Rational Disinhibition and Externalities in Prevention

Flavio Toxvaerd*

January 14, 2017

Abstract. This paper studies a model of disease propagation in which individuals can control their exposure to infection by engaging in costly preventive behavior. Individuals are fully rational, strategically sophisticated and forward-looking. Equilibrium outcomes under decentralized decision making are characterized and contrasted to the outcomes chosen by a benevolent social planner. In general, individuals over-expose themselves to infection, leading to suboptimally high disease prevalence. This stems from two separate effects, a pure externality effect and a smallness effect. The model is applied to study the welfare effects of pre-exposure prophylaxis, which reduces transmission between serodiscordant couples and causes disinhibition. It is shown that a decrease in the induced infection risks increases disease prevalence and can lead to decreases in overall welfare.
JEL Classification: C73, I18.
Keywords: Economic epidemiology, preventive behavior, disinhibition, risk compensation, second-best policies.

1. Introduction

Since 1981, the AIDS epidemic has claimed more than 25,000,000 lives and ruined the lives of innumerably more.[1] From an historical perspective, these magnitudes are not unusual. Between 1348 and 1350, the bubonic plague (the Black Death) caused the death of approximately a third of Europe's total population of 100,000,000, while the influenza epidemic (the Spanish flu) between 1918 and 1920 claimed a world death toll estimated at 50,000,000.[2] In more recent times, the advent of bird flu and swine flu have served as salutary reminders that communicable diseases are of first order importance as a matter of public policy. Furthermore, they constitute a promising and worthwhile field of inquiry for economists and other social scientists.

While the study of infectious diseases has traditionally belonged to the domain of biology and medical science, policy and welfare concerns have always been a raison d'être of epidemiological inquiry. In motivating his formal analysis of infectious diseases, the father of mathematical epidemiology Daniel Bernoulli (1766) stated that

> "I simply wish that, in a matter which so closely concerns the well-being of mankind, no decision shall be made without all the knowledge which a little analysis and calculation can provide."

In order to formulate sensible public policies to manage infectious diseases through preventive measures, knowledge of two central issues is necessary. First, the policy maker must determine what the optimal outcomes are. Even if such outcomes are beyond reach

*Faculty of Economics, University of Cambridge. Email: fmot2@cam.ac.uk.
[1] See UNAIDS 2008 Report on the Global AIDS Epidemic.
[2] See Langer (1970) and Johnson and Mueller (2002).

in practice, they serve as a useful benchmark for formulating policy and for understanding when and why markets fail in implementing such outcomes. Second, the policy maker must determine which equilibrium outcomes would emerge in a setting with fully decentralized, forward-looking and non-cooperative decision making by individuals.

In almost all instances, public policy is to some extent implemented by influencing the tradeoffs faced by individuals (rather than by mere fiat) and thus it is important to fully understand these tradeoffs and the resulting equilibrium patterns of prevention. Once this is achieved, the design of policy becomes better informed and can be carried out on a more solid foundation.

There is a growing literature on the design of very sophisticated policy interventions such as taxes and subsidies, that may help implement socially desirable outcomes (see Rowthorn and Toxvaerd, 2015). But less is known about the outcomes and desirability of simpler and more easily implementable second-best policy interventions. This paper seeks to fill this void and contribute to the policy debate by analyzing one of the main such policies, known as pre-exposure prophylaxis (aka PrEP).

On July 16, 2012, the Food and Drug Administration announced its approval of the drug Truvada for use as a pre-exposure prophylaxis against HIV.[3] Drugs like Truvada have been shown to reduce the probability of contracting HIV from non-protected sex with an infected individual and have therefore been welcomed as a potentially important new tool in controlling the HIV epidemic.[4] Significant undertakings to combat HIV/AIDS, such as the Getting to Zero campaign promoted by UNAIDS (2010), are centered around the extensive use of pre-exposure prophylaxis. In fact, PrEP has been widely lauded as a "game changer" in the battle against HIV infection, raising expectations that it may radically alter the calculus of HIV control.[5]

Regrettably, there is a considerable lack of clarity about both the likely effects and the desirability of such medical innovations and the issue of so-called risk compensation remains controversial. Risk compensation (or disinhibition) is said to occur, if a decision maker responds to decreased risk, by engaging in *more* risky behavior. With refreshing candor, Susan Buchbinder of the San Francisco Department of Public Health, in discussing the introduction of Truvada, stated that

> "We don't really know what is going to be happening with risk compensation in a real world setting".[6]

While progress in determining the real world effects of this introduction will no doubt be made over the next few years, it is important to lay the theoretical foundations for critically evaluating such innovations. Unfortunately, not much progress has been made on this front. In his paper, I consider second-best interventions such as medical innovations that permanently reduce the infectiousness of a disease and their relation to disinhibition. Almost universally, the policy debate on disinhibition focuses only on the presence or otherwise of disinhibition. But arguably, whether there is disinhibition is a

[3] See http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm312210.htm.

[4] See Karim et al. (2010).

[5] See e.g. The Guardian, August 2, 2016: *Judge to rule on NHS funding for 'game-changing' HIV treatment* and BBC, February 25, 2015: *Analysis: An HIV 'game changer'?*

[6] Quoted in Werble (2012).

moot point, unless its welfare consequences can be assessed. Contributions to the literature that engage with such welfare issues are few and far between and will be discussed below. In almost all analyses, disinhibition is implicitly taken to be an undesirable behavior response that should be avoided or mitigated through parallel interventions, such as risk counselling. I show that this perspective may be misleading and that disinhibition may indeed be socially optimal.

My analysis contributes to this important policy debate by addressing the following two fundamental questions: *(i) Does the introduction of pre-exposure prophylaxis to the uninfected cause disinhibition? (ii) Does the introduction of pre-exposure prophylaxis to the uninfected decrease social welfare?*

To preview the answer to the first question, my analysis confirms that a permanent decrease in the infectiousness of the disease will prompt an increase in exposure and in steady state disease prevalence. The reason is that such a change alters the tradeoffs faced by decision makers in favour of increased exposure. Although the decrease in infectiousness decreases the amount of transmission per exposure, the exposure itself increases so much that the net effect is to increase disease incidence. This outcome is not pathological and holds both for a utilitarian social planner and for self-interested individuals.

To preview the answer to the second question, my analysis shows that such a change in the infectiousness of the disease may, but need not, lead to a decrease in social welfare. The reason is that individual decision makers make different choices than those preferred by a social planner. This is because of the existence of a pure externality effect and of a so-called smallness effect, which together drive a wedge between private and social incentives to engage in costly preventive behavior. The strength of these effects changes across the stages of the epidemic and therefore influence the ex-ante intervention and the ex-post intervention steady state values under centralized and decentralized decision making, respectively. As a consequence, welfare along transition paths induced by interventions may be different under centralized and decentralized decision making. For this reason, the introduction of pre-exposure prophylaxis may well lead to a decrease in social welfare. That this possibility can materialize is verified numerically.

These findings point to two important policy conclusions. First, the customary focus on disinhibition in the policy debate is patently misplaced. When once decreases the infectiousness of a disease, it may be perfectly desirable that individuals increase exposure and hence cause disease prevalence to rise. After all, to the extent that individuals derive utility from such exposure, it is right that optimal policy should take that into account when trading off the costs and benefits of such exposure. If one is unwilling to trade off infection risk with the private benefits of exposure, then one must accept the conclusion that there should be no exposure at all.

Second, the presence of external effects may lead policy interventions to have undesirable consequences. In particular, a well-intentioned second-best policy to reduce infections may in fact exacerbate the problem of over-exposure and lead to socially inferior outcomes. These conclusions are a testament to the fact that controlling infectious diseases is at best complicated. Careful analysis often leads to counter-intuitive conclusions and policy proposals must therefore be carefully evaluated before they are acted on, as called for in the excerpt of Bernoulli's original work.

1.1. Related Literature. This paper contributes to an important and growing literature that studies the effects and desirability of permanently decreasing the infectiousness of diseases through medical innovations. Lakdawalla et al. (2006) consider the effects of simultaneously introducing antiretroviral therapy to susceptible and infected individuals in a *susceptible-infected* (SI) type model. They note that there are several confounding effects. Ceteris paribus, the susceptible individuals may benefit, because the introduction of antiretrovirals effectively reduces the probability of transmission per risky sexual act. On the other hand, antiretrovirals increase the survival probability of infected individuals, thereby increasing the source of infection. Unfortunately, Lakdawalla et al. (2006) do not distinguish between equilibrium outcomes and socially optimal outcomes and it is therefore not a priori clear that policy interventions are warranted in the first place.[7] Last, their welfare statements focus on steady state values. As shown in the present paper, focusing on steady states can lead to very misleading conclusions. Chan et al. (2012) consider the effects of introducing antiretroviral therapy to susceptible individuals and conclude that they are likely to benefit from reduced infectivity. Gersovitz (2010) considers innovation in an infection context, but mainly offers results for technological improvements that do not interact directly with individual's decisions. The technological improvements he considers cannot be interpreted as PrEP and the analysis is therefore not directly comparable to mine.

Disinhibition and related effects have been noted in several different contexts.[8] In energy and conservation economics, the so-called rebound effect has been recognized as a real possibility (see Gillingham et al., 2015 for a review of this literature). The rebound effect is said to be present if an innovation that is intended to reduce the use of a resource, instead acts to increase the use of said resource. The rebound effect is also known as the Peltzman effect, articulated in the context of safety equipment and driving. In traffic and network economics, the "fundamental law of road congestion" posits that an improvement of a road network may lead to increased congestion (Duranton and Turner, 2011). In the context of protection against infectious diseases, such effects are known as disinhibition.[9] Interestingly, an almost universal feature of papers that treat disinhibition and related effects is that they focus almost exclusively on whether the effect is present and do not deal with the welfare consequences of such effects (see e.g. Richens et al., 2000, Cassell et al., 2006 and Blumenthal and Haubrich, 2014). A notable exception is Chan and Gillingham (2015), who analyze the welfare consequences of the rebound effect in a static model of energy use.[10] Outside the realm of economics, the theory of risk homeostasis has been proposed (see Wilde, 1982). This theory, which is not based on maximization, holds that individuals are characterized by the amount of risk that they are willing to bear and that an innovation that reduces the risk of an activity, may induce the individual to engage in more of said activity.

It is difficult to analyze fully dynamic models of disease propagation analytically and it is therefore very common in the literature to focus the analysis on the steady state(s) of

[7] As shown elsewhere in this paper, in the model when no recovery is possible, i.e. in a standard SI model, equilibrium is socially optimal unless the population is heterogeneous.

[8] See Hedlund (2000) for a non-technical survey of the literature on risk compensation.

[9] Autier et al. (1999) find that the introduction of high SPF sun creams can increase sunbathing and associated risks.

[10] Duranton and Turner (2009) perform a related static welfare analysis in the context of road congestion.

such systems. Papers that take this route include Kremer (1996), Barrett (2003), Kremer and Snyder (2003), Auld (2003), Lakdawalla et al. (2006), Stashko (2012), Greenwood et al. (2013), Cerdeiro (2015) and Bate et al. (2016). No doubt, the steady state(s) of a dynamic system are important and merit attention. But to focus on steady states only can be misleading, a fact seen very clearly in the welfare analysis of policy interventions designed to change behavior. One important contribution of this paper is to show just how misleading such steady state analysis can be. In the present analysis, the costs and benefits of policy interventions are dynamic in nature. Their merits must therefore be assessed by carefully analyzing not only different steady states, but also the transition paths between steady states. As I show below, a narrow focus on steady state welfare may yield misleading policy conclusions.

The formal framework of my analysis is that of an economic epidemiology model. The existing literature on economic epidemiology falls into two broad categories. First, there is a small literature that considers different models of treatment, including Sanders (1971), Sethi (1974), Sethi and Staats (1978) Goldman and Lightwood (1995, 2002), Rowthorn (2006), Toxvaerd (2009a, 2009b) and Rowthorn and Toxvaerd (2015). Treatment turns out to be fundamentally different from prevention and thus the policies discussed in that literature are orthogonal to the policies studied in the present work.

Second, there is a relatively large and growing literature on the effects and desirability of preventive measures such as quarantines, prophylaxis, vaccines and promiscuity/abstinence. Main contributions include Sethi (1978), Geoffard and Philipson (1996), Kremer (1996), Auld (2003), Gersovitz and Hammer (2004), Aadland et al. (2010), Francis (1997, 2004), Boulier et al. (2007), Brito et al. (1991), Barrett (2003) and Reluga (2009, 2010). Chen and Toxvaerd (2014) surveys the large literature on the economics of vaccination. Although many of these papers study interventions that fall under the broad category of "preventive measures" that may seem superficially similar, there are very important differences between them in terms of predictions and policy implications. It is therefore of crucial importance for practical applications to understand the delicate relations between them. Below, I offer a detailed comparison of (instantaneous) moment-by-moment prevention, as studied in the present paper, and more permanent prevention such as vaccination. I also discuss models of abstinence, which differ from the present model by changing not only the rate of transmission but also the contact pattern itself. Last, I briefly discuss models of quarantines.

In Section 2, I set out the classical and economic models and briefly summarize the classical results. In Section 3, I analyze the problem faced by a benevolent central planner, the solution of which serves as a benchmark. In Section 4, I analyze the equilibrium under non-cooperative decentralized decision making. In Section 5, I characterize the nature of external effects. In Section 6, I discuss welfare, derive comparative statics results and draw policy conclusions. In Section 7, I briefly discuss the extension to settings with a heterogeneous population in which there is no recovery. Section 8 contains a detailed comparison of the present model with related models of vaccination, abstinence and quarantines. In Section 9, I offer concluding remarks. The Appendix contains details of proofs and of the numerical experiments.

2. The Classical and Economic Models

To make the exposition self-contained, I start by expounding the classical epidemiological version of the model in some detail. This will not only aid in understanding the economic

models that follow, but also highlight the contrast in predictions based on the different modeling approaches.

The classical susceptible-infected-susceptible model is simple to describe.[11] Time is continuous and runs indefinitely. A population $\mathcal{P} = [0,1]$ consists of a continuum of infinitely lived individuals who can at each instant $t \geq 0$ each be in one of two states, namely *susceptible* or *infected*. The set of infected individuals is denoted by $\mathcal{I}(t)$ and has measure $I(t)$, while the set of susceptible individuals is denoted by $\mathcal{S}(t)$ and has measure $S(t)$. Because the population size is normalized to unity, these measures can be interpreted as fractions. In what follows, $I(t)$ will be referred to as *disease prevalence*.

At each instant, the population mixes homogeneously. This corresponds to pair-wise random matching, where each individual has an equal chance of meeting any other individual, irrespective of the health status of the two matched individuals. While a match between two infected individuals or two susceptible individuals does not create any new infection, a match between an infected and a susceptible individual may. The rate at which infection is transferred in such a match, is denoted by $\beta > 0$. This parameter captures the infectivity of the disease. Coupled with the assumption of homogeneous mixing, this means that the aggregate rate at which susceptible individuals become infected is given by the simple expression $\beta I(t)S(t)$. Thus the rate of new infection, or *disease incidence*, is proportional to disease prevalence.[12] Note that while disease incidence is a flow value, disease prevalence is a stock value.

Infected individuals recover spontaneously at rate $\alpha \geq 0$. This means that the aggregate rate at which infected individuals become susceptible is given by $\alpha I(t)$. The dynamics of the model are thus described by the following system of differential equations:

$$\dot{S}(t) = I(t)\left[\alpha - \beta S(t)\right] \tag{1}$$

$$\dot{I}(t) = I(t)\left[\beta S(t) - \alpha\right] \tag{2}$$

$$I(t) = 1 - S(t), \quad I(0) = I_0 \tag{3}$$

Using the normalization, this system reduces to the following simple logistic growth equation:

$$\dot{I}(t) = I(t)\left[\beta(1 - I(t)) - \alpha\right], \quad I(0) = I_0 \tag{4}$$

The steady states of this system are given by

$$I^* = 0, \quad I^* = \frac{\beta - \alpha}{\beta} \tag{5}$$

For $\beta > \alpha$, the stable steady state is endemic (or persistent), while for $\beta < \alpha$, the stable steady state involves eradication. In other words, if the rate at which individuals become infected surpasses the rate at which they recover, then some positive fraction of the population will always be infected. If recovery is not possible, then the entire

[11] See Anderson and May (1991), Daley and Gani (2001) or Keeling and Rohani (2008) for good introductions and applications to specific diseases.

