = \epsilon N - \alpha S\pi + \beta I$.

The first part of the right-hand side embodies the assumption that all newborns are susceptible. The second part reduces the number of susceptibles by those people who become infected. Under the assumption of random contacts, the probability per contact of a susceptible person's meeting an infected (and infectious) person, π, is the proportion of infected people in the population, $\pi = i = I/N = (1 - s)$ and is the same for all people. The product, $S\pi$, is the number of susceptibles who do so. The factor α is an adjustment incorporating both the rate of contact and the inherent infectiousness of an infected. The third part is the addition to the susceptible pool resulting from the recovery of a fraction, β, of the infecteds. Eqs. (1) - (3) can be solved for the change in the proportion s:

(4) $\dot{s} = -\alpha s(1-s) + (1-s)(\beta+\epsilon)$,

which is a differential equation in one variable if α, β and ϵ are fixed parameters.

Subsequent discussion replaces two parameters, α and β, by endogenous functions of preventive and therapeutic inputs, themselves ultimately optimally-chosen implicit functions of the single state s. The main purpose is to consider how preventive and therapeutic actions by governments and individuals endogenously affect the evolution of disease dynamics and how to design public health interventions. Without any such interventions, however, these parameters are fixed, and the model of eq. (4) evolves to one of two possible steady states, either s = 1, the infection dies out, or $s^* = (\beta + \epsilon)/\alpha$. If $0<s^*<1$ so that the infection is endemic, s^* is stable and s = 1 is unstable. Thus the problem of multiple endemic steady states does not arise in the simple SIS model without endogenous α and β.

1.2 The Costs and Benefits of Interventions and the Social Planner's Optimization:

The social planner's objective function sums the well-being of individuals whether susceptible or infected. Absent the disease, each person in the society would have a level of income of V_0. Utility is assumed linear in income, and so V_0 would also be their level of utility.

When the disease exists, however, the social planner will likely be spending resources on prevention to lower the chance of infection (α) and on therapies to increase the chance of recovery (β). Either intervention may be targeted so that only a proportion of the population receives the intervention and generates the associated costs. Let θ^j, j = a, b be the proportions of the population that generate either preventive or therapeutic costs associated with an infectious disease. The θ^j are targeting functions, and depend on s in special ways taking several forms with some people getting access to prevention and/or therapeutic inputs.

A natural form of targeting is for prevention to target susceptibles ($\theta^a = s$) and for therapeutic interventions to target infecteds ($\theta^b = 1-s$). Preventive targeting at everyone, $\theta^a = 1$, with therapy targeted only at the infected, $\theta^b = 1-s$, may, however, be at least as natural in practice. For instance, someone taking antibiotics to recover from cholera is hardly likely to think it a good time to take a break from trying to drink safe water. The difference between $\theta^a = 1$ and $\theta^a = s$ does have important implications for some qualitative properties of the model. Mass prevention and mass therapy may be appropriate for infections that are hard to diagnose, $\theta^a = 1$ and $\theta^a = 1$. Plausible alternatives also include: $\theta^a = 1-s$, or $\theta^a = s(1-s)$ and $\theta^b = 1$ in combination with the targeting functions just discussed; Gersovitz and Hammer (2004) provide more discussion of these alternatives. In this paper, I consider all eight possible combinations of these expressions in terms of s for these two targeting functions.

In the social planner's problem, the government directly controls the quantities of the two policy interventions, preventive effort of $a \geq 0$ units per person at whom prevention is targeted and therapeutic effort of $b \geq 0$ units per person at whom therapeutic effort is targeted. The total number of units of these interventions are therefore $a\theta^a N$ and $b\theta^b N$. The levels of these health inputs per targeted person affect the model through $\alpha(a)$ and $\beta(b)$, and thereby determine respectively the rate of new infections and the rate of recovery. Nothing precludes $\beta(0) = 0$, in which case there is no recovery without some therapy although formally this special case of the model would not be SIS in the absence of therapy. The controls exhibit positive but diminishing marginal products in the promotion of health through prevention and therapy so that: $\alpha'<0$, $\alpha''>0$, $\beta'>0$, $\beta''<0$. Preventive inputs cost p_a per unit and therapeutic inputs cost p_b per unit; total costs of the interventions are therefore $p_a a\theta^a N$ and $p_b b\theta^b N$.

To represent the possibility of improvements to prevention and therapy, I parameterize the α and β functions as follows:

$$(5a) \qquad \alpha = \alpha(a, \mu_\alpha), \quad \alpha_\mu = \partial\alpha/\partial\mu_\alpha < 0$$

and

$$(5b) \qquad \beta = \beta(b, \mu_\beta), \quad \beta_\mu = \partial\beta/\partial\mu_\beta > 0 \ ,$$

in which the μ_j are constants unless altered by technical change. Important special cases of these equations occur when either prevention or therapy (but not both) is entirely under discretionary private control and the other intervention is subject to the technical improvement. For instance, prevention may involve effort that is continuously variable (say hand washing that can be done more vigorously and more often). Therapy may, however, consist of taking an antibiotic or not, a

better antibiotic is an increase in μ_β, taking the antibiotic may be the obvious choice both before and after the innovation, and there may be no other therapeutic intervention so b = 0. The key interaction is then between the change in μ_β and the induced optimal changes in preventive effort, a. The reverse situation could also happen in which private choice is about therapy and the innovation affects prevention about which there is no private decision. So long as the μ_i are constant, however, these parameterizations do not place any restrictions on the properties of the model which remains as general as the SIS model in Gersovitz and Hammer (2004).

A further specialization of the prevention and therapeutic functions uses an exponential form and makes the μ_j additive so that $\alpha'_\mu = \partial\alpha/\partial a\partial\mu_\alpha = 0$ and $\beta'_\mu = \partial\beta/\partial b\partial\mu_\beta = 0$:[1]

$$(6a) \qquad \alpha = -\mu_\alpha + \gamma_\alpha e^{-a} - \gamma_\alpha \; ; \qquad -\mu_\alpha > \gamma_\alpha > 0 \; , \; a \geq 0 \; ,$$

and

$$(6b) \qquad \beta = \mu_\beta - \gamma_\beta e^{-b} + \gamma_\beta \; ; \qquad \mu_\beta \; , \; \gamma_\beta > 0 \; , \; b \geq 0 \; .$$

Although special, these formulations allow some unambiguous conclusions about technological change. Under these assumptions,

$$(7a) \qquad \alpha(0) = -\mu_\alpha \; ; \quad \alpha(\infty) = -\mu_\alpha - \gamma_\alpha \; ; \quad \alpha' = -\gamma_\alpha e^{-a} < 0 \; ; \quad \alpha'' = -\alpha' > 0 \; ,$$

and

[1]In principle, eqs. (6a) and (6b) could be generalized to include parameters σ_a and σ_b multiplying the exponents, -a and -b. For the comparative analysis presented in this paper, however, such parameters would only be nuisance parameters because they could be eliminated by a change in the units of measurement of a or b. Heffley(1982) uses a similar functional form in a discussion of prevention and therapy but one that does not model the dynamics of infectious diseases.

