Published in final edited form as:
J Health Econ. 2013 March ; 32(2): 440–451. doi:10.1016/j.jhealeco.2013.01.002.

Economic considerations for social distancing and behavioral based policies during an epidemic

Eli P. Fenichel*
Yale School of Forestry & Environmental Studies, New Haven, CT 06511, United States

Abstract

Public policies intended to induce behavioral change, specifically incentives to reduce interpersonal contacts or to "social distance," increasingly play a prominent role in public disease response strategies as governments plan for and respond to major epidemics. I compare social distancing incentives and outcomes under decentralized, full control social planner, and constrained social planner, without health class specific control, decision making scenarios. Constrained social planner decision making, based on non-health class specific controls, can in some instances make society worse off than decentralized decision making (i.e. no intervention). The oft neglected behavior of recovered and immune individuals is important for welfare and health outcomes.

Keywords

Public health; Infectious disease; Dynamic optimization; SIR model; Adaptive behavior

1. Introduction

Infectious disease epidemics are scary, and behavioral adaptation has been a part of human response to infectious disease for centuries. Recently, the World Health Organization (2006), governments (Stern and Markel, 2009), and public health experts (Ferguson et al., 2005; Glass et al., 2006; Halloran et al., 2008) have emphasized the potential importance of public policies designed to elicit behavioral changes in preparing for and responding to infectious disease epidemics. Specifically, these strategies provide incentives, some quite strong, to reduce interpersonal contacts or to "social distance." Social distancing policies can be in the form of public facility shut downs (e.g., mass transit and school closures), propaganda campaigns, and other attempts to reduce the ordinary contact rate among people. Economics has a clear role to play in mapping incentives through micro-level behaviors to macro-level outcomes for health and measures of social welfare.

If policies are aimed at changing individual behavior, then one must consider the ways in which heterogeneous agents will respond to infection risk (Funk et al., 2010; Gersovitz, 2011). Prior economic studies of infectious disease transmission have developed macro-dynamic models of socially optimal disease control and eradication in humans via treatment and vaccination (Barrett and Hoel, 2007; Boulier et al., 2007; Francis, 2004; Gersovitz and Hammer, 2003, 2004), in animals via culling strategies (reviewed in Horan et al., 2011b, 2010), and in vector-borne diseases (Gersovitz and Hammer, 2005). These models show that individuals under-invest in prevention and control of disease, and that policy intervention

*Tel.: +1 203 432 5114, eli.fenichel@yale.edu.

can be welfare enhancing. Moreover, Almond and Mazumder (2005) and Almond (2006) show that the impacts of an epidemic can persist long after the epidemic fades. Yet, public interventions must be undertaken with care. Smith et al. (2009) and Keogh-Brown et al. (2010) simulate social distancing interventions using a computable general equilibrium model and find that the costs of proposed public health interventions may outweigh the cost of the disease. Chen et al. (2011) show that social distancing interventions may have ambiguous results on endemic disease equilibria. Social distancing and quarantine policies may be "overdone" and decrease welfare (Fenichel et al., 2011; Keogh-Brown et al., 2010; Mesnard and Seabright, 2009).

Epidemiologists and economists increasingly recognize that consumer heterogeneity and micro-level decision making are essential in determining how epidemics evolve and how policy interventions affect this evolution. Economists have developed models to describe how individuals engage in adaptive or strategic behavior, including treatment and vaccination (Francis, 1997; Geoffard and Philipson, 1996; Philipson, 2000), reductions in risky sexual behavior (Auld, 2003; Kremer, 1996), migration behavior (Mesnard and Seabright, 2009), and generic risk mitigation through reducing social contacts (Chen et al., 2011; Fenichel et al., 2011). A key component of models with strategic behavior is that the current health state of an individual influences his incentives to engage in different behaviors (Auld, 2003; Fenichel et al., 2011). This type of adaptive response is largely missing from epidemiological models that express an individual's rate of social contacts as a function of observable attributes that are exogenous to the epidemic such as sex and age (reviewed by Funk et al., 2010).[1]