[12] The term $\beta I(t)S(t)$ should be thought of as the rate at which susceptible individuals have contact with other individuals, multiplied by the probability of the contact being with an infectious individual, multiplied by the probability that the infection is transmitted in such a contact. See e.g. Keeling and Rohani (2008) for a detailed derivation.

population ends up being infected. On the other hand, if individuals recover at a higher rate than the rate at which they become infected, then the disease eventually dies out. This completes the description of the classical SIS model.

Note the following straightforward but important comparative statics results, which I will later contrast with comparable results under optimizing behavior:

Result 0. The endemic steady state level of disease prevalence in the classical model is decreasing in the recovery rate α and increasing the infectivity rate β.

At the aggregate level, there is no uncertainty and thus the probability that an arbitrary individual is infected, must coincide with the fraction of infected individuals. From the perspective of an infected individual, the transition to susceptibility is governed by a Poisson process with rate α. Similarly, for a fixed level of aggregate infection $I(t)$, the transition to infectivity for a susceptible individual is governed by a Poisson process with rate $\beta I(t)$. Thus transition probabilities are memoryless, a fact that greatly simplifies the analysis that follows.

2.1. Extension to an Economic Model. To turn the classical model into a fully fledged economic model, I will proceed by assigning payoffs to the different disease states and assume that time is discounted. Specifically, I will assume that individuals earn a flow payoff $\pi_{\mathcal{S}} > 0$ per instant while susceptible, a flow payoff $\pi_{\mathcal{I}} < \pi_{\mathcal{S}}$ per instant while infected and that time is discounted at rate $\rho > 0$. For notational simplicity, let the quantity $\pi \equiv \pi_{\mathcal{S}} - \pi_{\mathcal{I}} > 0$ denote the health premium. The health premium should be thought of broadly as the benefits of not being infected, e.g. physical well-being, higher productivity etc.

To model the possibility of engaging in preventive behavior, assume that the individuals can affect the rate of infection by controlling the rate at which they expose themselves to infection. In particular, at each instant t, each individual $i \in \mathcal{S}(t)$ non-cooperatively chooses exposure level $\varepsilon_i(t) \in [0,1]$, at personal cost $(1-\varepsilon_i(t))c \geq 0$.[13] Effectively, this reduces the rate of infection to $\varepsilon_i(t)\beta I(t)$. This formalization captures the notion that, ceteris paribus, exposure is desirable. Equivalently, this means that engaging in preventive behavior is privately costly. Thus, in the absence of infection risk, it would be optimal to choose full exposure (i.e. to choose not to protect oneself). This completes the description of the economic version of the SIS model.

Throughout, the following assumptions are maintained:

Assumption 1

$$c < \pi\left(\frac{\beta}{\alpha+\beta+\rho}\right) \tag{6}$$

This assumption, which bounds the cost of prevention, ensures that prevention is potentially desirable from the perspective of both the social planner and the individuals.

Assumption 2

$$\alpha < \beta \tag{7}$$

[13] This formulation is equivalent to choosing protection level $\delta(t) \in [0,1]$ at cost $\delta(t)c$, which then yields the infection rate $(1-\delta(t))\beta I(t)$.

This assumption implies that the eradication steady state of the classical model is unstable and the relevant steady state is the endemic one. Note that if $\alpha > \beta$, then the infectious disease is eventually eradicated, even if no-one engages in any preventive behavior at all. The only possible role for preventive measures would then be to speed up the inevitable eradication of the disease.

3. Centralized Decision Making

In this section, I analyze the centralized problem. In this setup, a benevolent utilitarian social planner seeks to maximize the sum of the individuals' expected, discounted lifetime utilities through the direct control of the aggregate level of exposure $\varphi(t) \in [0,1]$. In solving this problem, the planner explicitly takes into account the fact that he influences aggregate disease incidence and prevalence, through the control of aggregate exposure. This will turn out to be a central difference between the scenarios with centralized and decentralized decision making, respectively.

Formally, the planner's problem is to solve the following programme:

$$\max_{\varphi(t)\in[0,1]} \int_0^\infty e^{-\rho t}[I(t)\pi_{\mathcal{I}} + S(t)\left[\pi_{\mathcal{S}} - (1-\varphi(t))c\right]]dt \tag{8}$$

$$s.t. \quad \dot{I}(t) = I(t)\left[\varphi(t)\beta S(t) - \alpha\right], \quad I(0) = I_0 \tag{9}$$

$$\dot{S}(t) = I(t)[\alpha - \varphi(t)\beta S(t)], \quad S(0) = 1 - I_0 \tag{10}$$

The constraints on the planner's problem are the laws of motion for the measures of infected and susceptible individuals, suitably modified to take into account that the rate at which infection occurs is a function of the centrally chosen aggregate exposure level.

Instead of solving this problem, the following simplified but equivalent programme will be considered (which differs only by the constant $\pi_{\mathcal{S}}$ in the objective and by using the normalization of the population size):

$$\max_{\varphi(t)\in[0,1]} \int_0^\infty e^{-\rho t}\left[-I(t)\pi - (1-I(t))(1-\varphi(t))c\right]dt \tag{11}$$

$$s.t. \quad \dot{I}(t) = I(t)\left[\varphi(t)\beta(1-I(t)) - \alpha\right], \quad I(0) = I_0 \tag{12}$$

An admissible policy is a pair of functions $(I(t), \varphi(t))$ in which for all $t \geq 0$, $I(t)$ satisfies the logistic growth equation and where $\varphi(t) \in [0,1]$. Furthermore, $\varphi(t)$ must be piecewise continuous.

To solve the problem, let $\lambda(t)$ denote the current-value multiplier on the constraint. The multiplier is required to be piecewise continuously differentiable. The current-value Hamiltonian is then given by

$$\begin{aligned} H^C \equiv & -I(t)\pi - (1-I(t))(1-\varphi(t))c \\ & +\lambda(t)I(t)\left[\varphi(t)\beta(1-I(t)) - \alpha\right] \end{aligned} \tag{13}$$

The evolution of the multiplier is then described by the differential equation

$$\dot{\lambda}(t) = \lambda(t)\left[\rho + \alpha + \varphi(t)\beta\left(2I(t)-1\right)\right] + \left[\pi - (1-\varphi(t))c\right] \tag{14}$$

Differentiating the current-value Hamiltonian with respect to $\varphi(t)$ (and supposing that

$I(t) < 1$), yields the following necessary Hamiltonian condition for optimality:

$$\lambda(t)\beta I(t) + c = 0 \tag{15}$$

The optimal policy for the planner is thus of the bang-bang type and given by

$$\varphi(t) = 0 \quad for \quad -\lambda(t)\beta I(t) > c \tag{16}$$
$$\varphi(t) \in [0,1] \quad for \quad -\lambda(t)\beta I(t) = c \tag{17}$$
$$\varphi(t) = 1 \quad for \quad -\lambda(t)\beta I(t) < c \tag{18}$$

This policy simply states that when the marginal cost of prevention is lower than the marginal benefit, then the planner fully prevents infection, i.e. he sets the exposure level to zero. Similarly, when the marginal benefit of prevention is lower than the marginal cost, then full exposure is optimal. Last, when the two coincide, then any policy yields the same welfare and hence is optimal. To see why the left-hand sides of these inequalities should be interpreted as the marginal benefits of prevention, recall that $\beta I(t)$ is the rate at which unprotected susceptible individuals become infected. But since $-\lambda(t)$ is the social value of a marginal decrease in disease prevalence, $-\lambda(t)\beta I(t)$ is the marginal benefit from a small decrease in exposure. Note that the relevant necessary transversality condition $\lim_{t\to\infty} e^{-\rho t} H^C(t) = 0$ is satisfied, when $\lambda(t)$ approaches a finite limit.

It is important to emphasize that while the optimal exposure is of the bang-bang type, there is nothing essential in the analysis that hinges on this feature. The important feature of the optimal policy to be noted, is that exposure is monotone decreasing in the risk faced by the decision maker. This feature generalizes straightforwardly to non-linear cost structures. To see this, consider the cost function $C(1-\varphi(t))$ with $C' > 0$ and $C'' \geq 0$ (of which linear costs are a special case). With this specification of costs, the dynamic equation for the evolution of the state variable is unchanged, but the Hamiltonian (or first-order) condition is now

$$\beta I(t)\lambda(t) + C'(1-\varphi(t)) = 0 \tag{19}$$

Since C' is increasing, it follows that the exposure level $\varphi(t)$ that satisfies the Hamiltonian condition is decreasing in the term $\lambda(t)\beta I(t)$, exactly as in the linear case.[14] It follows that when prevalence is above its steady state level, it is optimal to maintain an aggregate exposure level that is lower than the steady state value. Similarly, when prevalence is below its steady state level, it is optimal to maintain aggregate exposure at a level that is higher than its steady state value. It is thus clear that whether costs are linear or strictly convex, optimal exposure levels always bring infection towards the interior steady state level. This is the central result that drives the dynamics, not the linearity of costs.[15]

To characterize the singular steady state, I set $\dot{I}(t) = \dot{\lambda}(t) = 0$ and use the Hamiltonian

[14] Recall that $\beta I(t)\lambda(t) < 0$, so an increase in this term in absolute value calls for an increase in $C'(1-\varphi(t))$, that is for a decrease in exposure.

[15] The fact that linearity is not restrictive in this class of models is also found by Goldman and Lightwood (2002) in a model of treatment. The great advantage of the linearity assumption is that it allows me to derive closed form solutions for both the steady states and of exposure levels along the transition paths.

condition to derive the following steady state values:

$$\varphi_C^* \equiv \frac{\alpha(\pi - c)}{\beta\pi - (\beta + \rho)c} \tag{20}$$

$$I_C^* \equiv \frac{c\rho}{\beta(\pi - c)} \tag{21}$$

$$\lambda_C^* \equiv \frac{-(\pi - c)}{\rho} \tag{22}$$

This steady state is saddle path stable.[16]

As the following result shows, there is no boundary solution to the planner's problem:

Theorem 1. *The centralized problem has a unique steady state, which is interior.*

Proof: See Appendix A ■

This result has a straightforward intuition. When enough individuals are infected, the probability that an unprotected individual will become infected is so high, that the cost of prevention is outweighed by the expected welfare loss of becoming infected. Similarly, when only a few individuals are infected, the probability of becoming infected is too low to warrant engaging in costly preventive behavior. Thus the optimal policy always forces disease prevalence towards the interior of its domain.

Having characterized the unique steady state, all that remains is to characterize the optimal path towards the steady state. It takes a particularly simple form, as the next result shows:

Theorem 2. *An optimal path exists and is of the most rapid approach type. It is characterized by*

$$\varphi_C^* = 0 \quad for \quad I(t) > I_C^* \tag{23}$$

$$\varphi_C^* = \frac{\alpha(\pi - c)}{\beta\pi - (\beta + \rho)c} \quad for \quad I(t) = I_C^* \tag{24}$$

$$\varphi_C^* = 1 \quad for \quad I(t) < I_C^* \tag{25}$$

Proof: See Appendix B ■

This result follows as the necessary conditions for the optimality of a nearest approach path in Sethi (1977, Theorem 3.1) are met.[17]

Thus, whenever disease prevalence is above the steady state level, the planner optimally reduces exposure to zero until the steady state level is reached. At this point, the planner switches to partial exposure so as to maintain the steady state disease prevalence. Similarly, whenever disease prevalence is below the steady state level, the planner

[16] The eigenvalues of the Jacobian are given by $\frac{\rho(-\beta\pi + c(\rho+\beta-\alpha))}{-\beta\pi + c(\rho+\beta)}$ and $\frac{\alpha\rho c}{-\beta\pi + c(\rho+\beta)}$ respectively. Under Assumption 1, these are always of opposite sign.

[17] See also Hartl and Feichtinger (1987) for a more general result.

optimally chooses full exposure, until disease prevalence has increased to its steady state level, which is subsequently maintained through partial exposure.

It is worth emphasizing that in the present non-concave model, the Hamiltonian condition is necessary but not sufficient for optimality and that the Mangasarian and Arrow sufficiency conditions do not apply. One can therefore not rely on the standard results to ensure sufficiency, but must instead rely on alternatives, based on Green's Theorem. This is indeed the route I take and which is covered in Appendix B. I prove there that the optimal paths are indeed the ones I have characterized. The proof does not rely on second-order conditions (i.e. the relevant Legendre-type conditions), because they are inapplicable.

For emphasis, the optimality of the solution I have characterized can also be seen from the following alternative argument. It is known from Wagener (2003) that in the class of models to which this model belongs, limit sets are either equilibria or limit cycles. That limit cycles can be ruled out follows from Rowthorn and Toxvaerd (2015), who show the result in a more general infection model that encompasses the present protection-only model as a special case. This leaves equilibria as the only possible limit set. Rowthorn and Toxvaerd (2015) also show existence of an optimal policy, so the only remaining issue is to pin down the steady state. But as there is a unique steady state in this system, it follows that it is indeed the endpoint of any optimal path.[18]

4. Decentralized Decision Making

In this section, I analyze the decentralized problem. I will proceed by analyzing the problem of an individual and then aggregate across the entire population. From the perspective of an individual, the path of disease prevalence is exogenously given.[19] It will be assumed that each individual has perfect foresight, in the sense that conjectures about the aggregate evolution of the disease are confirmed in equilibrium.

Individual i's optimization problem is given by

$$\max_{\varepsilon_i(t)\in[0,1]} \int_0^\infty e^{-\rho t} v_i(t)^T \varsigma_i(t) dt \tag{26}$$

$$s.t. \quad \dot{\varsigma}_i(t) = Q_i(t)\varsigma_i(t) \tag{27}$$

where

$$v_i(t) = [\pi_{\mathcal{I}}, \pi_{\mathcal{S}} - (1-\varepsilon_i(t))c]^T \tag{28}$$

is the vector of state dependent flow utilities, the 2×1 vector $\varsigma_i(t)$ is a probability measure on the set of states and $Q_i(t)$ is the transition rate (or *intensity*) matrix, given by

$$Q_i(t) = \begin{pmatrix} -\varepsilon_i(t)\beta I(t) & \alpha \\ \varepsilon_i(t)\beta I(t) & -\alpha \end{pmatrix} \tag{29}$$

[18] In fact, that cycles cannot be present in this model also follows from the fact that in an autonomous, infinite horizon, one state/one control optimal control problem, it is known that the optimal policy is monotone in time. If there were cycles, they would contradict the non-monotonicity. See e.g. Hartl (1987) and Kamien and Schwartz (1991) for details.

[19] Because each agent is negligible and does not influence the aggregate, any feedback between an individual's action and other individuals' responses can be ignored by the individual. This effectively circumvents a major complication of analyzing continuous time games.