(7*b*) $\beta(0) = \mu_\beta \; ; \quad \beta(\infty) = \mu_\beta + \gamma_\beta \; ; \quad \beta' = \gamma_\beta e^{-b} > 0 \; ; \quad \beta'' = -\beta' < 0 \; .$

These functional forms place an upper bound on α and β. As most people are all too aware, it is often impossible either to ensure that one cannot be infected or to ensure that one will recover, no matter what one spends on one's health. Furthermore, if even when the social planner uses the maximum amount of inputs and

(8*a*) $$s^*_{\max} = \frac{\beta(\infty) + \epsilon}{\alpha(\infty)} = \frac{\mu_\beta + \gamma_\beta + \epsilon}{-\mu_\alpha - \gamma_\alpha} < 1 \; ,$$

then s = 1 must be a locally unstable equilibrium because it is unstable no matter the inputs a and b. Eq. (8a) gives the highest possible value of the proportion of susceptibles in the steady-state. Correspondingly, the lowest possible value of the steady-state proportion of susceptibles is:

(8*b*) $$0 < s^*_{\min} = \frac{\beta(0) + \epsilon}{\alpha(0)} = \frac{\mu_\beta + \epsilon}{-\mu_\alpha} < 1 \; ,$$

when the social planner does nothing. Taken together, the properties of eqs. (7a), (7b), (8a) and (8b) make this functional form both powerful in overcoming some intractabilities of the model and at the same time consistent with some intuitions, such as that infinite inputs may not be able to drive the model to a zero-infection steady state.

The social planner maximizes social welfare, W, as given by the present discounted value of total income net of the total costs of the disease and the total costs of the interventions:

(9) $$W = \int_0^\infty \{ V_0 N - [p_I (1-s) N + p_a a \theta^a N + p_b b \theta^b N] \} \, e^{-rt} dt$$

in which r is the discount rate, V_0 is income in the absence of the disease (received by everyone who is alive whether well or sick), p_I is the current money cost of being infected (and sick) such as foregone wages while ill and including the monetary equivalent of pain and suffering and (1-s)N is the total number of the sick, and the remaining two terms are the total costs of the preventive and therapeutic interventions. The integrand of eq. (9) is therefore the sum of the current incomes of the susceptible and the infected less the total current costs of illness and health interventions. Note that there is no cost to any innovation, μ_i,. These are assumed to be the product of pure science and would be undertaken regardless of cost. An alternative formulation would be to assume that innovation is the outgrowth of public choice about investment in innovative activity dependent on its costs and benefits inclusive of any second-best consequences such as the potential for immiserization

Eq.(9) therefore provides the objective function while eqs. (2) and (4) provide the dynamic constraints on the optimization problem. The current-value Hamiltonian, H, is:

$$(10) \qquad \begin{aligned} H = N \{ V_0 - [p_I i + p_a a\theta^a + p_b b\theta^b] \} \\ + (\lambda_s N)[-\alpha s\pi + (1-s)(\beta+\epsilon)] \\ + \lambda_N [\epsilon N] \end{aligned}$$

in which $(\lambda_s N)$ and λ_N are the current value multipliers and α and β are given by eqs. (5 and (5b). Because $\lambda_s N$ is the multiplier on the change in s, it has the interpretation of the value of a unit increase in s on social welfare, W, and therefore λ_s has the interpretation of the value of an increase in s on the welfare of the representative person, or alternatively as the value of an increase in one susceptible person on social welfare.

The first derivatives of H with respect to the controls, a and b, set equal to zero imply:

$$(11a) \qquad p_a\theta^a = -\lambda_s\alpha' s\pi$$

and

$$(11b) \qquad p_b\theta^b = \lambda_s\beta'(1-s)$$

Under the assumptions on the θ^j and on α' and β', the variable λ_s must be positive if the first-order conditions are to hold. The left hand sides of eqs. (11a) and (11b) show how the θ^j play a critical role in determining the marginal costs of the interventions by making them depend explicitly on the state of the epidemic, s. In the case of prevention, the relation between marginal cost and s takes many forms including a direct and an inverse relation.

In addition, the dynamic equations for the multipliers imply:

$$(11c) \qquad \dot{\lambda}_s = r\lambda_s - [p_I - p_a a\theta_s^a - p_b b\theta_s^b] + [\alpha(1-2s)+\beta]\lambda_s$$

and

$$(11d) \qquad \dot{\lambda}_N = \lambda_N(r-\epsilon) - [V_0 - (p_I i + p_a a\theta^a + p_b b\theta^b)] \quad .$$

The dynamic system of eqs. (4) and (11a-c) is independent of N and λ_N although not of ϵ and eq. (11d) can be discarded. (This result would not hold if people died of the disease, Gersovitz and Hammer, 2004). The social planner recognizes that $\pi = (1-s)$, hence the term (1 -2s) in eq. (11c).

Substitution of eqs. (11a) and (11b) into eqs. (4) and (11c) and setting the resulting equations to zero produces the two isoclines of the dynamic system. The simultaneous solution of the isoclines for s and λ_s yields all the steady states of the system. At this point, a lot of trouble

begins. In most economic problems, there is a unique optimal steady state. In the analysis of the control of infectious diseases, however, it is not clear what general condition on the primitives of the problem rules out the possibility of multiple optimal steady states other than the condition that the isoclines cross only once. Gersovitz (2012) discusses these issues further, but in this paper it is assumed that the steady state is unique, as it may be.

The slope of the s-isocline in s-λ space is:

$$(12a) \qquad \left[\frac{\partial \lambda_s}{\partial s}\right]_{\dot{s}=0} = \frac{-(\beta' b_s - \alpha - \alpha' s a_s)}{\beta' b_\lambda - \alpha' s a_\lambda} = \frac{-\Delta_{ss}}{\Delta_{s\lambda}} = \frac{?}{+},$$

and the slope of λ-isocline in s-λ space is:

$$(12b) \qquad \left[\frac{\partial \lambda_s}{\partial s}\right]_{\dot{\lambda}_s=0} = \frac{-\{p_a(a_s\theta^a_s + a\theta^a_{ss}) - \lambda_s[2\alpha - (1-2s)\alpha' a_s] + \beta' \lambda_s b_s\}}{[r - \alpha s + \beta + (1-s)(\alpha + \alpha' s a_s) + (\lambda_s \beta' + p_b \theta^b_s) b_\lambda]} = \frac{-\Delta_{\lambda s}}{\Delta_{\lambda\lambda}} = \frac{+}{?},$$

in which the Δ_{jk}, j, k = s, λ denote the corresponding numerators and denominators that precede their introduction, the j subscript denotes the isocline and the k subscript denotes a variable in that isocline. The terms a_j and b_j for j = s and λ refer to the partial derivatives of a and b calculated from eqs. (11a) and (11b), the first-derivatives of the Hamiltonian set to zero; Gersovitz and Hammer (2004, p.10) give these expressions for the case $\theta^b = 1 - s$ and all four values of θ^a. The denominator of eq. (12b) involves some algebraic manipulation using the expressions for the a_j and does not derive immediately from the differentiation of eq. (11c).