Susceptible individuals may have incentives to avoid infection for their own wellbeing even if they are not concerned with the overall state of public health. Hence, one can think of individuals as having partial ownership over the state of public health or alternatively the public health state can be modeled as an impure public good (Bell and Gersbach, 2009). However, public health interventions are difficult to design. For example, vaccinations are wasted if they are given to individuals who would not have become infected or have already contracted an infection. Treatment, if available, may be well targeted at infected individuals, but may not prevent latent individuals from spreading infection. These targeting challenges are exacerbated when considering social distancing policies, where the first best policy must account for the mapping between endogenous behavioral responses to the policy and the resulting spread of infection. Policies that abstract from this feedback mechanism could have perverse effects. There is a need for economic analysis, but to have policy impact such analysis must also be grounded in the nearly hundred-year-old epidemiological modeling tradition (Kermack and McKendrick, 1929) to gain traction in non-economic spheres – the spheres that dominate actual policy decision making.

In this paper, I develop an integrated epidemiological-economic model of social distancing and strategic economic behavior during an epidemic. I contrast three types of decision making: decentralized decision making, socially optimal decision making in the sense of maximizing the discounted net present utility of the *ex ante* representative agent (this defination of social welfare follows Chakraborty et al., 2010 and is explain in more detail below; Francis, 1997, 2004; Geoffard and Philipson, 1996), and a constrained social planner who tunes a policy instrument to be as efficient as possible given the inability to target the epidemic itself, specifically, I consider a social planner who constrains all individuals to the same behavior regardless of health class. This third mode of decision making approximates the thinking that drives many real-world social distancing policies (though these are seldom

[1]Recently, mathematical epidemiologists have begun to use game theoretic models to consider the role of adaptive or strategic behavior (Galvani et al., 2007; Reluga, 2010).

optimized), such as school closures, (Cauchemez et al., 2008; Glass et al., 2006), public transit shut downs, and other policies put forth by "frontline" infectious disease epidemiologists (Ferguson et al., 2005; Fraser et al., 2009). I show that targeting social distance policies by health class is important for maximizing social welfare and that it in some cases it is possible for indiscriminant, but optimally tuned, decision making to make society worse off than allowing for decentralized decision making. In other words, seemingly "second-best" interventions that tune a non-targeted control to maximize social welfare, conditional on the constrained choice set, can potentially lead to lower wellbeing than not intervening, suggesting that indiscriminant policies to reduce contacts that are not optimally tuned may not satisfy the creed "do no harm." A relative loss of welfare can potentially occur from non-targeted policies, compared to decentralized decision making, because designing policies for the "average" individual may impose a strong constraint that erodes welfare more than incomplete markets for disease prevention. This result, which is easily overlooked in infectious disease epidemiology, highlights the need for analysis that integrates economic modeling with epidemiological theory, and contributes to the broader literature on commons problems with partial ownership (Stavins, 2011).

2. The model and analysis

2.1. An epidemiological model with behavioral response

Most modern epidemiology builds on the compartmental epidemiological modeling framework introduced by Kermack and McKendrick (1929) and popularized by Anderson and May (1979). Our model follows in this tradition, and closely follows Fenichel et al. (2011). Specifically, I consider a communicable disease that causes significant utility loss to infected individuals, but does not cause mortality.[2] A population that is incurring an epidemic can be divided into health compartments. I restrict the analysis to the three basic epidemiological compartments: susceptible, S, infected and infectious, I (for the purpose of this article I use these terms interchangeably), and recovered with immunity, Z, in a fixed population, N. The epidemiological model is formalized as

$$\dot{S} = -\frac{C(\cdot)\beta S I}{N} \quad (1)$$

$$\dot{I} = \frac{C(\cdot)\beta S I}{N} - vI \quad (2)$$

$$\dot{Z} = vI \quad (3)$$

$C(\cdot)I/N$ is the rate susceptible individuals contact infectious individuals, where $C(\cdot)$ is a contact function. A special case of $C(\cdot)$ is a constant. $C(\cdot)$ is described in detail in the next paragraph. The parameter β represents the likelihood that contact with an infectious individual yields infection, i.e., the conditional "infectiveness" of a pathogen. The rate of recovery and acquired immunity is v, and I assume no loss of immunity. The model is constructed so that N is fixed and that any outbreak is temporary. Therefore, I focus on dynamics as opposed to steady states (see Francis, 2004 for a similar treatment). Individuals within a particular compartment are homogeneous.