Note that the individual's transition rate matrix $Q_i(t)$, is a function of the strategies adopted by the individual and by the population as a whole. This formulation of the individual's problem is analogous to that in Reluga (2009) and Reluga and Galvani (2011). To further analyze the individual's problem, it is useful to rewrite it as a standard optimal control problem with a single state variable.[20] First, note that at time $t \geq 0$, individual i's health status is given by the indicator random variable

$$h_i(t) = \begin{cases} 1 & if \quad i \in \mathcal{I}(t) \\ 0 & if \quad i \in \mathcal{S}(t) \end{cases} \tag{30}$$

Thus the probability that individual i is infected at instant $t \geq 0$, is given by

$$q_i(t) = E[h_i(t)] \tag{31}$$

One may therefore write the vector of state probabilities simply as

$$p_i(t) = [q_i(t), 1 - q_i(t)]^T \tag{32}$$

The probability $q_i(t)$, which I will take as the state variable in individual i's control problem, evolves according to a non-homogeneous continuous-time Markov process[21]

$$\dot{q}_i(t) = (1 - q_i(t))\varepsilon_i(t)\beta I(t) - \alpha q_i(t) \tag{33}$$

The individual's problem can now be rewritten as the following standard optimal control problem:

$$\max_{\varepsilon_i(t)\in[0,1]} \int_0^\infty e^{-\rho t} \left[q_i(t)\pi_{\mathcal{I}} - (1 - q_i(t)) \left[\pi_{\mathcal{S}} - (1 - \varepsilon_i(t))c\right]\right] dt \tag{34}$$

$$s.t. \quad \dot{q}_i(t) = \varepsilon_i(t)\beta I(t)(1 - q_i(t)) - \alpha q_i(t), \quad q_i(0) \in \{0, 1\} \tag{35}$$

This problem is amenable to standard optimal control techniques. An individual's problem can be further simplified, to yield the following final formulation:[22]

$$\max_{\varepsilon_i(t)\in[0,1]} \int_0^\infty e^{-\rho t} \left[-q_i(t)\pi - (1 - q_i(t))(1 - \varepsilon_i(t))c\right] dt \tag{36}$$

$$s.t. \quad \dot{q}_i(t) = \varepsilon_i(t)\beta I(t)(1 - q_i(t)) - \alpha q_i(t), \quad q_i(0) \in \{0, 1\} \tag{37}$$

For an infected individual at instant $t \geq 0$, $q_i(t) = 1$ and thus there is no decision to be made. For a susceptible individual at instant $t \geq 0$, $q_i(t) = 0$ and hence the individual trades off costs and benefits of controlling the rate of transition from $\mathcal{S}(t)$ to $\mathcal{I}(t)$.

The above problem is solved for each individual on the background of the aggregate evolution of the infectious disease. This is in turn described by the following modified

[20] This can be done because there are only two health states and the state probabilities must sum to one at all times.

[21] It is non-homogeneous because infection prevalence $I(t)$ changes over time.

[22] This objective is a simplification similar to that used in the centralized setting.

logistic growth equation:

$$\dot{I}(t) = I(t)\left[\boldsymbol{\varepsilon}(t)\beta(1 - I(t)) - \alpha\right], \quad \boldsymbol{\varepsilon}(t) \equiv \int_{i\in\mathcal{S}(t)} S(t)^{-1}\varepsilon_i(t)di \tag{38}$$

In this equation, $\boldsymbol{\varepsilon}(t)$ denotes the aggregate level of exposure resulting from the susceptible individuals' disaggregate exposure levels. Let $\mu(t)$ be the current-value multiplier on the constraint in the individual's problem.[23] The current-value Hamiltonian for the individual's problem is then given by

$$\begin{aligned} H_i^D \equiv & -q_i(t)\pi - (1 - q_i(t))(1 - \varepsilon_i(t))c \\ & +\mu(t)\left[\beta I(t)(1 - q_i(t))\varepsilon_i(t) - \alpha q_i(t)\right] \end{aligned} \tag{39}$$

The evolution of the multiplier is in turn governed by the differential equation

$$\dot{\mu}(t) = \mu(t)\left[\rho + \alpha + \varepsilon_i(t)\beta I(t)\right] + \left[\pi - (1 - \varepsilon_i(t))c\right] \tag{40}$$

Differentiating the current-value Hamiltonian with respect to $\varepsilon_i(t)$ (and supposing that $q_i(t) < 1$), yields the following necessary Hamiltonian condition for (private) optimality:

$$\mu(t)\beta I(t) + c = 0 \tag{41}$$

The best response of an individual can thus be characterized as follows:

$$\varepsilon_i(t) = 0 \quad for \quad -\mu(t)\beta I(t) > c \tag{42}$$

$$\varepsilon_i(t) \in [0, 1] \quad for \quad -\mu(t)\beta I(t) = c \tag{43}$$

$$\varepsilon_i(t) = 1 \quad for \quad -\mu(t)\beta I(t) < c \tag{44}$$

Again, $-\mu(t)\beta I(t)$ is to be interpreted as the marginal benefit of prevention, while c is the corresponding marginal cost. In words, the privately optimal policy is to fully expose oneself to infection, if the marginal cost of doing so outweighs the marginal benefit and to fully protect oneself, if the marginal cost is outweighed by the marginal benefit. In case of indifference, any level of prevention will do.[24]

An individual's steady state conditions imply that

$$q_i^* = \frac{\beta I(t)\varepsilon_i(t)}{\beta I(t)\varepsilon_i(t) + \alpha} \tag{45}$$

$$\mu^* = \frac{-\pi + (1 - \varepsilon_i(t))c}{\alpha + \rho + \beta I(t)\varepsilon_i(t)} \tag{46}$$

Thus the individual's problem can only be in steady state if the aggregate system is too and vice versa. That is, $\dot{q}_i(t) = \dot{\mu}(t) = \dot{I}(t) = 0$ in steady state. But the individual's problem can reach a "stationary state" for any number of different (constant) levels of

[23] An admissible policy for an individual is a pair of functions $(q(t), \varepsilon_i(t))$ in which for all $t \geq 0$, $q(t)$ satisfies the differential equation constraint and $\varepsilon_i(t) \in [0, 1]$ and is piecewise continuous. Furthermore, the multiplier $\mu(t)$ is required to be piecewise continuously differentiable.

[24] For completeness, the transversality condition $\lim_{t\to\infty} e^{-\rho t} H_i^D(t) = 0$ holds, and the Arrow sufficiency conditions for an individual optimum are also satisfied.

disease prevalence $I(t)$.

To characterize the singular solution, note that $\boldsymbol{\varepsilon}(t) = \varepsilon_i(t)$ for all $i \in \mathcal{S}(t)$ in symmetric equilibrium. Setting $\dot{\mu}(t) = \dot{I}(t) = 0$ and using the Hamiltonian condition, then yields the steady state values:

$$\varepsilon_D^* \equiv \frac{\alpha(\pi - c)}{\beta\pi - c(\alpha + \beta + \rho)} \tag{47}$$

$$I_D^* \equiv \frac{c(\alpha + \rho)}{\beta(\pi - c)} \tag{48}$$

$$\mu_D^* \equiv \frac{-(\pi - c)}{\alpha + \rho} \tag{49}$$

At the steady state level of disease prevalence I_D^*, the individuals are willing to use mixed strategies. Furthermore, the strategies ε_D^*, if followed by all, yields the steady state disease prevalence I_D^*. It should be noted that the equilibrium is saddle-path stable.[25] The singular (aggregate) solution just derived, fully characterizes the set of steady states under decentralized decision making. That this is the case follows from the next result, which rules out any further steady states:

Theorem 3. *The decentralized problem has a unique steady state, which is interior.*

Proof: See Appendix A ■

To complete the description of the decentralized equilibrium, the equilibrium path towards the steady state must be characterized. As in the centralized setup, equilibrium paths take a very simple form, as shown next:

Theorem 4. *In the unique decentralized equilibrium, for all times $t \geq 0$ and individuals $i \in \mathcal{S}(t)$, equilibrium strategies are given by*

$$\varepsilon_i^* = 0 \quad for \quad I(t) > I_D^* \tag{50}$$

$$\varepsilon_i^* = \frac{\alpha(\pi - c)}{\beta\pi - c(\alpha + \beta + \rho)} \quad for \quad I(t) = I_D^* \tag{51}$$

$$\varepsilon_i^* = 1 \quad for \quad I(t) < I_D^* \tag{52}$$

Proof: See Appendix B ■

Akin to the optimal policy in the centralized setup, in the decentralized equilibrium individuals play pure strategies outside of steady state in order to approach the singular solution as rapidly as possible. Once there, the individuals switch to mixed strategies to maintain the steady state level of disease prevalence. The proof follows similar steps as those in the centralized setup.

For reasons that parallel those in the centralized setup, the steady state is necessarily interior under decentralization. But because individuals do not internalize the effects

[25] The eigenvalues of the Jacobian are given by $\alpha+\rho+\beta I(t)\varepsilon(t) > 0$ and $-\alpha-\beta I(t)\varepsilon(t) < 0$ respectively.

that their preventive behavior has on other individuals and because each individual has a negligible effect on the aggregate, the centralized and aggregate decentralized steady state policies and disease prevalences differ. This finding will be explored further in the next section.

5. THE NATURE OF EXTERNALITIES

Many of the insights and recommendations in this paper stem from the fact that individual decision makers have different incentives and objectives than those of the social planner. Received wisdom has it that because individuals do not properly internalize the negative externalities that their decisions have on third parties, they effectively overexpose themselves to infection, thereby driving a wedge between equilibrium outcomes and those that are socially optimal. It turns out that externalities are only part of the reason why individuals make socially undesirable decisions. In what follows, I will build on results in Rowthorn and Toxvaerd (2015) and decompose the external effects into (i) a *pure externality* effect and (ii) a *smallness* effect. This decomposition shows that the discrepancy in incentives is not merely a result of uninternalized externalities and that received wisdom may therefore be overly simplistic.

5.1. A Typology of External Effects. As noted above, under decentralized decision making, the incentives of the individual are different from those of the central planner. As a result, they assign different shadow prices to infection. On an optimal path, the social planner's shadow price is state dependent and can be denoted by $\lambda(I)$. At the same level of aggregate infection, the individual's shadow price for its own infection can be denoted by $\mu(I)$. The gap between these shadow prices is $\lambda(I) - \mu(I) \leq 0$.

One can usefully think of this gap as follows: First, there is a pure externality effect, which stems from the fact that in deciding which exposure level to choose, each individual ignores the impact of its actions on the well-being of other individuals. Second, there is a smallness effect, which stems from the fact that the individual cannot affect the aggregate disease dynamics. Starting from any given level of aggregate infection I, the future time profile $I(t)$ is different under centralized and decentralized decision making. As a result, the individual faces a different future time profile of infection risk under the two scenarios. The smallness effect encapsulates the influence of this difference on the individual's shadow price.

To quantify these two effects, consider a so-called "maverick" individual. Suppose that under central direction, all individuals, except for one numerically insignificant maverick, behave in the socially optimal fashion. This ensures that aggregate infection $I(t)$ will exactly follow the socially optimal path. The maverick, in contrast, behaves in a purely selfish fashion and maximizes its own personal discounted expected utility, while facing the socially optimal time profile of future infection risk. The privately optimal solution for the maverick will then satisfy the system of equations and conditions that characterize decentralized dynamics, but on the assumption that aggregate infection $I(t)$ follows the socially optimal trajectory. Let $\eta(t)$ denote the shadow price of infection for the maverick. This can be written in the state dependent form $\eta(I)$.

Next, consider the difference between the optimization problem of the maverick and that of the central planner. By assumption, both the planner and the maverick face the same path of aggregate infection (namely the socially optimal one) and hence any differences in valuation stem from the fact that the maverick ignores the impact of its

actions on the well-being of others. Thus the difference is a pure externality effect. The maverick disregards the externality effect and as such, experiences a lower return on investment from prevention than does the planner. For this reason, the maverick invests less in such measures, even though it faces the same aggregate path of infection as the planner does. Next, consider the difference between the optimization problem of the maverick and that of an individual on the decentralized equilibrium path of aggregate infection. Both individuals face the same costs of prevention and both are indifferent to the harm their actions may impose on others, but they face different paths of future aggregate infection, namely the socially optimal one versus the decentralized equilibrium path. This difference in valuation captures the added risk that comes about because the individuals concerned face different future paths of aggregate infection.

More formally, compare the following three paths, expressing the costate variables as functions of the current aggregate infection rate I: (i) the decentralized equilibrium path facing an individual (the shadow price of an individual is then equal to $\mu(I)$); (ii) the centralized optimal path (along which the shadow price is equal to $\lambda(I)$); (iii) the centralized optimal path facing the maverick (along which the shadow price is equal to $\eta(I)$). One then has the following important result:

Theorem 5. *The shadow price gap can be decomposed into a pure externality effect* $[\lambda(I) - \eta(I)] \leq 0$ *and a smallness effect* $[\eta(I) - \mu(I)] \leq 0$*, such that*

$$\lambda(I) - \mu(I) = [\lambda(I) - \eta(I)] + [\eta(I) - \mu(I)] \tag{53}$$

where

$$\dot{\lambda}(t) = \lambda(t)\left[\rho + \alpha + (1 - (1 - \varphi(t)))\beta(2I(t) - 1)\right] + [\pi - (1 - \varphi(t))c] \tag{54}$$

$$\dot{\mu}(t) = \mu(t)\left[\rho + \alpha + (1 - (1 - \varphi(t)))\beta I(t)\right] + [\pi - (1 - \varphi(t))c] \tag{55}$$

$$\dot{\eta}(t) = \eta(t)\left[\rho + \alpha + (1 - (1 - \varphi(t)))\beta I(t)\right] + [\pi - (1 - \varphi(t))c] \tag{56}$$

Note that in the equations for $\lambda(t)$ and $\eta(t)$, aggregate infection $I(t)$ follows the socially optimal path. In the equation for $\mu(t)$, aggregate infection follows a decentralized equilibrium trajectory. The values of the instrument $\varphi(t)$ differ across equations.

Proof: See Appendix E ■

It is worth noting that $\eta(1) = \lambda(1)$, i.e. the shadow values coincide when the entire population is infected. This follows from the straightforward observation that there are no possible externalities when the entire population is already infected.[26]

As can be appreciated by inspecting the laws of motion, the magnitudes of the two effects are state dependent and thus vary across the stages of the epidemic. This turns out to be important from the perspective of second-best policy interventions, because such interventions shift the system from one equilibrium steady state to another, each having different magnitudes of these effects. Thus the overall impact of the intervention on social welfare is in general difficult to analyze analytically.

[26] In a similar fashion, note that for the special case $\alpha = 0$ and $I(t) = 1$, the individuals and the planner have the same shadow prices, i.e. $\mu(t) = \lambda(t)$. This case is relevant when prevention is too costly to be used in steady state and when spontaneous recovery is not possible. In this case, in steady state the entire population is infected and thus the planner and the individuals face the same path of infection. Since there are no susceptibles on which an infected individual can have external effects, there are no externalities. Therefore, the individuals and the planner value infection equally much.

5.2. Strategic External Effects. Last, I will briefly discuss the issue of strategic externalities. As shown above, the only thing that determines an individual's equilibrium exposure at a given point in time, is the risk of infection associated with exposure at that moment. Other individuals' past, present or future behavior is entirely irrelevant. The future plays no role because individuals reoptimize at each instant and have no ability to commit to future preventive behavior.[27] Next, note that past behavior of other individuals only feeds through to the present decisions via their effect on the present level of disease prevalence. The exact history is immaterial and all that matters is the present risk of infection. Last, present behavior by other individuals is irrelevant, because from the perspective of an individual, all that matters is the fraction of the population that is infected, not how the fraction of susceptible individuals is distributed between protected and non-protected individuals. This is because at a given point in time, no susceptible individual can transmit the disease, whatever his or her level of exposure.[28] Of course, other individuals' past and present decisions do influence a given individual's *future* choices.

In contrast to the lack of externalities on the transition path, in steady state the game is one of strategic substitutes, in the sense that the more others protect themselves against infection, the lower is the incentive of an individual to protect itself. As a consequence, the players necessarily use mixed strategies in steady state, i.e. they randomize between full prevention and full exposure.

To see this, consider an individual at time $t \geq 0$. From the Hamiltonian condition of the individual, one can determine how the marginal benefit of prevention, i.e. the quantity

$$-\beta I(t)\mu(t) \tag{57}$$

varies with aggregate exposure $\boldsymbol{\varepsilon}(t)$, where

$$I(t) = \frac{\beta\boldsymbol{\varepsilon}(t) - \alpha}{\beta\boldsymbol{\varepsilon}(t)}, \quad \mu(t) = \frac{-\pi + (1 - \varepsilon_i(t))c}{\alpha + \rho + \beta I(t)\varepsilon_i(t)} \tag{58}$$

These are the relevant state and costate values in a symmetric steady state equilibrium. Direct inspection of the derivative of the marginal benefit from prevention with respect to $\boldsymbol{\varepsilon}(t)$, shows that the sign is positive under assumption (6). In other words, the more others expose themselves, the less inclined will the remaining individual be to expose him or herself.