If $\theta^a = s$ or $s(1-s)$, then both isoclines have positive slopes about the steady state so long as $r \geq \epsilon$ and if the s-isocline is steeper than the λ -isocline then the steady state is stable as can be

proved by linearizing the dynamic eqs. (4) and (11c) about the steady state.[2] If $\theta^a = (1 - s)$ or $\theta^a = 1$, then both isoclines can have positive slopes about the steady state with the s-isocline steeper than the λ -isocline so that the steady state is stable, just as when $\theta^a = s$ or $s(1-s)$. These isoclines may have other combinations of slopes but none is stable.[3] For notational simplicity, define the difference between the slope of the s-isocline and the slope of the λ-isocline as Δ:

$$(13) \qquad \Delta = -\frac{\Delta_{ss}}{\Delta_{s\lambda}} + \frac{\Delta_{\lambda s}}{\Delta_{\lambda\lambda}} = \Delta_{ss}\left[\frac{-1}{\Delta_{s\lambda}} + \frac{\lambda}{\Delta_{\lambda\lambda}}\right] + \frac{\{p_a(a\theta^a_{ss} + \theta^a_s a_s) + \lambda[(1-s)\alpha' a_s - \alpha]\}}{\Delta_{\lambda\lambda}},$$

which is positive in the stable case when both isoclines are upward sloping and the s-isocline is steeper.

The exponential form of eqs. (6a) and (6b) and their associated properties makes it possible to derive explicit algebraic expressions for the two isoclines by using these special forms

---

[2] An important property of the linearized equations is that the sum of the coefficient on s in the linearized version of eq. (4) and of the coefficient on λ_s in the linearized version of eq. (11c) is positive so long as $r \geq \epsilon$. If, in addition, the slopes of the isoclines are positive and the slope of the s-isocline is steeper than the slope of the λ-isocline, then the dynamic system must have one negative and one positive root and so is saddlepoint stable. If both slopes are positive but the relative steepness of the slopes are reversed, the system has two positive roots and is unstable.

[3] Gersovitz and Hammer (2004, p. 13) incorrectly state that there can be a stable case with both isoclines negatively sloped and the s-isocline more negatively sloped; although both isoclines can be negatively sloped, the s-isocline cannot be the more negatively sloped and so such an intersection cannot be stable. From eqs. (12a) and (12b), for both isoclines to be negatively sloped requires $\Delta_{ss} > 0 \; and \; \Delta_{\lambda\lambda} < 0$, but for the s-isocline to be more steeply negatively sloped also requires $-\Delta_{ss}\Delta_{\lambda\lambda} + \Delta_{s\lambda}\Delta_{\lambda s} > 0$. This latter condition cannot hold if $\theta^a = (1 - s)$ or $\theta^a = 1$ as can be proved by fully expanding all the expressions including those for the a_j and b_j for j = s and λ, however, and these two cases are the only ones that allow negatively sloped isoclines. In other words, the configuration of Figure 1c in Gersovitz and Hammer (2004) cannot exist.

in eqs. (12a) and (12b):

$$(14a) \qquad \lambda_s = \frac{p_a\theta^a + p_b\theta^b}{(1-s)[(\mu_\beta+\gamma_\beta)+\epsilon+s(\mu_\alpha+\gamma_\alpha)]} = \frac{p_a\theta^a + p_b\theta^b}{(1-s)(\mu_\alpha+\gamma_\alpha)(s-s^*_{max})} > 0$$

and

$$(14b) \qquad [r-(\mu_\alpha+\gamma_\alpha)(1-2s)+(\mu_\beta+\gamma_\beta)]\ \lambda_s + [p_a\theta^a_s+p_b\theta^b_s]\ \ln\lambda_s$$
$$= p_I - \frac{p_a\theta^a(1-2s)}{s(1-s)} + \frac{p_b\theta^b}{(1-s)}$$
$$+ p_a\theta^a_s \ln\left[\frac{p_a\theta^a}{\gamma_\alpha s(1-s)}\right] + p_b\theta^b_s \ln\left[\frac{p_b\theta^b}{\gamma_\beta(1-s)}\right].$$

Both these isoclines are upward sloping regardless of the form of θ^a, something that is not true in the more general case of eqs. (4) and (11c).[4] These expressions prove useful in deriving further results.

1.3 Technological Change in Prevention and Therapy

I now turn to the effects of changes in the technological parameters, the μ_j's, on the infection rate, i, and on social welfare, W, (and the utility of a representative member of society), under both the general and the special functional forms. These effects are, of course, highly dependent on the configuration of the isoclines. It makes little sense to investigate the effect of

[4]For the functional forms of eqs (6a) and (6b), the expression $(\alpha+\alpha' sa_s)$ is always positive regardless of which of the four functional forms for prevention targeting, θ^a, is under consideration. In the more general expressions of eqs. (5a) and (5b) and for values of $\theta^a = 1$ or (1-s), $(\alpha+\alpha' sa_s)$ could be either positive or negative and therefore leads to other possible configurations of the isoclines.

changes in parameters from an unstable steady state, so the discussion proceeds under the assumption of a unique stable steady state.

The effects of changes in parameters on welfare, W, are simple and independent of the effects of changes in these parameters on the optimal steady-state rate of infection. For any parameter x the effect on W is given by the dynamic envelope theorem (Caputo, 2005, chs. 9 and 14):

$$(15) \qquad \frac{dW}{dx} = \int_0^{\infty} e^{-rt} \frac{\partial H}{\partial x} \, dt \, ,$$

in which H is given by eq. (10). Consequently, an increase in μ_j, j = α or β raises W. These effects are qualitatively the same regardless of the values of the targeting functions, θ^a and θ^b, and accord with the intuitive notion of an improvement in the technology of prevention or therapy.

Next, I turn to the effect of changes in the technological parameters on the (assumed unique and stable) optimal steady-state rate of infection. In the social planner's problem, these changes are of secondary importance to the effects on welfare, but they are still important. First, knowledge of whether implementation of the social planner's program is inconsistent with movement either up or down in the infection rate after an improvement could provide a weak but observable check on whether society is acting optimally. Second, it provides a contrast with the unintervened response of the private society. Information on the signs and relative magnitudes of the slopes of the isoclines helps to determine the effects of parameter changes on the steady state rate of infection, whereas the effects of parameter changes on welfare did not depend on such information beyond the assumption that the steady state is unique and stable. To conserve space,

this section assumes that the effects of the inputs, a and b, and the effects of exogenous technical change, μ_α and μ_β, are independent so that $\alpha'_\mu = \beta'_\mu = 0$; the discussion of private decisions without optimal interventions relaxes this assumption.