6. WELFARE, POLICY AND RATIONAL DISINHIBITION

In this section, I call on the results of the previous sections to draw implications about welfare and public policy towards preventive measures. The relative simplicity of the model makes it straightforward to derive comparative statics results, and to conduct welfare analysis and arrive at some interesting and surprising conclusions. I will start by characterizing and comparing the optimal solution with the equilibrium outcome and then derive some important comparative statics results. Next, I will use these to consider three specific policy measures, namely (i) medical interventions and innovations that permanently alter the infectiousness of the disease, (ii) medical interventions/innovations

[27] This is in sharp contrast to vaccination, which will be discussed in further detail below.

[28] Similar points are made by Francis (1997), in his analysis of the work by Brito et al. (1991).

that permanently alter the rate of recovery from infection (iii) the design of tax and subsidy schemes that align private and public incentives.

The first important result follows directly from the steady state level of disease prevalence of the planner's problem:

Result I. For any recovery rate $\alpha \in [0, \beta)$, eradication through continual prevention is suboptimal.

This result should not be taken as a statement about the undesirability of eradicating infectious diseases, but rather as a statement about the suitability of this particular means of obtaining it. If eradication is found to be desirable, then alternative means such as vaccination and/or treatment may be more suitable for achieving this goal (if feasible).[29]

Next, I turn attention to a comparison of the centralized outcome and the decentralized equilibrium. It follows from inspection that for $\alpha > 0$, it is the case that $\varepsilon_D^* = \boldsymbol{\varepsilon}^* > \varphi_C^*$, so the central planner chooses lower exposure than do individuals in the decentralized equilibrium. As a consequence, $I_D^* > I_C^*$ so steady state disease prevalence is lower under centralized decision making than under decentralized decision making. These observations follow from the fact that the steady state multipliers are ranked so that $\lambda_C^* < \mu_D^*$ for $\alpha > 0$. This means that the shadow cost of infection is smaller from the perspective of the individual than it is from that of the population as a whole.

Turning to the exposure levels along the equilibrium path, note that for all levels of disease prevalence $I(t) \notin [I_C^*, I_D^*]$, equilibrium exposure levels are socially optimal and hence active public intervention is unnecessary. This is because for prevalence levels outside of this interval, decentralized equilibrium levels are already moving prevalence in the right direction (from a social perspective) as fast as possible.

For the special case $\alpha = 0$, i.e. when there is no possible recovery from the disease, inspection of the different steady state values shows that $I_D^* = I_C^*$, $\varepsilon_D^* = \boldsymbol{\varepsilon}^* = \varphi_C^*$ and $\lambda_C^* = \mu_D^*$. In other words, the outcomes under centralized and decentralized decision making coincide.[30]

The reason that the decentralized equilibrium outcome is socially optimal when $\alpha = 0$, can be understood in terms of the externalities that flow from individual's decentralized decisions. Ceteris paribus, even when recovery is not possible, susceptible individuals' decisions have external effects on other susceptible individuals. But in a large homogeneous population like the one considered here, the only possible equilibria are symmetric ones. In particular, this means that all individuals decide to switch from no prevention to full prevention at exactly the same critical level of disease prevalence. As a consequence, when one individual decides to start preventing infection, so does everyone else and there are therefore no positive external effects on them. To sum up, only individuals that are exposed to infection, can benefit from others' preventive efforts. But in a symmetric equilibrium, everyone starts protecting themselves at the same time, thereby ruling out external effects.[31] It should be emphasized that this result is not robust to

[29] Rowthorn and Toxvaerd (2015) show that a policy that involves eradication, does not satisfy the relevant transversality condition. Furthermore, they show that with sufficiently ineffective prevention, there may be steady states in which full prevention is optimal. But under such scenarios, the disease is never eradicated.

[30] The outcome of this special case of the model thus mirrors that obtained by Geoffard and Philipson (1996).

[31] Chen and Toxvaerd (2014) show, in the context of vaccination models, that inefficiencies stem from

population heterogeneity. This is because heterogeneity induces asymmetric prevention patterns, which in turn means that there is strategic interaction on the transition path. Such interaction creates non-internalized externalities from the individuals' prevention decisions. This is discussed further in Section 7.

The above results are summarized as follows:

Result II. If $\alpha > 0$, then in the decentralized equilibrium, both steady state exposure and disease prevalence are strictly higher than in the centralized solution. If $\alpha = 0$, then the decentralized equilibrium coincides with the solution of the centralized problem.

Turning to welfare, in the centralized setting, the steady state flow welfare is given by

$$W_C \equiv \frac{-c(\beta + \rho - \alpha)}{\beta} \tag{59}$$

while in the decentralized equilibrium, it is given by

$$W_D \equiv \frac{-c(\beta + \rho)}{\beta} = W_C - \frac{\alpha c}{\beta} \tag{60}$$

The former is obtained by evaluating the planner's current-value Hamiltonian at the centralized solution, while the latter is obtained by evaluating the planner's current-value Hamiltonian at the aggregate decentralized exposure level and the corresponding value for disease prevalence. These values lead to the following straightforward but important result:

Result III. If $\alpha > 0$, then in the decentralized equilibrium, steady state flow utility is lower than in the solution to the centralized problem, while if $\alpha = 0$, then they coincide.

It follows immediately from inspection of W_C and W_D that the steady state welfare loss due to decentralized decision making, is increasing in the recovery rate and decreasing in the infectivity of the disease. Interestingly, neither W_C nor W_D depend on the magnitude of the health premium π (although the first-best optimal and the decentralized equilibrium exposure levels do). Last, steady state welfare is decreasing in the prevention cost c in both settings.

Next, I turn to the comparative statics of the model, which I will then use for an analysis of public policy. I start by stating comparative statics results that have no counterpart in the classical analysis, namely for the "economic" parameters c, π and ρ:

Result IV. In both the centralized and decentralized settings, steady state exposure and

either ex ante heterogeneity or from strategic interactions. Their insights apply in a straightforward manner to the present setting. Recall that the population is homogeneous. As I have argued, there is never strategic interaction on transition paths, whatever the magnitude of the recovery rate α. This is in contrast to vaccination models, in which individually optimal policies are typically not myopic. Consider the case where $\alpha = 0$. In this case, all individuals fully protect themselves in steady state and since this protection offers perfect immunity from infection, no one individual is affected by the prevention decisions of other individuals. In this sense, when $\alpha = 0$, there is never strategic interaction on the equilibrium path (in or out of steady state) and thus the equilibrium under decentralized decision making is socially optimal). Next, consider the case where $\alpha > 0$. As before, on the transition path there is no strategic interaction, but once steady state is reached, agents switch to a mixed strategy in which the individually optimal level of prevention depends on other individuals' choices through aggregate disease prevalence. Thus there are eventually strategic interactions, which render equilibrium outcomes socially suboptimal.

	Steady State Exposure		**Steady State Prevalence**		**Steady State Welfare**	
	Centralized	Decentralized	Centralized	Decentralized	Centralized	Decentralized
	$\frac{\alpha(\pi-c)}{\beta\pi-(\beta+\rho)c}$	$\frac{\alpha(\pi-c)}{\beta\pi-(\alpha+\beta+\rho)c}$	$\frac{c\rho}{\beta(\pi-c)}$	$\frac{c(\alpha+\rho)}{\beta(\pi-c)}$	$\frac{-c(\beta+\rho-\alpha)}{\beta}$	$\frac{-c(\beta+\rho)}{\beta}$
α	$+$	$+$	0	$+$	$+$	0
β	$-$	$-$	$-$	$-$	$+$ for $\rho > \alpha$	$+$
π	$-$	$-$	$-$	$-$	0	0
c	$+$	$+$	$+$	$+$	$-$	$-$
ρ	$+$	$+$	$+$	$+$	$-$	$-$

Table 1: Comparative statics results.

disease prevalence are increasing in the prevention cost c and in the discount rate ρ and decreasing in the health premium π. Furthermore, steady state welfare is decreasing in c and ρ and is independent of π.

I next turn to the basic epidemiological parameters α and β. Recall for comparison that in the classical model, increased infectivity β or decreased recovery rate α, lead to higher (endemic) steady state disease prevalence. The comparative statics results for infectivity and the recovery rate in the economic model are as follows:

Result V. In the centralized setting, steady state exposure and disease prevalence are decreasing in infectivity β. Furthermore, steady state exposure is increasing in α, while steady state prevalence is independent of the recovery rate α.

Interestingly, as α is increased, the optimal centralized policy is modified exactly to keep prevalence unchanged. Next, I turn to the decentralized setting.

Result VI. In the decentralized setting, steady state equilibrium exposure and disease prevalence are decreasing in infectivity β and increasing in the recovery rate α.

The interpretation of these results is straightforward. Increasing the infectivity or decreasing the recovery rate, effectively makes exposure more unappealing, which is in turn reflected in decreased steady state exposure and disease prevalence. Before discussing the significance of these results, I will outline the effects of such changes on steady state flow welfare.

Result VII. In the decentralized setting, steady state flow welfare W_D is increasing in β and constant in α while in the centralized setting, steady state flow welfare W_C is increasing in α and increasing in β when $\rho > \alpha$.

For ease of reference and comparison, the comparative statics results for the centralized and decentralized settings are summarized in Table 1. In the following subsections, I will consider different policies aimed at managing the infectious disease, by directly influencing the tradeoffs faced by the individuals in the population.

6.1. Interventions that Influence Infectiousness. In this section, I will offer a detailed analysis of the effects and desirability of permanently changing the infectiousness of the disease. Before doing so, there are a number of common (and somewhat overlapping) misconceptions that must be addressed.

The first mistake, which is now recognized by some in both the policy and the academic communities, is the failure to explicitly account for behavioral responses to changes

in parameters characterizing the environment in which decisions are made. Recall that in the classical model, steady state disease prevalence is increasing in infectivity β. Without further analysis, this finding may lead to the premature conclusion that a worthy policy aim is to seek to decrease β, with a view to decrease disease incidence and prevalence. A possible rationale would be as follows. Given a fixed number of unprotected contacts, reducing infectivity β must decrease the number of new infections (i.e. incidence). Furthermore, since infection is undesirable, decreasing β must increase overall welfare. As the present analysis shows, this view is at best misleading and may prompt policy makers to introduce policies that do more harm than good. Specifically, such policies could be misguided on two separate counts. First, it is not necessarily the case that a benevolent social planner would always wish to reduce steady state prevalence, as evidenced by the fact that the steady state level of infection under central planning is interior. Second, and more importantly, the comparative statics of the classical model, on which such policies are based, are inadequate as they take individuals' behavior as given and fixed. As is clear from the present analysis, any decision maker that trades off costs and benefits of exposure, will respond to changes in infectivity β. Thus the mechanistic disease accounting inherent in the classical approach, is inappropriate for policy analysis. In short, while decreasing infectivity β indeed reduces overall disease incidence, *given* exposure levels, these in turn increase in response to the decrease in infectivity and thus cannot be taken as given.[32]

The second mistake, which is still ubiquitous in policy debates on risk compensation, is to confound statements about the presence of behavioral disinhibition with statements about the desirability of interventions that reduce risk. That is, it is common to see judgements about social welfare based solely on the presence (or lack of) disinhibition.[33] As shown in the present analysis, such a conclusion is wholly unwarranted, simply because both the central planner and the individuals respond to decreased infectivity by engaging in more risky behavior (i.e. to increase the exposure to infection). Thus the presence (or lack of) disinhibition is *not* an appropriate yardstick with which to measure changes in social welfare.

Rather than lumping the response of the social planner and that of the individuals into the common category "behavioral disinhibition", it is useful to distinguish between *optimal disinhibition* (exhibited by the central planner) and *rational disinhibition* (exhibited by individuals), the latter being rational in the sense that it is consistent with individual discounted expected utility maximization. With these concepts in place, it follows that behavioral disinhibition should in general be expected both theoretically and empirically (it is a simple consequence of the law of demand), but that it need not per se be a cause for concern.[34] As will be shown shortly, optimal disinhibition can never

[32] This type of effect is also found in Kremer (1996) in a model of abstinence (but without prevention), but he does not formally consider the welfare consequences of such changes in behavior.

[33] See e.g. Abbas et al. (2007), Over (2008), Over et al. (2004) and Szekeres et al. (2004).

[34] Risk compensation is documented by Cohen et al. (2009), Paltiel et al. (2009) and Crepaz et al. (2004). See also discussions in Philipson and Posner (1993), Szekeres et al. (2004), Eaton and Kalichman (2007) and Blower et al. (2000). There are some studies which fail to find evidence of risk compensation, in particular in the context of male circumcision (see e.g. Wilson et al., 2012 and references therein). The problem in interpreting such results, is that many interventions that seek to reduce the transmission rate of infectious diseases are bundled with other simultaneous interventions, such as "risk reduction counselling", as recommended by Martin et al. (2004) and Cassell et al. (2006). It is therefore difficult

be detrimental to welfare, while rational disinhibition might, depending on parameter values.

The third common mistake, is the failure to identify the source of potential gains from permanently reducing infectivity. Unless exposure decisions can be dictated by the central planner, the welfare gains flowing from a permanent decrease in infectivity stem from the benefits of increased exposure, rather than from permanently lower steady state disease prevalence. To see that this is indeed the case, one need only recall that permanently decreasing β unambiguously leads to rational disinhibition, which in turn lowers steady state welfare. In the policy debate surrounding risk compensation, it is commonplace to confound increased welfare with decreases in disease prevalence. As the present work shows, this is misleading, if one accepts the premise that the social planner (and the individuals) trade off both costs and benefits of risky behavior. In conclusion, reducing infectivity β may increase welfare because it allows individuals to enjoy the benefits of engaging in more risky behavior, not because such an intervention actually decreases disease prevalence.

In what follows, I will proceed to make the above observations more formally. Concretely, consider the following thought experiment. As a benchmark, take a situation in which the infectivity parameter is given by value β_1, which leads to some associated steady state prevalence level I_1 under decentralized decision making (i.e. $I_1 = I_D^*|_{\beta=\beta_1}$). Assume that the system has reached this steady state. Next, consider an exogenous, costless and permanent reduction of infectiousness to some lower rate $\beta_2 < \beta_1$ and denote by I_2 the associated steady state prevalence level (i.e. $I_2 = I_D^*|_{\beta=\beta_2}$).

To evaluate the welfare consequences of reducing infectivity β in this way, I compare the discounted social value of staying indefinitely at steady state I_1 under infectivity β_1, with the discounted social value of the path that takes the system from the old steady state I_1 to the new steady state I_2 and stays there indefinitely, the latter path calculated under infectivity β_2. As shown above, $I_2 > I_1$ and the associated steady state welfare levels are such that $W_D|_{\beta=\beta_2} < W_D|_{\beta=\beta_1}$. In words, under decentralized decision making, a reduction in infectivity increases both equilibrium steady state exposure and disease prevalence, leading to a decrease in steady state social welfare. If the discounted social welfare associated with the former path is higher than the discounted social welfare of the latter path, then reducing infectivity is not a welfare enhancing policy.

Formally, let $V(I, I'; \beta)$ denote the discounted social value of following the most rapid approach path that starts at I and ends (and stays perpetually) at I', under infectivity β. With this notation, there is immiserization of the population if for some values (β_1, β_2) with $\beta_2 < \beta_1$, it is the case that

$$V(I_1, I_2; \beta_2) < V(I_1, I_1; \beta_1) \tag{61}$$

where the latter term is simply the discounted social welfare of staying indefinitely at steady state I_1.[35]

to pinpoint the pure effect of reducing infectivity and to disentangle it from the effects of education to change risky behavior directly.

[35] Gersovitz (2010) states that "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 [..] that the rate of infection fall in the long-run steady state". As seen above, this condition is *never* satisfied in the present model and thus has no bite, i.e. immiserization cannot be ruled out.