An increase in μ_α: The effect on the steady-state proportion of susceptibles, s, is given by:

$$(16) \qquad \frac{ds^0}{d\mu_\alpha} = \frac{-\alpha_\mu}{\Delta}\left[\frac{s}{\Delta_{s\lambda}} + \frac{(1-2s)\lambda_s}{\Delta_{\lambda\lambda}}\right],$$

with the 0 superscript indicating that this expression only applies if $\alpha'_\mu = 0$. The first term in the brackets results from the unambiguous shift downward in the s-isocline in s-λ space (e. g. in Figure1) consequent on an increase in μ_α. This effect follows from the decrease in the ease of transmission of infection via eq. (4) everything else equal, and corresponds qualitatively to the only effect on welfare of an increase in μ_α as given by eq. (15).

But there are secondary effects on the equilibrium level of infection via the effect of μ_α on the shadow valuation of an additional susceptible person, λ_s. In the case of the social planner, this secondary effect may be opposite in sign to the impact effect, in the only possible stable case when the isoclines are sloping upward and the term $\Delta_{\lambda\lambda}$ is positive. There are two components to the effect of an increase in μ_α operating through the shadow price (and the resultant shift in the λ-isocline) and they are summarized in the term (1-2s)λ_s which is the coefficient on α in the

equation for this isocline. The term (1-2s) has its origin in the effect of a change of s on $s\pi$, the coefficient on α in the co-state eq. (11c):

$$(17) \qquad \frac{\partial s\pi}{\partial s} = \pi\frac{\partial s}{\partial s} + s\frac{\partial \pi}{\partial s} = (1-2s) \ .$$

On the one hand, the left hand term in the central piece of eq. (17) is positive. This effect arises because an increase in μ_α induces an increase in the steady-state shadow value of being susceptible (given s). The risk of infection is lower given a meeting with an infected person and so it is worth more to be susceptible because one can expect to stay in this desirable state longer. An increase in μ_α therefore shifts the λ-isocline up and increases the steady-state value of s. On the other hand, if s is higher, i and π must be lower, and therefore there is less benefit to undertaking prevention because the chance of meeting an infected person and becoming infected is lower and hence the right hand term in the central piece of eq. (17) is positive. This part of eq. (17) means that an increase in μ_α shifts the λ-isocline down and decreases the steady-state value of s. Thus to this point, it would seem possible that this last effect could not only offset the shift up in the λ-isocline coming from the first term in eq. (17) but also the shift down in the s-isocline associated with the term s in eq. (16). Based on these arguments, the net shift in the λ-isocline seems ambiguous, and the balance would be determined by the value of (1-2s) so that a value of s ≤ 0.5 is a sufficient condition for the proportion of susceptibles to rise with an increase in μ_α.

The stability of the steady state, however, allows further inferences. The expression for the effect of a change in μ_α is closely related to the relative slopes of the isoclines. In particular,

from eqs. (13) and (16):

$$(18)\qquad \frac{ds^0}{d\mu_\alpha} = \frac{-s\alpha_\mu}{\Delta}\left[\frac{-\Delta}{\Delta_{ss}} + \frac{\psi_\alpha^{SP}}{\Delta_{\lambda\lambda}\Delta_{ss}}\right]; \qquad \psi_\alpha^{SP} = sp_a(a_s\theta_s^a + a\theta_{ss}^a) - \lambda\alpha + \lambda(1-s)\beta' b_s \ .$$

The right-hand side of eq. (18) is positive in the case of stable, upward-sloping isoclines because $\Delta>0,\ \Delta_{\lambda\lambda}>0,\ \Delta_{ss}<0$ *and* $\psi_\alpha^{SP}<0$; the last inequality can be proved by substituting for the targeting functions. So despite the seemingly ambiguous shifts in the isoclines, the net effect on s of an increase in the technology of prevention is always positive.

An increase in μ_β: If both isoclines are upward sloping, as in Figure 1, an increase in μ_β results in an unambiguous lowering of the s-isocline via eq. (4). All other things equal, it makes it easier to recover and lowers the infection rate. But all other things are not equal, and an increase in μ_β also unambiguously lowers the λ-isocline for a given value of s. Intuitively, for a given value of s, a higher value of μ_β makes it easier to recover and less valuable to be susceptible (or, equivalently, less expensive to be infected). On balance, however, the effect of an increase in μ_β on the steady-state rate of susceptibility, s, is:

$$(19)\qquad \frac{ds^0}{d\mu_\beta} = \frac{\beta_\mu}{\Delta}\left[\frac{1}{\Delta_{s\lambda}} - \frac{\lambda_s}{\Delta_{\lambda\lambda}}\right] = \frac{-\beta_\mu}{\Delta}\left[\frac{\Delta}{\Delta_{ss}} - \frac{\psi_\beta^{SP}}{\Delta_{\lambda\lambda}\Delta_{ss}}\right],$$

in which $\beta_u' = 0$ *and* $\psi_\beta^{SP} = p_a(a_s\theta_s^a + a\theta_{ss}^a) - \lambda[\alpha - (1-s)\alpha' a_s] < 0$ except, possibly but not necessarily, if targeting is $\theta^a = 1$ and s >0.5. Except possibly in this one sub-case, therefore, an

improvement in the technology of therapy increases the proportion of susceptibles in the case of the social planner. Furthermore, in the exponential case, an increase in μ_β increases the steady-state rate of susceptibility, s, even in this one special sub case of $\theta^a = 1$ and $s > 0.5$ as can be shown by using the expressions for the exponential case, eqs. (6a), (6b) and (14a) in eq. (19).

2. Private Choice, Disinhibition and Immiserization

2.1 The Formulation of Private Choice:

The social planner directly controls the values of a and b in a model without people who make decisions that affect their own health. The next step is to consider private decisions and their implications for public health and government policy. In their decisions about prevention and therapy, people generate a negative externality if they fail to take into account the effect of their becoming and staying infected on the risk of infection to others. After identifying the externality, one can examine how government interventions with subsidies or taxes can decentralize the social planner's first-best solution, something done in Gersovitz and Hammer (2004). This paper examines the effect of changes in the technological parameters on people's decisions and welfare when the social planner is not intervening to correct the externality and compares the outcomes to the social planner's.

The simplest way to illustrate the infection externality and its implications for policy is to assume that private decisions are made by a group of people termed the household; see Gersovitz and Hammer (2004) on this representative agent. The only distinction between the social planner and the representative agent is that the household is assumed to be small relative to the population

as a whole, in this case so that the infected proportion of the household does not affect the infected proportion of the population as a whole. The household takes as exogenous to its decisions the proportion of the population that is infected, which equals the probability, π, that any random contact is with an infected person. The household is assumed to be sufficiently large that it can fulfill the role of a representative agent and therefore that the infected proportion of the household is identical to the corresponding population proportion. Finally, it is this household that takes decisions about the interventions, a and b. Because the instantaneous utility function is linear, there is no sense in which the household is performing any implicit insurance function for its members. The equations for the model of private decision making are therefore the same as in section 1.1, except that the household takes the probability of infection, $\pi = i$, to be exogenous.

2.2 Private Choice, Disinhibition and Immiserization:

The Hamiltonian for the household's problem is the same as the social planner's, eq. (10), except that π is exogenous. All functions of variables (θ, α, and β) are evaluated at the household values of their arguments and ϵ is a constant common to both the social planner's and the household's problems.

The first derivatives of H with respect to the controls, a and b, set equal to zero imply eqs. (11a) and (11b) and the condition for the change in the multiplier on population, eq. (11d), remains the same, although as before it is not relevant to the solution for the variables of interest. What is different is the dynamic equation for the multiplier on the proportion of the population that is susceptible, s, which yields:

$$(20) \qquad \dot{\lambda}_s = r\lambda_s - [p_1 - p_a a\theta_s^a - p_b b\theta_s^b] + [\alpha\pi + \beta]\lambda_s \quad .$$

Because the household is representative of society, π must equal (1-s). Once this substitution is made, the only difference between eqs. (11a-c), the planner's problem, and eqs. (11a), (11b) and (20), the private problem, is the (1-s) term at the end of eq. (20) rather than the (1-2s) term at the end of equation (11c). This latter difference arises because the household takes the society-wide infection rate as exogenous in its decisions. The social planner's first best can be decentralized by subsidies at equal rates to preventive and therapeutic activities; see Gersovitz and Hammer (2004) for proof and intuition.

In the absence of subsidies, however, the first-best is not realized. The isocline for s is identical to that of the social planner's and therefore traces out the same curve, for example as in Figure 1 if the s-isocline slopes upward. The λ-isocline for the general case is given by setting the right hand side of eq. (20) to zero. The λ-isocline has slope given by:

$$(21) \qquad \left[\frac{\partial \lambda_s}{\partial s}\right]_{\lambda_s=0} = \frac{-\{p_a(a_s\theta_s^a + a\theta_{ss}^a) - \lambda_s[\alpha - (1-s)\alpha' a_s] + \beta'\lambda_s b_s\}}{[r - s\alpha + \beta + (1-s)(\alpha + \alpha' s a_s) + s(\alpha + \lambda\alpha' a_\lambda) + (\lambda_s\beta' + p_b\theta_s^b)b_\lambda]}$$

$$= \frac{-\Gamma_{\lambda s}}{\Gamma_{\lambda\lambda}} = \frac{?}{?},$$

in which the Γ_{jk}, j = λ and k = s, λ denote the corresponding numerator and denominator that precede their introduction. The slope of the λ-isocline is more difficult to characterize in the case of private decision making than in the case of the social planner, eq. (12b), and there are more sub-cases, only one of which is considered in this paper:

First, the term $\Gamma_{\lambda s}$ need no longer be negative as $\Delta_{\lambda s}$ had to be in eq (12b). In fact, however, this term can be positive only in the case $\theta^a = 1$, and even then a sufficient (but not necessary) condition for it to be negative is $s \leq 0.5$ as can be proved by substitution using eqs.

(11a) and (11b). Assumption P.1 states that: the numerator of eq. (21) is positive as it is in eq. (12b).

Second, the term $\Gamma_{\lambda\lambda}$ need not be positive about the steady state even if $(\alpha+\alpha'sa_s) > 0$ and $r \geq \epsilon$ as the corresponding term, $\Delta_{\lambda\lambda}$, is in the social planner's problem, eq. (12b). In the characterization of the social planner's problem, this property is important in reducing the number of cases because $(\alpha+\alpha'sa_s)> 0$ also ensures that the s-isocline is upward sloped and furthermore, this property must obtain for θ^a = s or s(1-s). But in the private problem, the term $(\alpha+\lambda\alpha'a_\lambda)$ appears in the denominator of eq. (21) and can be negative or positive for θ^a = 1, (1-s) or s(1-s) regardless of the sign of $(\alpha+\alpha'sa_s)$. If θ^a = s and if $r \geq \epsilon$, however, the denominator of eq. (21) must be positive as can be shown by expanding all the terms. Assumption P.2 states that: the denominator of eq. (21) is positive if the numerator of eq. (12a) is.[5]

Define the difference between the slope of the s-isocline and the slope of the λ-isocline as Γ:

$$(22) \qquad \Gamma = -\frac{\Delta_{ss}}{\Delta_{s\lambda}} + \frac{\Gamma_{\lambda s}}{\Gamma_{\lambda\lambda}} = \Delta_{ss}\left[\frac{-1}{\Delta_{s\lambda}}+\frac{\lambda}{\Gamma_{\lambda\lambda}}\right] + \frac{[p_a(a\theta^a_{ss}+\theta^a_s a_s)+\lambda\alpha'a_s]}{\Gamma_{\lambda\lambda}}$$

The following analysis only applies if Assumptions P1 and P2 hold. Under P1 and P2, qualitative

[5]Assumption P.2 rules out the case in which the s-isocline is upward sloped and the λ-isocline is downward sloped. This configuration produces a stable steady state in which the real parts of the roots of the characteristic equation are both negative. It seems possible for these roots to be either real or imaginary, and so the steady state is either a star or the focus of stable cycles. But this information is not adequate to characterize the optimal path; see the references in Gersovitz (2012).

properties of the isoclines discussed in section 1.2 carry over from the social planner's problem to the private problem. In particular, the only saddlepoint stable configuration of the isoclines is that both are upward sloping and the s-isocline is steeper than the λ-isocline. Furthermore, if the slopes of the two isoclines are positive and produce a stable configuration, the private λ-isocline is positioned below the social planner's λ-isocline so that the steady-state value of s is lower in the unintervened private case than in the social planner's case. This result accords with the intuition that the externality leads to private individuals doing too little to avoid being infected, either through prevention or therapy.

For the exponential, the λ-isocline is given by:

$$(23) \qquad [r-(\mu_\alpha+\gamma_\alpha)(1-s)+(\mu_\beta+\gamma_\beta)]\ \lambda_s + [p_a\theta_s^a+p_b\theta_s^b]\ \ln\lambda_s = p_I - \frac{p_a\theta^a}{s} + \frac{p_b\theta^b}{(1-s)} + p_a\theta_s^a \ln\left[\frac{p_a\theta^a}{\gamma_\alpha s(1-s)}\right] + p_b\theta_s^b \ln\left[\frac{p_b\theta^b}{\gamma_\beta(1-s)}\right]$$

which differs from eq. (14b) of the social planner's problem to reflect the appearance of (1-s) rather than (1-2s) as appropriate. Eq. (23) always slopes upward in s-λ_s space, but is positioned lower than eq. (14b) and eq. (14a) slopes upward. The exponential case therefore satisfies Assumptions P1 and P2. Thus, whether decisions are taken by the social planner or by private households without government intervention, both isoclines slope up in the exponential case.