Turning to the social planner, let $I_1^* = I_C^*|_{\beta=\beta_1}$ and $I_2^* = I_C^*|_{\beta=\beta_2}$ denote the end points of the optimal paths under centralized decision making. Because the planner can directly control the path of aggregate disease prevalence, he can ensure himself *at least* the value $V(I_1^*, I_1^*; \beta_2)$, i.e. the value of staying forever at the original steady state, but calculated using the lower infectivity rate β_2. By reoptimizing over possible paths, he will typically do even better. In other words, the following result holds:

Theorem 6. *Under centralized decision making,*

$$V(I_1^*, I_2^*; \beta_2) \geq V(I_1^*, I_1^*; \beta_2) > V(I_1^*, I_1^*; \beta_1) \tag{62}$$

In other words, under central planning, it follows from revealed preferences that reducing the infectiousness of the disease cannot decrease social welfare.[36]

In summary, a planner that decreases infectivity permanently and optimally steers disease prevalence from some steady state I_1^* to I_2^*, necessarily increases overall discounted social welfare. This is in contrast to the situation under decentralized decision making, where the reduction in infectivity causes a shift from steady state I_1 to I_2. From a social perspective, neither steady state I_1 nor I_2 is the end point of an optimal path. Thus the reduction in infectivity involves moving from one socially suboptimal steady state to another socially suboptimal steady state. There is therefore no reason why one should in general expect such a transition to be welfare enhancing.

The reason that the welfare implications are less clear cut in the decentralized setting, is that there is no revealed preference argument equivalent to that of the planner. Controlling for the behavior of other individuals, an individual cannot be harmed by lower levels of infectivity, for reasons mirroring those pertaining to the planner's response. However, in equilibrium, all individuals increase their exposure in response to decreased infectivity, thereby leading to suboptimally high levels of disease incidence and prevalence. It is therefore not clear in general what the net effect on social welfare of infectivity reducing measures is under decentralized decision making.

Numerical Experiments. Unfortunately, it is not possible to make further progress analytically and therefore I will report on several numerical experiments.[37] These confirm that depending on parameter values, reducing infectivity may indeed reduce overall welfare under decentralized decision making. For specific values of the parameters (α, ρ, c, π), I calculate the discounted social welfare under the scenarios described above, for a range of infectivity parameters β_1 and β_2. That is, I calculate the discounted social value of staying at the original steady state I_1 under infectivity β_1 and compare it to the discounted social value of the path that starts at the original steady state I_1 and moves to the new steady state I_2 (and stays there), calculated using infectivity β_2. Last, I repeat the experiment but vary each of the parameters (α, ρ, c, π) in turn, in order to check that the results are robust. Other than the discount rate ρ, the parameter values of the model are likely to differ significantly from disease to disease. Uncovering these from data is an ambitious and worthwhile exercise, but beyond the scope of the present work. For that reason, I select parameter values simply in order to illustrate the workings of the

[36] This result also follows from a straightforward application of the dynamic envelope theorem.

[37] In carrying out the experiments, a fourth-order Runge-Kutta procedure was used to approximate the solutions of the relevant ordinary differential equations. For further details, see Appendix D.

model. I have normalized the rate of recovery to unity and chosen prevention costs that are considerably smaller than the health premium. Last, I have set a rate of time preference corresponding to a discount factor of 0.9, in line with estimates in the empirical literature.

Wefare Gains from Reducing Infectiousness under Decentralized Decision Making

β_1 / β_2	4.0	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0
4.0	0.00	-	-	-	-	-	-	-	-	-	-	-	-
4.5	-0.48	0.00	-	-	-	-	-	-	-	-	-	-	-
5.0	0.43	0.14	0.00	-	-	-	-	-	-	-	-	-	-
5.5	1.04	-0.04	-0.20	0.00	-	-	-	-	-	-	-	-	-
6.0	1.98	0.89	-0.06	-0.69	0.00	-	-	-	-	-	-	-	-
6.5	3.17	1.30	0.33	-0.30	-0.44	0.00	-	-	-	-	-	-	-
7.0	3.82	1.91	0.92	0.29	0.15	-0.25	0.00	-	-	-	-	-	-
7.5	5.35	2.67	1.67	0.23	0.07	-0.34	-0.09	0.00	-	-	-	-	-
8.0	6.26	3.55	1.75	1.10	0.12	-0.30	-0.05	-0.83	0.00	-	-	-	-
8.5	7.27	4.53	2.73	1.26	0.27	-0.15	0.09	-0.69	-0.74	0.00	-	-	-
9.0	8.35	4.83	2.99	1.51	1.34	0.08	0.32	-0.46	-0.50	-0.65	0.00	-	-
9.5	8.77	5.96	4.12	2.64	1.65	0.39	0.62	-0.16	-0.20	-0.35	-0.58	0.00	-
10.0	9.97	6.38	4.51	3.02	2.02	0.75	0.13	0.21	0.16	0.02	-0.22	-0.53	0.00

Wefare Gains from Reducing Infectiousness under Centralized Decision Making

β_1 / β_2	4.0	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0
4.0	0.00	-	-	-	-	-	-	-	-	-	-	-	-
4.5	23.81	0.00	-	-	-	-	-	-	-	-	-	-	-
5.0	44.01	19.45	0.00	-	-	-	-	-	-	-	-	-	-
5.5	60.81	35.22	14.15	0.00	-	-	-	-	-	-	-	-	-
6.0	75.18	49.67	28.54	11.94	0.00	-	-	-	-	-	-	-	-
6.5	87.84	61.16	40.40	21.84	10.09	0.00	-	-	-	-	-	-	-
7.0	97.97	71.58	50.86	33.93	20.55	8.64	0.00	-	-	-	-	-	-
7.5	107.74	80.82	60.34	42.71	26.02	16.11	7.48	0.00	-	-	-	-	-
8.0	116.23	88.90	67.81	50.68	36.23	22.63	14.00	6.67	0.00	-	-	-	-
8.5	123.60	96.65	74.96	57.23	43.28	31.35	19.75	12.42	5.88	0.00	-	-	-
9.0	130.64	102.66	81.05	63.82	50.12	37.43	24.84	17.52	10.99	5.23	0.00	-	-
9.5	135.67	108.56	87.65	70.06	55.35	43.00	33.36	22.07	15.54	9.79	4.68	0.00	-
10.0	141.20	113.88	92.33	74.76	60.34	48.01	37.58	26.15	19.63	13.89	8.78	4.21	0.00

Table 2: Welfare gains from permanently decreasing infectiousness. Rows correspond to initial levels of infectiousness β_1, while columns correspond to reduced levels β_2. Numbers reported are scaled by a factor 10^3.

Table 2 illustrates the results of the experiment for $\alpha = 1$, $\rho = 0.11$, $c = 0.1$, $\pi = 10$ and $\beta_1, \beta_2 \in [4, 10]$, with increments of 0.5. Each row corresponds to a different initial infectivity parameter β_1, while each column corresponds to a different reduced infectivity parameter β_2. Since I only consider the effects of reducing the infectiousness of the disease, only entries below the diagonal are relevant and thus reported. For a pair of infectivity parameters (β_1, β_2), the entry in the table reports the magnitude of $\Delta(I_1, I_2; \beta_1, \beta_2) \equiv V(I_1, I_2; \beta_2) - V(I_1, I_1; \beta_1)$, i.e. of the net gain in discounted social welfare from decreasing infectiousness. The upper panel in the table shows the effects on welfare under decentralized decision making, while the lower panel shows the effects under centralized decision making.

Interpreting how the scope and severity of immiserization varies with the parameters of the model is a delicate matter, because the quantity $\Delta(I_1, I_2; \beta_1, \beta_2)$ involves the comparison of discounted social welfare along two entire paths, whose start and end points all depend on the parameters in question. Nevertheless, it is useful to decompose the effects that changes in the parameters have on immiserization into two effects. Namely, there is an effect on the steady state flow welfare and an effect on the transition path. The former constitutes the cost of decreasing infectiousness, while the latter can be thought of as a rough proxy for the associated benefits. To see this, note that since $I_2 > I_1$, along the

entire path from the old to the new steady state, it is the case that the marginal benefit to the individual of full exposure, is larger than the marginal cost in terms of increased infection risk. The reason that $|I_2 - I_1|$ is not an exact measure of these benefits, is that they are not liner in the base level I_1.

With this in mind, I now turn to the findings of the numerical experiments. The general patterns are as follows:

1. In order to ensure a positive effect on discounted social welfare, a sizeable decrease in infectiousness is necessary. In contrast, a moderate decrease in infectivity decreases discounted social welfare and thus creates immiserization. It is interesting that the immiserization region in (β_1, β_2)-space becomes broader as the level of the initial infectivity parameter β_1 increases. In other words, the higher initial infectiousness β_1 is, the harder it becomes to achieve positive welfare effects. Note that $\partial^2 W_D/\partial\beta^2 < 0$. Thus the larger initial infection is, the smaller the decrease in steady state welfare results from reducing infectivity to some lower level β_2. Similarly, since $\partial^2 I_D^*/\partial\beta^2 > 0$, a higher initial level of infectiousness causes a smaller increase in steady state prevalence from decreasing infectiousness. For the parameter ranges considered in the experiments, the former effect dominates for small decreases in infectiousness, while the latter effect dominates for larger decreases. It should be noted that for sufficiently low infectiousness, equilibrium steady state exposure (and hence steady state prevalence) is socially optimal.

2. The higher initial infectiousness β_1 is, the smaller in absolute value does $\Delta(I_1, I_2; \beta_1, \beta_2)$ tend to be. In other words, whether the welfare gains are negative or positive (i.e. whether there is immiserization or not), these seem to become more important the lower the initial level of infectiousness is. A possible reason for this is that for very high infectiousness levels, the difference between the socially optimal and the equilibrium steady state prevalence level is small. Thus, even if a decrease in infectiousness exacerbates this difference, it does so from a low base. For lower initial levels of infectiousness, the discrepancy between the socially optimal and the equilibrium steady state level is further amplified.

3. An increase in the recovery rate α, makes the immiserization region in (β_1, β_2)-space narrower, with immiserization ceasing altogether for low levels of β_1. That is, for a given β_1, the range of β_2 for which there is immiserization becomes narrower. Nonetheless, conditional on immiserization occurring both before and after the change in recovery rate α, the severity of immiserization is largely unaffected. Note that the rate α does not affect steady state welfare and thus any effects of changing it, stem from the transition path. Since $\partial^2 I_D^*/\partial\beta\partial\alpha < 0$, increasing the recovery rate exacerbates the increase in steady state disease prevalence brought about by decreasing the infectiousness of the disease. For the parameter values considered in the experiments, the net effects of changing the recovery rate are very small.

4. An increase in the health premium π, does not materially alter the immiserization region in (β_1, β_2)-space, but tends to worsen immiserization when it occurs. Note that the health premium does not influence the flow welfare levels in steady state. This means that the overall effects of changing π must be attributed to the effects on the transition path. It is easily verified that $\partial^2 I_D^*/\partial\beta\partial\pi > 0$. This means that

increasing the health premium in fact attenuates the increase in steady state prevalence brought about by decreasing infectiousness. In short, the health premium has no impact on the steady state, but shortens the transition path (and the associated benefits), thus worsening the severity of the immiserization.

5. An increase in the prevention cost c, does not materially change the immiserization region in (β_1, β_2)-space. It does though, change the magnitude of the immiserization, with higher prevention costs c leading to a more severe drop in social welfare. Since $\partial^2 W_D/\partial\beta\partial c > 0$, increasing c exacerbates the steady state welfare loss brought about by decreasing infectivity. On the other hand, $\partial^2 I_D^*/\partial\beta\partial c < 0$, which means that increasing c also exacerbates the increase in steady state prevalence brought about by decreased infectivity. For the parameterizations considered in the experiments, the former effect dominates, thereby tending to worsen the degree of immiserization.

6. An increase in the discount rate ρ, does not materially change the immiserization region in (β_1, β_2)-space. Furthermore, for reasonable values of ρ, the severity of immiserization does not change significantly. Increasing the discount rate exacerbates the drop in steady state welfare caused by decreased infectiousness, since $\partial^2 W_D/\partial\beta\partial\rho > 0$. On the other hand, $\partial^2 I_D^*/\partial\beta\partial\rho < 0$, so increasing ρ also exacerbates the increase in steady state disease prevalence brought about by decreasing infectiousness. For the parameter values considered in the experiments, the net effects of changing the discount factor within the range 0.8-0.99 are very small.

7. As a check, the experiments were also run for the case of centralized decision making. As can be seen in the lower panel of Table 2, in this scenario, the welfare gain $\Delta(I_1^*, I_2^*; \beta_1, \beta_2)$ is found to always be decreasing in β_2, confirming that the planner can never be worse off by decreasing infectiousness. In this context, it is worth noting that the social welfare gains under decentralized decision making are substantially lower than those under centralized decision making. This illustrates both the severity of the non-internalized externality and smallness effects and the desirability of well-designed policy intervention.

For a concrete example of how welfare losses can result from a permanent lowering of the infectiousness of the disease, consider a reduction from $\beta_1 = 6$ to $\beta_2 = 5.5$, with the other parameters given as before. Under centralized decision making, this reduction of infectiousness takes the system from steady state prevalence $I_1^* = 0.000187$ to $I_2^* = 0.000204$, in the process increasing discounted social welfare, from $V(I_1^*, I_1^*, \beta_1) = -0.766667$ to $V(I_1^*, I_2^*; \beta_2) = -0.754725$.

Turning to decentralized decision making, the same reduction in infectiousness takes the system from steady state disease prevalence $I_1 = 0.001871$ to $I_2 = 0.002041$. In contrast to the the situation under central planning, under decentralized decision making this reduction in infectiousness *decreases* discounted social welfare, from $V(I_1, I_1; \beta_1) = -0.916667$ to $V(I_1, I_2; \beta_2) = -0.917353$. For ease of comparison, I summarize these numbers in the table below:

	Status Quo	**Intervention**	**Status Quo**	**Intervention**
Dec.	$I_1 = 0.001871$	$I_2 = 0.002041$	$V(I_1, I_1; \beta_1) = -0.916667$	$V(I_1, I_2; \beta_2) = -0.917353$
Cen.	$I_1^* = 0.000187$	$I_2^* = 0.000204$	$V(I_1^*, I_1^*, \beta_1) = -0.766667$	$V(I_1^*, I_2^*; \beta_2) = -0.754725$

Although $W_D|_{\beta=\beta_1} > W_D|_{\beta=\beta_2}$, so the decrease in infectiousness decreases steady state welfare under decentralized decision making, $W_C|_{\beta=\beta_1} < W_C|_{\beta=\beta_2}$ for this parameterization (since $\alpha = 1 > 0.11 = \rho$). This means that under centralized decision making, there is no tradeoff between gains on the transition path and losses from being in a new steady state. Both contribute towards increased social welfare. Under decentralized decision making though, the new steady state is indeed worse and in this case the decrease in discounted steady state welfare, dominates the benefits of increased exposure on the transition path.

To reiterate a point made above, it should be emphasized that *if* a decrease in infectivity indeed increases overall welfare under decentralized decision making, then the increase is derived from the fact that on the path towards the new (and higher) level of steady state prevalence, the individuals fully expose themselves to infection (i.e. they do not engage in any preventive behavior at all). It is precisely the immediate welfare gains associated with these higher exposure levels on the transition path, that account for the overall increase in welfare. In short, welfare is increased *despite* the fact that it leads to higher levels of infection.

For completeness, it should be emphasized that the criterion used to detect welfare improvements used in this thought experiment is very strict, because it assumes that infectivity can be reduced at *no cost*. In other words, if there are parameters for which reducing infectivity does not improve welfare, then this means that it would be a socially unattractive policy, even if it could be implemented for free. Clearly, introducing the more realistic assumption that reducing infectivity is in fact costly, would only further reduce the desirability of the policy in question. The costs of permanently reducing the infectivity of a disease may be several orders of magnitude larger than the cost of moment-by-moment preventive effort.

It follows clearly from this example that although reducing the infectiousness of a disease permanently has strong intuitive appeal, a more careful analysis is warranted. Such an analysis shows that the policy may lead to immiserization and therefore be wholly counterproductive, negating the intended policy aim.

Since an important infection against which pre-exposure prophylaxis is indicated is HIV, it is proper to discuss this special case separately. There is currently no cure for HIV and it is therefore properly modeled as an SI type disease rather than as an SIS type disease (as done e.g. in Geoffard and Philipson, 1996). That is, the relevant special case of the present model is that in which $\alpha = 0$. But in this special case, it has been shown that decentralized equilibrium outcomes are socially optimal and thus there cannot be any immiserization. As noted elsewhere, this conclusion is not warranted in general, since the rather striking efficiency result is an artifact of population homogeneity. Once heterogeneity is introduced to the model, inefficiencies in equilibrium outcomes would persist even when recovery is no longer possible. For completeness, such a setting is discussed in Section 7. In a model with population heterogeneity, a thought experiment analogous to the one presented above can be carried out numerically.

6.2. Interventions that Influence Recovery. As shown above, an important source of distortion of decentralized exposure decisions is the possibility of recovery from infection, captured by the parameter α. At first blush, reducing α may seem like a worthy objective, since it would appear to narrow the gap between public and private incentives. There are several reasons why this may not be a sensible policy. The first is that the

social optimality of the decentralized equilibrium in the SI model is not robust to the introduction of population heterogeneity (see Section 7 for details). Thus the idea of achieving the first best through reducing the recovery rate, may be a red herring and only eliminate one source of unrealized positive externalities. Second, observe that in practice, reducing the recovery rate is likely achieved by reducing the intensity of treatment of infected individuals (that is, if treatment is feasible). But if treatment is feasible and chosen optimally, then the analysis in the present paper no longer applies and more complicated policies must be considered. Once prevention and treatment are chosen in conjunction, the dynamics of the model become very complicated and there may be multiple steady states across which the comparative statics results vary (see Rowthorn and Toxvaerd, 2015 for details). Furthermore, if treatment is available and optimally chosen, then the optimal level of prevention may no longer correspond to the level derived here.

Last, it should also be noted that it may be very difficult to persuade the public of the wisdom of seeking to increase overall welfare by decreasing the recovery rate. Thus this type of policy seems untenable in practice, even if found to be socially desirable.

6.3. Interventions Through Dynamic Subsidies and Taxes. As emphasized above, the main source of distortions to individual exposure levels, is that individuals do not properly internalize the pure externality effect and the smallness effect. This suggests considering policies that seek to reduce these distortions directly. One such possibility is to align individuals' incentives with those of the central planner, through subsidies and taxes. Rowthorn and Toxvaerd (2015) derive such optimal schemes in more general settings and show that the socially optimal outcomes can be achieved through either (i) direct subsidies to prevention, (ii) a tax on infected individuals or (iii) a subsidy to uninfected individuals. In all cases, the schemes are fairly complicated objects, being both dynamic and state dependent in nature. In particular, they are not simple Pigouvian schemes that internalize the pure externality effect, but designed to also correct for the additional smallness effect. Even though such schemes may work in principle, there are both practical and moral considerations that may make them difficult to implement in practice.

6.4. Social and Private Usefulness of Prevention. To complete the analysis, I will briefly discuss the social usefulness of continual prevention as a tool to control infection. First, note that the eradication steady state, which is obtained through sustained protection, is ruled out under both centralized and decentralized decision making, regardless of parameter values. This means that in steady state, prevention effort must necessarily be less than 100%. In contrast, the endemic steady state with full exposure (i.e. with no preventive effort at all) may be viable, if prevention costs are large enough. It turns out that optimal full exposure (i.e. no prevention at all) and the resulting endemic steady state is viable for a wider range of parameters in the decentralized setting than in the centralized setting. Specifically, in the decentralized setting, full exposure is privately optimal for the individual if

$$c > \pi \left(\frac{\beta - \alpha}{\rho + \beta} \right) \tag{63}$$

which is ruled out by assumption (6).[38] In the centralized setting, full exposure is optimal if

$$c > \pi \left(\frac{\beta - \alpha}{\rho + \beta - \alpha} \right) \tag{64}$$

In other words, for prevention to be privately optimal under decentralized decision making, the costs of prevention must be lower than the level that would make prevention socially optimal with centralized decision making. This means that there are cost levels for which the planner would choose positive levels of prevention in steady state, while the equilibrium under decentralized decision making would involve no prevention at all.

7. Model with Heterogeneous Population and Without Recovery

For analytical convenience, this paper has adopted the framework of the SIS model, which greatly simplifies the analysis by making the problem stationary. But the ideas hold more generally and a similar analysis can be performed in other environments, such as the SI model, which I will briefly discuss in this section. Consider the model in the main text and assume that there is no recovery, i.e. that $\alpha = 0$. Assume further that there are two types of individuals, which for simplicity are assumed to differ only with respect to their prevention costs. A fraction of the population has cost $c > 0$, while the remainder has cost $k > 0$, with $k < c$. It is immediately clear that the structure of the planner's solution to this problem and that of the equilibrium under decentralized decision making is the same. Under decentralization, there will be two thresholds I_D^k and I_D^c such that individuals with cost k switch from no protection to full protection when disease prevalence reaches I_D^k, while individuals with cost c switch when prevalence reaches I_D^c. In each case, because the individuals ignore the effects that their decisions have on the other individuals, the threshold is calculated using the formula for I_D^* evaluated at the relevant cost (and with $\alpha = 0$).

In the planner's problem, the solution is given by two thresholds I_C^k and I_C^c at which k and c cost individuals switch from no prevention to full prevention, respectively. Again, the threshold I_C^c is calculated using the formula for I_C^*, evaluated at cost c (which will coincide with I_D^*). But since the planner takes into account that non-protected k cost individuals have negative externalities on non-protected c cost individuals, the social planner will choose a threshold $I_C^k < I_C^*$ (evaluated at cost k). In other words, when there is no recovery but the population is heterogeneous, the decentralized equilibrium pattern of protection is no longer socially efficient.

Turning to the issue of immiserization, a decrease in infectiousness β would increase all the critical thresholds $(I_D^k, I_D^c, I_C^k, I_C^c)$. Thus the same thought experiment as that performed in Section 6 can be carried out. That is, one can compare the discounted social value of staying at the original steady state, with the old infectiousness parameter, with the social value of reducing the infectiousness parameter and going to the new steady state. As in the SIS model, the two effects analyzed in Section 5 are still present and these are still state dependent. This suggests that similar welfare effects and tradeoffs identified within the SIS setting continue to be relevant more broadly.

[38] See Appendix C for details of the parameter restrictions.

8. Continual Prevention, Permanent Prevention and Abstinence

In this section, I offer a detailed discussion of the relation between different ways that preventive behavior has been modeled in the literature. In particular, I will juxtapose moment-by-moment measures that only yield short-term protection (such as condoms), with long-term measures such as vaccination. It turns out that under very restrictive assumptions, the predictions of the two types of model coincide. I show that this is merely a curiosity and the effect of simplifying modeling assumptions. Last, I discuss the relation to models of abstinence and of quarantines.

In the present model, preventive behavior is modeled as an ongoing activity that must be sustained through time, in order to remain effective against infection. In other words, the protective effects are modeled as being entirely transitory, as is the case with e.g. condoms. The upshot is that prevention decisions are particularly simple, depending only on the current population-wide infection level. By contrast, vaccinations provide a more prolonged (or even permanent) protection. Vaccination decisions are therefore necessarily forward-looking in nature.

If decision makers can perfectly commit to full (continual) prevention from some point in time onwards, then the model with no spontaneous recovery (i.e. with $\alpha = 0$) reduces to a model with vaccination. To be more precise, at a given instant $t \geq 0$, individuals that are protected at instant t cannot be infected. In a model with perfect vaccination, no individual that has been vaccinated at any time $s < t$ in the past can be infected at time t. Thus the temporal extent of the effects of preventive behavior are a key ingredient in understanding the relation between the different models and their predictions.

Most dynamic models of vaccination following Anderson and May (1991), such as Geoffard and Philipson (1997), Francis (1997) and Francis (2004), are essentially modified *susceptible-infected-recovered* models (also known as SIR models). In the classical SIR model, infected individuals recover spontaneously and thereby acquire permanent immunity to further infection. In vaccination type models, vaccination takes susceptible individuals directly to the class of recovered (and thus immune) individuals, thereby bypassing a spell of infection for those vaccinated individuals before they acquire immunity. It is well known that disease prevalence is non-monotonic in the classical SIR model; it first increases and then decreases, due to the effects of herd immunity.

In the important analysis of vaccination by Francis (1997), individuals are homogeneous and as a result, equilibrium (centralized as well as decentralized) involves all individuals vaccinating at exactly the same point in time, namely when a critical level of disease prevalence is reached. Specifically, this means that until that level is reached, there is no vaccination at all, whereas after that point in time, all remaining susceptible individuals vaccinate. The equilibrium path of the model is therefore quite simple to describe. Until the critical threshold is reached, the equilibrium path of disease prevalence of the homogeneous model with vaccination, is identical to that of the simple SI model, i.e. it is increasing over time. Once the threshold is reached, all uninfected individuals vaccinate and therefore disease prevalence remains constant thereafter.[39]

Turning to continual prevention without the possibility of recovery or immunity, the decision to protect oneself at a point in time rests solely on disease prevalence at that point in time. Specifically, full prevention is optimal for prevalence levels above a critical threshold, while zero prevention is optimal for prevalence levels below it. But as long as

[39] This discussion is confined to models of closed populations.

no individual engages in preventive behavior, the equilibrium path of disease prevalence coincides with that of the SI model, i.e. it is increasing over time. Therefore, if prevention is optimal at any point in time, it remains optimal in perpetuity. In other words, a "commitment" to maintain full preventive effort is credible, because the incentive to prevent infection becomes stronger as time passes and disease prevalence increases.

The upshot is that in the special case $\alpha = 0$, the equilibrium paths of the two models coincide exactly.[40] Despite the discussion above, this close correspondence between the two models is rather striking. The reason that this correspondence is non-trivial is that while in the continual prevention framework individuals' optimal behavior is necessarily myopic, in the vaccination model individuals must solve a potentially complicated continuation game at each point in time (because the decision to vaccinate is forward-looking). It is therefore not a priori clear why the conditions that prompt individuals to vaccinate are exactly the conditions prompting them to engage in preventive behavior. Chen and Toxvaerd (2014) contains a detailed exposition of the externalities involved in vaccination decisions under decentralized decision making.

Although this discussion has mainly focused on the special case of the SIS model in which $\alpha = 0$, it should be emphasized that the vaccination model to which this setting is compared, is also a special case of a more general model. Specifically, the model in which only vaccination confers immunity from further infection, is a special case of the setting studied in Francis (2004), in which immunity can be achieved both through vaccination and through spontaneous recovery at exogenous rate $\gamma > 0$. The steady state of that (more general) model is very similar to that of the classical SIR model. Specifically, the disease eventually dies out and individuals are either immune (through recovery) or susceptible in perpetuity (but protected through herd immunity). In other words, if there is a positive rate of recovery and associated acquired immunity, then the vaccination model's predictions differ radically from those of the continual prevention model.

A more direct and important distinction between the two modes of prevention is that while continual prevention can sensibly be modeled within an SIR or an SIS framework, vaccination makes sense only within an SIR setting (and not within an SIS setting). Therefore, received wisdom from research on vaccination models, is of little use in a framework with recovery like the one studied here.

For completeness, it should be noted that the reason that the Francis (1997) model has a monotonic prevalence path in equilibrium (although it is formally an SIR type model), is that once infected, individuals cannot recover. Infected individuals therefore constitute a perpetual source of infection to the susceptible population. In contrast, in the classical SIR model, recovered (and hence immune) individuals serve to stifle the propagation of infection in the population, eventually halting it altogether.

To sum up this discussion, there is not in general a direct correspondence between the models of continual and permanent protection, i.e. of moment-by-moment prophylaxis and vaccination. In the special case of the former model in which $\alpha = 0$, the steady state outcome resembles that of the special case of the latter model in which $\gamma = 0$. If either $\alpha > 0$ or $\gamma > 0$ (or both), then the predictions of the two models are incomparable.

[40] Indeed, the correspondence is exact. Letting $\pi = (\overline{u} - \underline{u})$, $N = 1$, $\rho = \delta$ and $c/\rho = \theta$, equation (14) in Francis (1997) is identical to equation (18) in the present paper, showing that the threshold prevalence above which individuals choose to vaccinate, is the same as the level above which they engage in preventive behavior.

Next, I turn to the model of Kremer (1996).[41] To understand the model and its conclusions, it is useful to consider the following analogy. The underlying model is of the susceptible-infected (or SI) variety, but with the following addition. Before contact between susceptible and infected individuals takes place, each decision maker decides between staying at home (the safe option) and exposing oneself to infection (the risky option). Only those susceptible individuals that choose the risky option can get infected. Were it not for the risk of infection, the risky option would dominate the safe option and so a susceptible individual faces a tradeoff between desirable exposure and the risk of welfare reducing infection.

It is immediately clear that all infected individuals will choose the risky option since they have nothing to lose from doing so. The decision of a typical susceptible individual is more complicated, depending on the assessed probability of infection. It is precisely the determination of this probability that makes the Kremer (1996) setting different from other SI type models like the one of Geoffard and Philipson (1996).

Kremer assumes that the probability of infection is proportional to the number of infected individuals *as a fraction of people choosing the risky option*. In particular, this means that the decision to stay at home (i.e. to choose the safe option) changes the rates of contact between susceptible and infected individuals that choose the risky option.

Rather than being a model of preventive measures (such as condom use), the model is more like a model of abstinence. Since the probability of getting infected is a decreasing function of the fraction of susceptible individuals that chooses the risky option, the game has strategic complementarities in the sense that the more susceptible individuals choose to expose themselves, the more attractive does exposure become to other susceptible individuals. In turn, this means that there can be multiple (self-confirming) equilibria. In one equilibrium, all individuals in the population expose themselves to infection. In the other equilibrium, only infected individuals expose themselves, whereas all susceptible individuals choose the safe option.

It is interesting to note that with a slight addition to the Kremer (1996) model, its predictions can be reconciled with those of the present model. Suppose that those individuals who choose the risky option, have the additional possibility of engaging in costly preventive behavior as modeled in the present paper. Furthermore, assume that prevention is not too costly, so that for sufficiently high probability of infection, it is still desirable to choose the risky option, but to do so accompanied by preventive measures. In this extended model, choosing the risky option at the first stage becomes a dominant strategy. At the second stage, prevention is chosen only if disease prevalence is sufficiently high and not otherwise. Thus the individuals never make use of the safe option and both optimal decisions and the evolution of the infectious disease exactly mirror those derived in the present paper.

In a sense, abstinence can be thought of as a self-imposed quarantine. The actual quarantine of whole segments of a population (typically of the already infected) is itself a type of preventive policy that may be considered. This is done e.g. by Sethi (1978). This type of policy differs from prevention is several respects. First, it targets the infected rather than the susceptible individuals, even if the latter are the intended beneficiaries. Second, in the prevention setting, it is the case of centralized decision making that is an idealized benchmark for the study of the more realistic decentralized decision making set-

[41] Related models are those of Auld (2003) and Goyal and Vigier (2015).

ting. With quarantines, only the case of centralized decision making seems sensible, since it is difficult to envision equilibrium outcomes in which infected individuals collectively quarantine themselves for the benefit of the susceptible. For these reasons, it seems that the study of optimal quarantines has very little direct relevance to the study of optimal prevention as formalized in the present work.

9. Conclusion

In this paper, I analyze a simple model of disease propagation in which individuals may engage in privately costly preventive behavior (but cannot control their rate of recovery). I analyze the outcomes under both centralized decision making by a benevolent social planner and under decentralized decision making by non-cooperative, forward-looking individuals. I find that in the decentralized equilibrium, individuals over-expose themselves to infection compared to the socially optimal level, and that this leads to socially suboptimal steady state disease prevalence and welfare. This is a result of the fact that individuals do not internalize the external effects of their private prevention decisions and because each have negligible effects on aggregate disease dynamics.

Next, I consider the desirability of different policy measures aimed at reducing the wedge between private and public incentives to engage in prevention, such as permanently reducing the infectiousness of the disease through pre-exposure prophylaxis (PrEP). I show that such a policy intervention may actually reduce overall discounted social welfare. This stems from the possibility of disinhibition (or risk compensation), which means that decision makers respond to lower infectiousness by increasing their exposure to infection. When considering such policies, the policy maker would therefore do well to keep in mind the maxim *"First, do no harm"*. Encapsulating a central concept in medical ethics, the advice reminds the physician that interventions have both costs and benefits and that inactivity may be preferable to prescribing a course of action that does more harm than good, however well-intentioned. The same rule also applies at the population level. That is, any policy measure intended to improve public health, must first meet this basic test. In this work, I explicitly show that for some parameter ranges, immiserization may well be the consequence of permanently reducing the disease's infectiousness. This calls for careful analysis before the introduction of such second-best policy interventions.