2.3 Technological Change in Prevention and Therapy

Starting from a steady state, the effect of a change in a parameter, x, on welfare, W, is

more complicated when decisions are made by private individuals without the social planner intervening with optimal first-best subsidies. The equation determining the outcome is:

$$(24) \qquad \frac{dW}{dx} = \int_0^\infty e^{-rt}\left[\frac{\partial H}{\partial x} + \frac{\partial H}{\partial \pi}\frac{d\pi}{dx}\right] dt = \int_0^\infty e^{-rt}\left[\frac{\partial H}{\partial x} + \lambda_s N \alpha s \frac{ds}{dx}\right] dt \ .$$

This result is an extension of the dynamic envelope theorem to account for the fact that the household ignores its effect on s through its effect on π (the derivation is the same as Caputo, 2005, chs. 9 and 14). The term $\partial H/\partial x$ has the same sign regardless of whether it derives from the social planner's or the household's problem, although the values at which it is evaluated differ between the two situations. Because the model has only one state variable, the state moves monotonically from one steady state to another so the sign of ds/dx is the same along the path as the sign of the difference between steady states. A sufficient condition for an improvement in welfare caused by a change that is a technological improvement (one that increases welfare in the social planner's problem) is therefore that the rate of infection fall in the long-run steady state so that ds/dx is always positive. Eq. (24) shows that there is no other potential source for immiserization. It does not seem possible to use $\partial H/\partial x > 0$ to strengthen this sufficient condition.

Immiserization arises if the expression involving ds/dx in eq. (24) is of the opposite sign and dominates in magnitude $\partial H/\partial x$ in the private problem. In this situation a change in a parameter that increases welfare when the social planner controls health inputs (or equivalently decentralizes the social planning problem through first-best interventions) decreases welfare when the household makes decisions in the absence of any interventions by the social planner. In the case of immiserization, the change in the parameter which would otherwise be advantageous is

exacerbating the externality to such a degree that this latter, seemingly secondary, effect dominates. For instance, a bio-medical innovation that makes either prevention or therapy more likely to succeed, other things equal, and thereby must improve welfare under the social planner, may encourage so much private behavior that increases the size of the infection pool (and the externality) that on net society is worse off. Eq. (24) makes this notion precise. The next step is to understand how changes in the technological parameters affect the (unique, stable) steady-state rate of infection.

An increase in μ_α: The effect on the proportion of susceptibles of an improvement in prevention technology if the effects of the prevention input and the improvement in prevention technology are separable ($\alpha_\mu < 0, \alpha'_\mu = 0$) is:

$$(25) \qquad \frac{ds^0}{d\mu_\alpha} = \frac{-\alpha_\mu}{\Gamma}\left[\frac{s}{\Delta_{s\lambda}} + \frac{(1-s)\lambda_s}{\Gamma_{\lambda\lambda}}\right]$$

This expression is strictly positive if the isoclines are positively sloped with the s-isocline steeper (the case of the saddlepath stable steady state) and so the improvement in the technology of prevention unambiguously improves welfare with private decision making. The first term in the brackets results from the unambiguous shift downward in the s-isocline (Figure 1) consequent on an increase in μ_α and has the same rationale as in the case of the social planner. In contrast to the situation of the social planner, however, there is only one secondary effect on the equilibrium level of infection via the effect of μ_α on the shadow valuation of an additional susceptible person, λ_s.

The fact that the private decision makers ignore the external effects of their actions is

paradoxically what makes it easier to demonstrate that an improvement in prevention technology rules out an increase in the infection rate in the case of private decision makers and therefore the possibility of immiserization. Because the isocline slopes upward (the term $\Gamma_{\lambda\lambda}$ is positive), the right hand piece in brackets in eq. (25) involving $(1\text{-}s)\lambda_s$ is positive. Consequently, an increase in μ_α increases the shadow value of being susceptible for a given value of s, and shifts the λ-isocline upward. The risk of infection is lower given a meeting with an infected person and so it is worth more to be susceptible because one can expect to stay in this desirable state longer. Thus an increase in μ_α shifts the s-isocline downward and the λ-isocline upward so that the steady state values of both s and λ_s rise.

Although the effect on the infection rate and welfare is therefore unambiguous in the saddlepoint stable configuration of the isoclines, how this change comes about can be quite varied. In particular, the inputs controlled by private decision makers, a and b, can respond in different ways to the increase in μ_α. The only constraint is that the steady state value of the proportion of susceptibles, $s = [\beta(b) + \epsilon]/\alpha(a)$, increases so the changes in a and b taken together with the increase in μ_α must produce this result. Under the assumption of the independence of μ_α and a, a change in μ_α does not affect directly either of the first-order conditions, eqs. (11a) and (11b). The adjustment of the inputs to a change in μ_α therefore operates via these equations only through the effects of μ_α on s and λ_s:

$$(26a) \qquad \frac{da}{d\mu_\alpha} = a_s \frac{ds}{d\mu_\alpha} + a_\lambda \frac{d\lambda}{d\mu_\alpha}$$

and

$$(26b) \qquad \frac{db}{d\mu_\alpha} = b_s \frac{ds}{d\mu_\alpha} + b_\lambda \frac{d\lambda}{d\mu_\alpha} .$$

From eqs. (11a) and (11b), a_λ and $b_\lambda > 0$. Two unambiguous cases of what can happen are: if $\theta^a = (1-s)$ so that $a_s > 0$ or $\theta^a = s(1-s)$ so that $a_s = 0$, and if $\theta^b = (1-s)$ so that $b_s = 0$ then both a and b increase when μ_α increases with a technological innovation that makes the disease less infectious.

In these cases, there is the opposite of any kind of disinhibition as the use of both health inputs increases, not just an absence of the kind of disinhibition that is so extreme that the infection rate rises. If $\theta^b = (1-s)$, the use of therapy increases in response to the increase in μ_α regardless of what happens to the use of prevention and (its targeting).

The effect on the proportion of susceptibles of an improvement in prevention technology if the effects of the prevention input and the improvement in prevention technology are not separable ($\alpha'_\mu \neq 0$) is:

$$(27) \qquad \frac{ds}{d\mu_\alpha} = \frac{-1}{\Gamma}\left[\frac{s(\alpha_\mu+\alpha' a_\mu)}{\Delta_{s\lambda}} + \frac{(1-s)\lambda_s(\alpha_\mu+\alpha' a_\mu)+p_a\theta_s^a a_\mu}{\Gamma_{\lambda\lambda}}\right]$$

$$= \frac{ds^0}{d\mu_\alpha} - \frac{\alpha' a_\mu}{\Gamma\Delta_{s\lambda}\Gamma_{\lambda\lambda}}\left[s(r-\epsilon+\alpha)+\lambda_s\beta' b_\lambda\left(1+\frac{s(1-s)\theta_s^b}{\theta^b} - \frac{s(1-s)\theta_s^a}{\theta^a}\right)\right],$$

in which $ds^0/d\mu_\alpha$ is given by eq. (25). This expression is strictly positive if $\alpha'_\mu < 0$, implying $\partial a/\partial \mu_a = a_\mu > 0$ from differentiation of eq. (11a), because the expression in parentheses on the

last line of eq. (27) is always positive regardless of the targeting functions. If $\alpha'_\mu > 0$ so that an improvement lowers the marginal product of the preventive health input, however, it is possible that s decreases and immiserization cannot be ruled out.