Appendix

A. Proofs that Steady States are Interior

A.1. The Centralized Setting. Because of the bang-bang nature of the optimal policy, there are only three policies to consider, namely the singular policy and the two

corner policies.[42] For $\varphi(t) = 1$ to be optimal in perpetuity, it must be that $\lambda(t)\beta I(t)+c > 0$ for all t. In this case, the laws of motion become

$$\begin{aligned} \dot{\lambda}(t) &= \lambda(t)\left[\rho + \alpha + \beta\left(2I(t) - 1\right)\right] + \pi \quad (65) \\ \dot{I}(t) &= I(t)\left[\beta(1 - I(t)) - \alpha\right] \quad (66) \end{aligned}$$

Next, consider the rate of change (over time) in the optimality condition. It is given by

$$\begin{aligned} \frac{d}{dt}\left[\lambda(t)\beta I(t) + c\right] &= \beta\left[\dot{\lambda}(t)I(t) + \lambda(t)\dot{I}(t)\right] \\ &= \beta I(t)\left[\lambda(t)(\rho + \beta I(t)) + \pi\right] \quad (67) \end{aligned}$$

The bracketed expression is positive if and only if

$$\lambda(t) \geq \frac{-\pi}{\rho + \beta I(t)} \quad (68)$$

This condition is implied by the inequality $\lambda(t) \geq \lambda_C^*$ whenever disease prevalence is above the singular solution, i.e. when $I(t) \geq I_C^*$. When disease prevalence is above this level, setting $\varphi(t) = 1$ perpetually is optimal. From full exposure, it then follows that in steady state $I(t) = (\beta - \alpha)/\beta$ and thus the multiplier is given by

$$\lambda(t) = \frac{-\pi}{\rho + \beta - \alpha} \quad (69)$$

But then

$$\lambda(t)\beta I(t) + c = \frac{-\pi(\beta - \alpha) + c\left(\rho + \beta - \alpha\right)}{\rho + \beta - \alpha} \quad (70)$$

This expression is positive if and only if

$$\pi < c\left(\frac{\beta + \rho - \alpha}{\beta - \alpha}\right) < c\left(\frac{\beta + \rho}{\beta - \alpha}\right) \quad (71)$$

But this inequality is inconsistent with (6) and thus a switch to the policy $\varphi(t) = 0$ is optimal. To rule out that setting $\varphi(t) = 0$ perpetually can be optimal, a slightly more delicate argument is needed. This is because as $I(t) \to 0$, the multiplier $\lambda(t) \to -\infty$ and hence it is not necessarily the case that $[\lambda(t)\beta I(t) + c] \to c > 0$, which would render $\varphi(t) = 1$ optimal. In other words, there is a discontinuity at zero prevalence.[43]

Instead, suppose that $\lambda(t)\beta I(t) + c < 0$, which makes $\varphi(t) = 0$ optimal. In this scenario, the laws of motion reduce to

$$\begin{aligned} \dot{\lambda}(t) &= \lambda(t)\left[\rho + \alpha\right] + \left[\pi - c\right] \quad (72) \\ \dot{I}(t) &= -\alpha I(t) \quad (73) \end{aligned}$$

[42] Note also that because of monotonicity of the optimal path $I(t)$, there can be at most one switch between these policies.

[43] I am grateful to Robert Rowthorn for pointing this out.

The rate of change (over time) in the optimality condition is given by

$$\begin{aligned}\frac{d}{dt}\left[\lambda(t)\beta I(t)+c\right] &= \beta\left[\dot{\lambda}(t)I(t)+\lambda(t)\dot{I}(t)\right]\\ &= \beta I(t)\left[\lambda(t)\rho+(\pi-c)\right]\\ &= \beta I(t)\left[\lambda(t)-\lambda_C^*\right] \qquad (74)\end{aligned}$$

Note that if the bracketed expression is negative and the supposition that $\lambda(t)\beta I(t)+c<0$ holds, then indeed setting $\varphi(t)=0$ will remain optimal in perpetuity. When $\varphi(t)=0$ for all t, the growth rate of the multiplier is given by $\dot{\lambda}(t)/\lambda(t)=\alpha+\rho$. As a result, when $I(t)\to 0$ it follows that $\lambda(t)\to-\infty$. Next, recall that $\lambda(t)$ is the current value multiplier and hence the present value multiplier is given by $\eta(t)\equiv e^{-\rho t}\lambda(t)$. But $\eta(t)$ grows at rate α which implies that $\lim_{t\to\infty}\eta(t)=-\infty$, thus violating the transversality condition $\lim_{t\to\infty}\eta(t)=0$.[44] In conclusion, the steady state must necessarily be interior (in both policy and state variable) ■

A.2. The Decentralized Setting. With some modifications, the proof follows similar steps as those in the centralized setting. For it to be optimal to set $\varepsilon_i(t)=1$ perpetually for all $i\in\mathcal{S}(t)$, it must be that $\mu(t)\beta I(t)+c>0$. In symmetric equilibrium, the laws of motion become

$$\dot{\mu}(t) = \mu(t)\left[\rho+\alpha+\beta I(t)\right]+\pi \qquad (75)$$

$$\dot{I}(t) = I(t)\left[\beta(1-I(t))-\alpha\right] \qquad (76)$$

The rate of change in the optimality condition is given by

$$\begin{aligned}\frac{d}{dt}\left[\lambda(t)\beta I(t)+c\right] &= \beta\left[\dot{\mu}(t)I(t)+\mu(t)\dot{I}(t)\right]\\ &= \beta I(t)\left[\mu(t)(\rho+\beta+\beta(1-I(t)))+\pi\right] \qquad (77)\end{aligned}$$

The bracketed expression is positive whenever

$$\mu(t)\geq\frac{-\pi}{\rho+\beta+\beta(1-I(t))}\equiv\widehat{\mu}(t) \qquad (78)$$

Note that $\widehat{\mu}(t)\geq\mu_D^*$ if

$$c\leq\pi\left(\frac{2\beta-\alpha-\beta I(t)}{2\beta+\rho-\beta I(t)}\right) \qquad (79)$$

which is implied by (6). Next, for $\mu(t)\geq\widehat{\mu}(t)$ setting $\varepsilon_i(t)=1$ in perpetuity is optimal. Therefore, in symmetric equilibrium full exposure implies that in steady state $I(t)=(\beta-\alpha)/\beta$. Then the multiplier is finite and given by

$$\mu(t)=\frac{-\pi}{\rho+\beta} \qquad (80)$$

[44] See Seierstad and Sydsaeter (1987, pp. 244-245). For any candidate endemic steady state, the multiplier is finite and hence the transversality condition is satisfied.

Using this in the optimality condition, gives

$$\mu(t)\beta I(t) + c = \frac{-\pi(\beta - \alpha) + c\,(\rho + \beta)}{\rho + \beta} \tag{81}$$

This expression is negative under (6) and therefore $\mu(t)\beta I(t) + c < 0$, making a switch to the policy $\varepsilon_i(t) = 0$ optimal. For $\mu(t) < \widehat{\mu}(t)$, setting $\varepsilon_i(t) = 1$ for all t cannot be an optimal path and the system returns to the singular solution.

For $\varepsilon_i(t) = 0$ for all t to be optimal, it must be that $\mu(t)\beta I(t) + c < 0$. In this case, the laws of motion in symmetric equilibrium are given by

$$\dot{\mu}(t) = \mu(t)\,[\rho + \alpha] + [\pi - c] \tag{82}$$
$$\dot{I}(t) = -\alpha I(t) \tag{83}$$

Again, consider the rate of change of the optimality condition:

$$\frac{d}{dt}\left[\lambda(t)\beta I(t) + c\right] = \beta\left[\dot{\mu}(t)I(t) + \mu(t)\dot{I}(t)\right] \tag{84}$$
$$= \beta I(t)\left[\mu(t)\rho + (\pi - c)\right] \tag{85}$$
$$= \beta I(t)\left[\mu(t) - \lambda_C^*\right] \tag{86}$$

For $\mu(t) \leq \lambda_C^*$, an analogous argument to the one in the centralized case shows that an optimal path with perpetual full prevention cannot satisfy the transversality condition.[45] Last, for $\mu(t) > \lambda_C^*$, setting $\varepsilon_i(t) = 0$ for all t cannot be an optimal path and the system returns to the singular solution. This completes the proof ■

B. Proofs that Optimal Policies Exist and are of the MRAP Type

B.1. The Centralized Setting. This appendix confirms that the conditions of Sethi (1977), Theorem 3.1 (i)-(vi) are satisfied for the centralized and decentralized problems respectively. I start with the former of these. From the logistic growth equation, solve for the exposure rate to get the expression

$$\varphi(t) = \frac{\dot{I}(t) + \alpha I(t)}{\beta I(t)(1 - I(t))} \tag{87}$$

Substituting this into the objective function and rearranging, yields the modified objective function

$$\int_0^\infty e^{-\rho t}\left[-I(t)[\pi - c] - c\left(\frac{\beta - \alpha}{\beta}\right) + \left(\frac{c}{\beta I(t)}\right)\dot{I}(t)\right]dt = \int_0^\infty e^{-\rho t}\left[M(I(t)) + N(I(t))\dot{I}(t)\right]dt \tag{88}$$

where

$$M(I(t)) \equiv -c\left(\frac{\beta - \alpha}{\beta}\right) - I(t)(\pi - c) \tag{89}$$
$$N(I(t)) \equiv \frac{c}{\beta I(t)} \tag{90}$$

[45] This is because for $\varepsilon_i(t) = 0$, the growth rate of $\mu(t)$ coincides with that of $\lambda(t)$ when $\varphi(t) = 0$, i.e. it is given by $\alpha + \rho$.

Next, define

$$\begin{aligned}\Delta^C(I(t)) &\equiv -\left(\rho N(I(t)) + M'(I(t))\right) \\ &= \frac{-\rho c}{\beta I(t)} + (\pi - c) \end{aligned} \tag{91}$$

First, note that $I(t) = I_C^*$ is the unique solution to the equation

$$\Delta^C(I(t)) = 0 \tag{92}$$

Next, given $I(t) = I_C^*$, $\varphi(t) = \varphi_C^*$ is the unique solution to the equation

$$\dot{I}(t) = 0 \tag{93}$$

Last, it is easy to verify that $\Delta^C(I(t)) > 0$ for $I(t) > I_C^*$ while $\Delta^C(I(t)) < 0$ for $I(t) < I_C^*$. This proves that conditions (i)-(iii) are satisfied. Conditions (iv)-(vi) hold trivially ■

B.2. The Decentralized Setting. Turning to the decentralized problem, solve the differential equation governing the probability of being infected for the individual exposure rate, to obtain

$$\varepsilon(t) = \frac{\dot{q}_i(t) + \alpha q_i(t)}{\beta I(t)(1 - q_i(t))} \tag{94}$$

Substitute into the individual's objective function and rearrange to get

$$\int_0^\infty e^{-\rho t}\left[-q_i(t)\left(\pi - c - \frac{\alpha c}{\beta I(t)}\right) - c + \left(\frac{c}{\beta I(t)}\right)\dot{q}_i(t)\right] dt = \int_0^\infty e^{-\rho t}\left[\widehat{M}(I(t)) + \widehat{N}(I(t))\dot{q}_i(t)\right] dt \tag{95}$$

where

$$\widehat{M}(I(t)) \equiv -q_i(t)\left(\pi - c - \frac{\alpha c}{\beta I(t)}\right) - c \tag{96}$$

$$\widehat{N}(I(t)) \equiv \frac{c}{\beta I(t)} \tag{97}$$

Next, define

$$\begin{aligned}\Delta^D(I(t)) &\equiv -\left(\rho \widehat{N}(I(t)) + \widehat{M}'(I(t))\right) \\ &= \frac{-(\rho + \alpha)c}{\beta I(t)} + (\pi - c) \end{aligned} \tag{98}$$

Here, I have made the substitution $q_i(t) = I(t)$. It is easily verified that $I(t) = I_D^*$ is the unique solution to the equation

$$\Delta^D(I(t)) = 0 \tag{99}$$

Given $q_i(t) = I(t) = I_D^*$, $\varepsilon_i(t) = \varepsilon_D^*$ is the unique solution to the equation

$$\dot{q}_i(t) = 0 \tag{100}$$

The last step is to note that $\Delta^D(I(t)) > 0$ for $I(t) > I_D^*$ while $\Delta^D(I(t)) < 0$ for $I(t) < I_D^*$. Conditions (iv)-(vi) hold trivially. This completes the proof ■

C. Parameter Restrictions

The following inequalities ensure that the endemic steady state solutions are indeed interior in the centralized and decentralized settings respectively:

$$I_C^* \leq 1 \quad if \quad c \leq \pi\left(\frac{\beta}{\beta+\rho}\right) \tag{101}$$

$$I_C^* \leq \frac{\beta-\alpha}{\beta} \quad if \quad c \leq \pi\left(\frac{\beta-\alpha}{\beta-\alpha+\rho}\right) \tag{102}$$

$$\varphi_C^* \leq 1 \quad if \quad c \leq \pi\left(\frac{\beta-\alpha}{\beta-\alpha+\rho}\right) \tag{103}$$

$$I_D^* \leq 1 \quad if \quad c \leq \pi\left(\frac{\beta}{\alpha+\beta+\rho}\right) \tag{104}$$

$$I_D^* \leq \frac{\beta-\alpha}{\beta} \quad if \quad c \leq \pi\left(\frac{\beta-\alpha}{\beta+\rho}\right) \tag{105}$$

$$\varepsilon_i^* \leq 1 \quad if \quad c \leq \pi\left(\frac{\beta-\alpha}{\beta+\rho}\right) \tag{106}$$

All these conditions are simultaneously satisfied under the assumption in (6).

D. DETAILS OF NUMERICAL EXPERIMENTS

The numerical experiments documented in the text were carried out with MATLAB R2011b. In order to make the welfare assessments, the value of the following two paths must be calculated. First, one needs to find the value $V(I_1, I_1; \beta_1)$ of staying indefinitely at I_1. This poses no particular computational problem, since it is simply a discounted integral of a constant. Second, one needs to find the value of $V(I_1, I_2; \beta_2)$. This integral can conveniently be decomposed into the discounted value of moving from steady state I_1 to steady state I_2, and the discounted value of staying indefinitely at I_2, i.e. $V(I_2, I_2; \beta_2)$. The latter is again a discounted integral of a constant and is straightforward to calculate. This leaves the calculation of the transition path, for which a fourth-order Runge-Kutta procedure was used. Because the new steady state I_2 is saddle-path stable, it is very difficult to initialize the algorithm at I_1 and ensure convergence to I_2. Instead, the following procedure is used. The algorithm is initialized at the new steady state I_2 (and its corresponding costate μ_2), and run backwards until the old steady state disease prevalence level I_1 is reached (with its corresponding costate $\hat{\mu}_1$, which constitutes the initial condition of the transition path of interest). Once this is done, the algorithm is initialized at point $(I_1, \hat{\mu}_1)$ and let to run forward to the new steady state I_2. This procedure ensures convergence, by essentially finding the appropriate initial condition for the algorithm. The experiments were run with a precision tolerance of 10^{-6}. The time interval for integration was set to $h = 0.01$.

The baseline parameterization was $\alpha = 1$, $\rho = 0.12$, $c = 0.1$ and $\pi = 10$. The values of the infectivity parameters where chosen such that $\beta_1, \beta_2 \in [4, 10]$, with increments of 0.5. The experiments were also run with finer grids, but this did not alter the results, while significantly increasing the number of computations. For each set of parameters, the program checked that Assumptions 1 and 2 were satisfied both at I_1 and I_2. To check robustness, the following alternative parameters were chosen: For the recovery rate, I chose values $\alpha \in [0.2, 3.8]$, with increments of 0.2. For the discount rate, I chose values of the discount factor $\delta \equiv 1/(1+\rho) \in \{0.8, 0.85, 0.9, 0.95, 0.99\}$. For the prevention cost and the health premium, I chose the alternative values $c = 1.5$ and $\pi = 15$, respectively.

E. THE MAVERICK'S PROBLEM

To formally derive the shadow price of the maverick, consider a situation in which, under central direction, all individuals except for the maverick behave in the socially optimal fashion. Denote the aggregate level of infection on the resulting optimal path by $I(t)$. The maverick individual evades instructions and chooses his or her own level of exposure. As before, this individual takes the trajectory of aggregate infection $I(t)$ as given. This maverick's objective is to solve

$$\max_{\varepsilon_i(t)\in[0,1]} \int_0^\infty e^{-\rho t} \left[-q_i(t)\pi - (1-q_i(t))(1-\varepsilon_i(t))c\right] dt \tag{107}$$

$$s.t. \quad \dot{q}_i(t) = \varepsilon_i(t)\beta I(t)(1-q_i(t)) - \alpha q_i(t), \quad q_i(0) \in \{0,1\} \tag{108}$$

But note that in this problem, I control for optimal aggregate infection, i.e. the path of infection is not allowed to follow its decentralized equilibrium path.

The current-value Hamiltonian for the maverick's problem is given by

$$H_M^D \equiv -q_i(t)\pi - (1-q_i(t))(1-\varepsilon_i(t))c \tag{109}$$
$$+\eta_i(t)\left[(1-q_i(t))\varepsilon_i(t)\beta I(t) - q_i(t)\alpha\right] \tag{110}$$

Note that in this equation, the path $I(t)$ is the socially optimal one and thus H_M^D differs from H_i^D only because they are evaluated along different paths for aggregate disease prevalence. Next, the costate variable for the maverick's problem evolves according to the differential equation

$$\dot{\eta}_i(t) = \eta_i(t)\left[\rho + \alpha + \varepsilon_i(t)\beta I(t)\right] + \pi - (1-\varepsilon_i(t))c \tag{111}$$

Note that because of symmetry, I can drop the subscript i on the costate and control variables. For completeness, the transversality condition $\lim_{t\to\infty} e^{-\rho t} H_M^D(t) = 0$ holds, and the Arrow sufficiency conditions for an individual optimum are also satisfied ■

REFERENCES

[1] AADLAND, D., D. FINNOFF AND K. X. D. HUANG (2010): Syphilis Cycles, *mimeo.*

[2] ABBAS, U. L., R. M. ANDERSON AND J. W. MELLORS (2007): Potential Impact of Antiretroviral Chemoprophylaxis on HIV-1 Transmission in Resource-Limited Settings, *PLoS ONE*, 2(9): e875.

[3] ANDERSON, R. M. AND R. M. MAY (1991): Infectious Diseases of Humans: Dynamics and Control, *Oxford University Press.*

[4] AULD, M. C. (2003): Choices, Beliefs, and Infectious Disease Dynamics, *Journal of Health Economics*, 22(3), 361-377.

[5] AUTIER, P., J.-F. DORE, S. NEGRIER, D. LIENARD, R. PANIZZON, F. J. LEJEUNE, D. GUGGISBERG AND A. M. M. EGGERMONT (1999): Sunscreen Use and Duration of Sun Exposure: A Double-Blind, Randomized Trial, *Journal of the National Cancer Institute*, 91(15), 1304-1309.

[6] BARRETT, S. (2003): Global Disease Eradication, *Journal of the European Economic Association*, 1(2-3), 591-600.

[7] BATE, A. M., G. JONES, A. KLECZKOWSKI, A. MACLEOD, R. NAYLOR, J. TIMMIS, J. M. TOUZA AND P. C. L. WHITE (2016): Modelling the Impact and Control of an Infectious Disease in a Plant Nursery with Infected Plant Material Inputs, *mimeo.*

[8] BERNOULLI, D. (1766): An Attempt at a New Analysis of the Mortality Caused by Smallpox and of the Advantages of Inoculation to Prevent It, in L. Bradley (1971): Smallpox Inoculation: An Eighteenth Century Mathematical Controversy, *University of Nottingham.*

[9] BLOWER, S. M., H. B. GERSHENGORN AND R. M. GRANT (2000): A Tale of Two Futures: HIV and Antiretroviral Therapy in San Francisco, *Science*, 287(5453), 650-654.

[10] BLUMENTHAL, J. AND R. H. HAUBRICH (2014): Will Risk Compensation Accompany Pre-Exposure Prophylaxis for HIV?, *AMA Journal of Ethics*, 16(11), 909-915.

[11] BOULIER, B. L., T. S. DATTA AND R. S. GOLDFARB (2007): Vaccination Externalities, *B.E. Journal of Economic Analysis & Policy*, 7(1), (Contributions), Article 23.

[12] BRITO, D. L., E. SHESHINSKI AND M. D. INTRILIGATOR (1991): Externalities and Compulsory Vaccinations, *Journal of Public Economics*, 45(1), 69-90.

[13] CASSELL, M. M., D. T. HALPERIN, J. D. SHELTON AND D. STANTON (2006): Risk Compensation: The Achilles' Heel of Innovations in HIV Prevention? *BMJ*, 332(7541), 605-607.

[14] CERDEIRO, D. (2015): Contagion Exposure and Protection Technology, *mimeo.*

[15] CHAN, N. W. AND K. GILLINGHAM (2015): The Microeconomic Theory of the Rebound Effect and its Welfare Implications, *Journal of the Association of Environmental and Resource Economists*, 2(1), 133-159.

[16] CHAN, T. Y., B. H. HAMILTON AND N. W. PAPAGEORGE (2012): The Value of Pharmaceutical Innovation: AIDS, Antivirals and the Joy of (Risky) Sex, *mimeo.*

[17] CHEN, F. AND F. TOXVAERD (2014): The Economics of Vaccination, *Journal of Theoretical Biology*, 363, 105-117.

[18] COHEN, C. R., M. MONTANDON, A. W. CARRICO, S. SHIBOSKI, A. BOSTROM, A. OBURE, Z. KWENA, R. C. BAILEY, R. NGUTI AND E. A. BUKUSI (2009): Association of Attitudes and Beliefs towards Antiretroviral Therapy with HIV-Seroprevalence in the General Population of Kisumu, Kenya, *PLoS ONE*, 4(3): e4573.

[19] CREPAZ, N., T. A. HART AND G. MARKS (2004): Highly Active Antiretroviral Therapy and Sexual Risk Behavior: A Meta-Analytic Review, *Journal of the American Medical Association*, 292(2), 224-36.

[20] DALEY, D. J. AND J. GANI (2001): Epidemic Modelling: An Introduction, *Cambridge Studies in Mathematical Biology.*

[21] DURANTON, G. AND M. A. TURNER (2009): The Fundamental Law of Road Congestion: Evidence from US Cities, *NBER Working Paper No. 15376.*

[22] DURANTON, G. AND M. A. TURNER (2011): The Fundamental Law of Road Congestion: Evidence from US Cities, *American Economic Review*, 101(6), 2616-2652.

[23] EATON, L. A. AND S. C. KALICHMAN (2007): Risk Compensation in HIV Prevention: Implications for Vaccines, Microbicides, and Other Biomedical HIV Prevention Technologies, *Current HIV/AIDS Reports*, 4(4), 165-172.

[24] FRANCIS, P. J. (2004): Optimal Tax/Subsidy Combinations for the Flu Season, *Journal of Economic Dynamics & Control*, 28(10), 2037 - 2054.

[25] FRANCIS, P. J. (1997): Dynamic Epidemiology and the Market for Vaccinations, *Journal of Public Economics*, 63(3), 383-406.

[26] GEOFFARD, P.-Y. AND T. PHILIPSON (1996): Rational Epidemics and Their Public Control, *International Economic Review*, 37(3), 603-624.

[27] GERSOVITZ, M. (2010): Disinhibition and Immiserization in a Model of Susceptible-Infected-Susceptible (SIS) Diseases, *mimeo.*

[28] GERSOVITZ, M. AND J. S. HAMMER (2004): The Economical Control of Infectious Diseases, *Economic Journal*, 114(492), 1-27.

[29] GILLINGHAM, K., D. RAPSON AND G. WAGNER (2015): The Rebound Effect and Energy Efficiency Policy, *Review of Environmental Economics and Policy*, 10(1), 68-88.

[30] GOLDMAN, S.M. AND J. LIGHTWOOD (1995): The SIS Model of Infectious Disease with Treatment, *mimeo.*

[31] GOLDMAN, S.M. AND J. LIGHTWOOD (2002): Cost Optimization in the SIS Model of Infectious Disease with Treatment, *Topics in Economic Analysis and Policy*, 2(1), 1-22.

[32] GOYAL, S. AND A. VIGIER (2015): Interaction, Protection and Epidemics, *Journal of Public Economics*, 125(), 64–69.

[33] GREENWOOD, J., P. KIRCHER, C. SANTOS AND M. TERTILT (2013): An Equilibrium Model of the Malawian HIV/AIDS Epidemic, *mimeo.*

[34] HARTL, R. F. (1987): A Simple Proof of the Monotonicity of the State Trajectories in Autonomous Control Problems, *Journal of Economic Theory*, 41(1), 211-215.

[35] HARTL, R. F. AND G. FEICHTINGER (1987): A New Sufficient Condition for Most Rapid Approach Paths, *Journal of Optimization Theory and Applications*, 54(2), 403-411.

[36] HEDLUND, J. (2000): Risky Business: Safety Regulations, Risk Compensation, and Individual Behavior, *Injury Prevention*, 6(2), 82-89.

[37] JOHNSON, N. P. A. S. AND J. MUELLER (2002): Updating the Accounts: Global Mortality of the 1918-1920 "Spanish" Influenza Pandemic, *Bulletin of the History of Medicine*, 76(1), 105-115.

[38] KARIM, Q. A. ET AL. (2010): Effectiveness and Safety of Tenofovir Gel, an Antiretroviral Microbicide, for the Prevention of HIV Infection in Women, *Sciencexpress Report*, 20 July 2010.

[39] KEELING, M. J. AND P. ROHANI (2008): Modeling Infectious Diseases in Humans and Animals, *Princeton University Press.*

[40] KREMER, M. (1996): Integrating Behavioral Choice into Epidemiological Models of AIDS, *Quarterly Journal of Economics*, 111(2), 549-573.

[41] KREMER, M. AND C. M. SNYDER (2003): Why Are Drugs More Profitable Than Vaccines?, *NBER Working Paper*, No. 9833.

[42] LAKDAWALLA, D., N. SOOD AND D. GOLDMAN (2006): HIV Breakthroughs and Risky Sexual Behavior, *Quarterly Journal of Economics*, 121(3), 1063-1102.

[43] LANGER, W. L. (1970): The Black Death, *Scientific American*, 222, 114-121.

[44] MARTIN J. N., M. E. ROLAND, T. B. NEILANDS, M. R. KRONE, J. D. BAMBERGER, R. P. KOHN, M. A. CHESNEY, K. FRANSES, J. O. KAHN, T. J. COATES AND M. H. KATZ (2004): Use of Postexposure Prophylaxis Against HIV Infection Following Sexual Exposure Does Not Lead to Increases in High-Risk Behavior, *AIDS*, 18(5), 787-92.

[45] OVER, M. (2008): Opportunities for Presidential Leadership on AIDS: From an "Emergency Plan" to a Sustainable Policy, in The White House and the World: A Global Development Agenda for the Next U.S. President, N. Birdsall (ed.), *Center for Global Development, Washington, D.C.*

[46] OVER, M., P. HEYWOOD, J. GOLD, I. GUPTA, S. HIRA AND E. MARSEILLE (2004): HIV/AIDS Treatment and Prevention in India. Modeling the Costs and Consequences, *The World Bank*.

[47] PALTIEL, A. D., K. A. FREEDBERG, C. A. SCOTT, B. R. SCHACKMAN, E. LOSINA, B. WANG, G. R. SEAGE III, C. E. SLOAN, P. E. SAX, AND R. P. WALENSKY (2009): HIV Preexposure Prophylaxis in the United States: Impact on Lifetime Infection Risk, Clinical Outcomes, and Cost-Effectiveness, *Clinical Infectious Diseases*, 48, 806-815.

[48] PHILIPSON, T., AND R. A. POSNER (1993): Private Choices and Public Health: An Economic Interpretation of the AIDS Epidemic. Cambridge, MA: Harvard University Press.

[49] RELUGA, T. C. (2009): An SIS Epidemiology Game with Two Subpopulations, *Journal of Biological Dynamics*, 1751-3766.

[50] RELUGA, T. C. (2010): Game Theory of Social Distancing in Response to an Epidemic, *PLoS Computational Biology*, 6(5).

[51] RELUGA, T. C. AND A. P. GALVANI (2011): A General Approach for Population Games with Application to Vaccination, *Mathematical Biosciences*, 230(2), 67-78.

[52] RICHENS, J., J. IMRIE AND A. COPAS (2000): Condoms and Seat Belts: The Parallels and the Lessons, *The Lancet*, 355, 400-403.

[53] ROWTHORN, R. (2006): The Optimal Treatment of Disease Under a Budget Constraint, in R. Halvorsen and D. Layton (eds), *Explorations in Environmental and Natural Resource Economics: Essays in Honor of Gardner M. Brown, Jr,* Edward Elgar.

[54] ROWTHORN, R. AND F. TOXVAERD (2015): The Optimal Control of Infectious Diseases via Prevention and Treatment, *mimeo*.

[55] SANDERS, J. L. (1971): Quantitative Guidelines for Communicable Disease Control Programs, *Biometrics*, 27(4), 883-893.

[56] SEIERSTAD, A. AND K. SYDSAETER (1987): Optimal Control Theory with Economic Applications, *North Holland*.

[57] SETHI, S. P. (1974): Quantitative Guidelines for Communicable Disease Control Program: A Complete Synthesis, *Biometrics*, 30(4), 681-691.

[58] SETHI, S. P. (1977): Nearest Feasible Paths in Optimal Control Problems: Theory, Examples, and Counterexamples, *Journal of Optimization Theory and Applications*, 23(4), 563-579.

[59] SETHI, S. P. (1978): Optimal Quarantine Programmes for Controlling an Epidemic Spread, *Journal of the Operational Research Society*, 29(3), 265-268.

[60] SETHI, S. P., P. W. STAATS (1978): Optimal Control of Some Simple Deterministic Epidemic Models, *Journal of the Operational Research Society*, 29(2), 129-136.

[61] STASHKO, A. (2012): The Effects of Prevention and Treatment Interventions in a Microeconomic Model of HIV Transmission, *mimeo*.

[62] SZEKERES, G., T. J. COATES, S. FROST, A. LEIBOWITZ AND S. SHOPTAW (2004): Anticipating the Efficacy of HIV Pre-Exposure Prophylaxis (PrEP) and the Needs of At-Risk Californians, *Center for HIV Identification, Prevention, and Treatment Services (CHIPTS)*.

[63] TOXVAERD, F. (2009a): Recurrent Infection and Externalities in Treatment, *mimeo*.

[64] TOXVAERD, F. (2009b): Infection, Acquired Immunity and Externalities in Treatment, *mimeo*.

[65] UNAIDS (2008): Report on the Global AIDS Epidemic.

[66] UNAIDS (2010): Getting To Zero, 2011–2015 Strategy, available at www.unaids.org/sites/default/files/.../JC2034_UNAIDS_Strategy_en.pdf.

[67] WAGENER, F. O. O. (2003): Skiba Points and Heteroclinic Bifurcations, with Applications to the Shallow Lake System, *Journal of Economic Dynamics and Control*, 27(9), 1533-1561.

[68] WERBLE, C. (2012): FDA De-Risks "Risk Compensation" Argument in Review of Truvada Prophylaxis Claim, *Elsevier Business Intelligence*, Friday, May 11, 2012.

[69] WILDE, G. J. S. (1982): The Theory of Risk Homeostasis: Implications for Safety and Health, *Risk Analysis*, 2(4), 209-225.

[70] WILSON, N., W. XIONG AND C. MATTSON (2012): Is Sex Like Driving? Risk Compensation Associated with Male Circumcision in Kisumu, Kenya, *mimeo*.