An increase in μ_β: If the effects of the therapeutic input and the improvement in therapeutic technology are separable ($\beta_\mu > 0,\ \beta'_\mu = 0$), then the effect of an increase in μ_β on the steady state proportion of people who are susceptible is:

$$(28) \qquad \frac{ds^0}{d\mu_\beta} = \frac{\beta_\mu}{\Gamma}\left[\frac{1}{\Delta_{s\lambda}} - \frac{\lambda_s}{\Gamma_{\lambda\lambda}}\right] = \frac{-\beta_\mu}{\Gamma}\left[\frac{\Gamma}{\Delta_{ss}} - \frac{\psi^P_\beta}{\Gamma_{\lambda\lambda}\Delta_{ss}}\right],$$

in which $\psi^P_\beta = p_a(a_s\theta^a_s + a\theta^a_{ss}) + \lambda\alpha' a_s$. In the case of upward sloping stable isoclines ($\Delta_{ss} < 0,\ \Gamma_{\lambda\lambda} > 0,\ \Gamma > 0$), a sufficient condition for an increase in μ_β to increase s is that ψ^P_β is negative, which is the case regardless of θ^b when: $\theta^a = 1$ and $s < 0.5$, $\theta^a = (1-s)$ or $\theta^a = s(1-s)$. In the other cases, $\theta^a = 1$ and $s > 0.5$ or $\theta^a = s$, an increase in μ_β may still increase s but $\psi^P_\beta > 0$ so that it depends on the relative magnitudes of both terms inside the brackets in eq. (28). These ambiguous cases have in common that they are the only ones in which $a_s < 0$, so that an increase in s lowers preventive effort all other things equal.

In the ambiguous case when $\theta^a = s$, further re-arrangement of the terms in eq. (28) shows that:

$$(29) \qquad \frac{ds^0}{d\mu_\beta} = \left\{ (r-\epsilon) + \alpha - s\alpha \frac{\eta_\alpha}{\eta_{\alpha'}} - \frac{\beta\theta_s^b \eta_\beta}{\eta_{\beta'}} \right\} \frac{1}{\Gamma\Delta_{s\lambda}\Gamma_{\lambda\lambda}} .$$

The higher is the difference between r and ϵ, the lower the absolute value of the elasticity of α with respect to a ($\eta_\alpha = \alpha' a/\alpha$) and the higher the absolute value of the elasticity of α' with respect to a ($\eta_{\alpha'} = \alpha'' a/\alpha'$), the more likely is an increase in μ_β to increase s. If $\theta^b = (1-s)$, the lower is the absolute value of the elasticity of β with respect to b ($\eta_\beta = \beta' b/\beta$) and the higher is the absolute value of the elasticity of β' with respect to b ($\eta_{\beta'} = \beta'' b/\beta'$), the more likely is an increase in μ_β to increase s; if $\theta^b = 1$, however, these two elasticities are irrelevant to the outcome. An increase in μ_β is more likely to increase s when the initial value of s is low. As with all elasticity conditions, the results in this paragraph all presuppose other things equal in what is actually an endogenous system, so these inferences need to be interpreted cautiously. For instance, an increase in r -ϵ may simultaneously affect the elasticities. That eq. (29) seems to allow for the possibility that the infection rate may actually rise with an improvement in therapy means that there is a risk of immiserization and that the need to investigate the question is justified; the answer is not obvious even under the assumption of independence of b and μ_β.

For the exponential of eqs. (6a) and (6b), however, the infection rate decreases with an improvement in the technology of therapy, $ds^0/d\mu_\beta > 0$, regardless of targeting. The middle expression in brackets in eq. (28) is always positive regardless of targeting as can be proved by

substituting the expressions for the exponential under the different targeting cases and collecting terms. Even if $\psi_{\beta}^{P} > 0$ in the general case as must always be true for $\theta^{a} = s$, the exponential makes clear there is no presumption that $ds^{0}/d\mu_{\beta} < 0$ let alone that this term could be sufficiently negative to be the dominant effect in the expression for the change in welfare as given in eq. (24).

If the effects of the therapeutic input and the improvement in therapeutic technology are not separable ($\beta'_{\mu} \neq 0$), then the effect of an increase in μ_{β} on the steady state proportion of people who are susceptible is:

$$(30) \qquad \frac{ds}{d\mu_{\beta}} = \left[\frac{\beta_{\mu}+\beta' b_{\mu}}{\beta_{\mu}}\right]\frac{ds^{0}}{d\mu_{\beta}} - \left[\frac{p_{b}\theta_{s}^{b} b_{\mu}}{\Gamma\Gamma_{\lambda\lambda}}\right].$$

If $\beta'_{\mu} > 0$ so that $\partial b/\partial \mu_{\beta} = b_{\mu} > 0$ from differentiation of eq. (11b) and if in addition $ds^{0}/d\mu_{\beta} > 0$ as given in eq. (28), then an improvement in therapeutic technology increases the proportion of susceptibles because the term in the right hand brackets in eq. (30) is either positive or zero. If, however, $\beta'_{\mu} < 0$ so that $b_{\mu} < 0$ and to such a degree that $\beta_{\mu}+\beta' b_{\mu} < 0$ and if $ds^{0}/d\mu_{\beta} > 0$, the effect of an improvement in therapeutic technology could be to decrease the proportion of susceptibles and immiserization cannot be ruled out.

Recall that if there is no private choice with regard to the input for which there is technological improvement, then the condition for separability between the private input and the innovation is met automatically. This observation has an important practical implication because

the case in which technological improvement is to the therapy and the individual's choice is about prevention is likely to be a common one.

3. Conclusions

In the case of an SIS disease, this paper provides guidance on when technological improvements in prevention or therapy can lead to an increase in infection rates and even a decrease in welfare (immiserization) when decisions are taken privately without optimal government interventions that internalize externalities. The results are sensitive to assumptions about how prevention and therapies are targeted. In many cases, the condition for disinhibition sufficient to raise the infection rate and therefore the necessary condition for immiserization is not met. If health inputs and the innovation are separable, immiserization is impossble in the case of preventive innovation and only possible in the case of therapeutic innovation for a minority of preventive targeting possibilities. If the relations between preventive and therapeutic inputs and the corresponding outputs take the special form of the exponential which includes the assumption that health inputs and the innovation are separable, immiserization is impossible regardless of targeting.

A welfare-improving outcome is more likely when the improvement is to prevention rather than to therapy. On first consideration, this result seems counter-intuitive. A lower value of the infection-generating parameter, α, consequent on an improvement to prevention and a higher value of the recovery-generating parameter, β, consequent on an improvement to therapy both tend to lower the infection rate for an SIS disease. In the former case, an improvement keeps people from becoming infected while in the latter case, an improvement gets people out of the

infected pool. This latter effect immediately suggests that the results of the SIS model may not generalize to a disease such as HIV in which an improvement in therapy keeps people in the infected pool raising the infection rate other things equal.[6] But it is not these impact effects that differ between prevention and therapy in an SIS model. Rather the asymmetry arises because an improvement in prevention tends to make being uninfected more valuable, thereby promoting additional effort to stay uninfected, whereas an improvement in therapy tends to make being uninfected less valuable. What the corresponding results are for other, non-SIS diseases remains to be investigated, building on the models of Gersovitz and Hammer (2004).

Of course, it is always best if the internalizing interventions are implemented. Not only is immiserization from innovation impossible, but the full benefits of the innovation are certain to be realized. Sometimes, however, it may not be possible to implement the internalizing interventions, especially when the activities that spread infection may be expensive to monitor or even effectively impossible to monitor as when these activities are intimate and private. In these circumstances, if disinhibition is so extreme as to produce immiserization, the only option may be to suppress innovations that would otherwise raise welfare. It is the prospect of the need for such a dire second-best response that motivates investigation of the conditions under which immiserization could occur.

4. References

Caputo, M.R., Foundations of Dynamic Economic Analysis (Cambridge: Cambridge University Press, 2005).

[6]See Lakdawalla et al (2006) for an analysis that relates to this question.

Gersovitz, M., “Births, Recoveries, Vaccinations and Externalities,” in Essays in Honor of Joseph E. Stiglitz eds., R. J. Arnott, et al. (Cambridge: MIT, 2003) pp. 469-483.

Gersovitz, M., “Infection Control and Uniqeness in a Model of Susceptible-Infected-Susceptible (SIS) Diseases,” mimeo, Johns Hopkins University, 2012.

Gersovitz M, Hammer JS. 2003. Infectious diseases, public policy, and the marriage of economics and epidemiology. The World Bank Research Observer. 18: 129-57

Gersovitz, M. and J. S. Hammer, "The Economical Control of Infectious Diseases," The Economic Journal, vol.114, (2004), pp. 1-27.

Heffley, D. R., “Allocating Health Expenditures to Treatment and Prevention,” Journal of Health Economics, vol. 1 (1982), pp. 265-290.

Lakdawalla, D. et al, “HIV Breakthroughs and Risky Sexual Behavior,” Quarterly Journal of Economics (2006), pp. 1063-1102.

Ma, Z. and J. Li (eds.), Dynamical Modeling and Analysis of Epidemics (New Jersey: World Scientific, 2009).

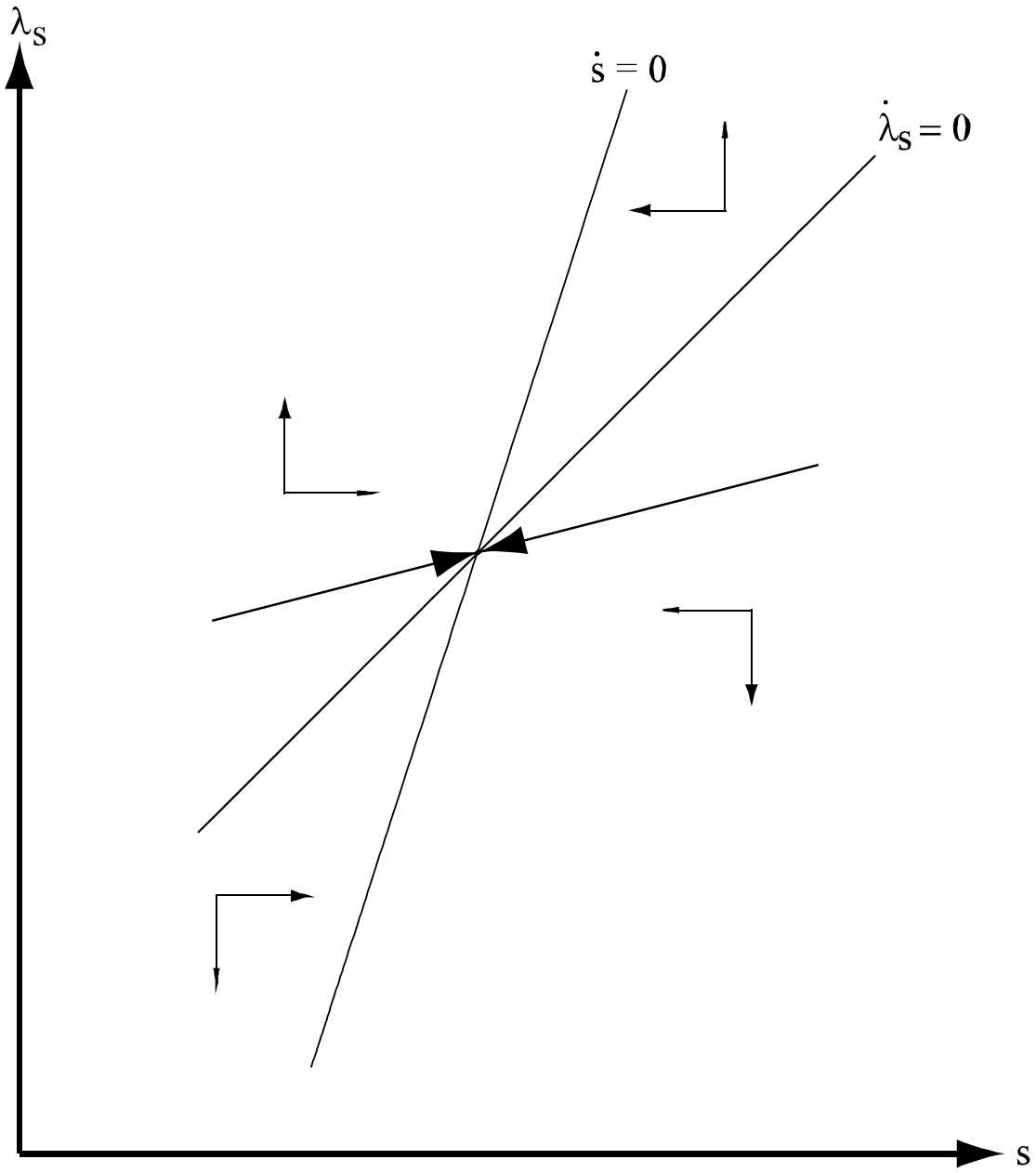

Fig. 1
The Phase Diagram of an SIS Disease: A Stable Case