[2] Our framework is easily adapted to handle disease-induced mortality, but this requires tracking changes in the total population thereby adding a state variable. It is also possible to include population turn-over, see supplemental material in Fenichel et al. (2011). However, our primary goal is to explore social distancing policies in epidemics such as flu, which are often managed as if they will eventually die out.

Most epidemiological models assume that all individuals in the population behave identically (exceptions include Auld, 2003; Fenichel et al., 2011). However, the basic model structure itself introduces situational heterogeneity among health classes. There is reason to believe that otherwise identical individuals in different health classes face different incentives and therefore behave differently.[3] To relax the assumption of homogeneous behavior, first index individuals by health type, denoting $Y = \{s,i,z\}$ to be the set of possible health types (corresponding to S, I, and Z). Next, define contacts between m-type and n-type individuals, with m, $n \in Y$, as

$$C^{mn}(\cdot)=\frac{C^m C^n N}{S C^s+IC^i+ZC^z} \quad (4)$$

C^m is the expected number of contacts made by a type-m individual. When $m = s$ and $n = i$, $C^{mn}(\cdot) = C^{si}$ and corresponds to $C(\cdot)$ in Eqs. (1) and (2). I emphasize that C^m is a *choice* made by a type-m individual. C^m may be chosen directly or by engaging in certain activities, e.g., taking public transportation. I assume individuals know their own health type, but not the health type of others. Accordingly, Eq. (4) implies conditional proportional mixing. Mixing is proportional, but also conditional on the behaviors and the distribution of individuals of different health types. In what follows, I simplify notation by scaling N to unity so that S, I, and Z are fractions of the total population.

If all types choose the same number of contacts $C^h = c$, $\forall h \in Y$, irrespective of health class, then $C^{mn}(\cdot)$ simplifies to the constant c. Accordingly, transmission takes the classic form $c\beta SI$. Furthermore, if behavior is assumed to be constant over time, then c is a parameter, and β and c can be combined into $\hat{\beta}$ so that transmission takes the common form $\hat{\beta}SI$ (Begon et al., 2002; McCallum et al., 2001). It is not possible to model endogenous social distancing with such a formulation. At a minimum, transmission must be modeled as $c\beta SI$ to investigate the role of behavior on the transmission process.

I follow Geoffard and Philipson (1996) and model a representative agent whose current-period utility depends on his current health state, $h \in Y$, and current-period contacts with others. Specifically, a type-h individual's instantaneous utility is $u(h,C^h)$. Instantaneous utility is assumed to be a concave and single peaked function in contacts, and infection reduces instantaneous utility, so that $u(i,c) < u(s,c)$ and $u(i,c) < u(z,c)$.[4]

2.2. Individual decentralized decisions

Prior work has examined individual decision making in the social distancing context (Chen et al., 2011; Fenichel et al., 2011; Reluga, 2010). These prior papers focus on individual incentives and, with the exception of Fenichel et al. (2011), focus on equilibrium outcomes, but transient dynamics are important during an epidemic (Francis, 2004).

If agents act in their own best interest and ignore externalities associated with contacts, then each individual type behaves as if he solves a dynamic problem formalized by the Hamilton–Jacobi–Bellman equation

$$\rho V(h)=\max_{C^h}\left\{u(h,C^h)+\sum\nolimits_Y\left[P^{h,j}(C^h;S,I,Z,C^{-h},\overline{C}^h)\times(V(j)-V(h))\right]\right\} \quad (5)$$

[3] Additional compartments or distributed parameters (Veliov, 2005) are required to model heterogeneous behaviors (and hence heterogeneous infection risks) within a health class.

[4] We may also expect that $u(z,c) \le u(s,c)$ suggesting the possibility of lasting effects from infection.

where ρ is the discount rate. $P^{h,j}$ is the probability of transition from state $h \in Y$ to state $j \in Y$ conditional on choice C^h at a given instant. This probability may be derived from Eqs. (1)–(3) (see Fenichel et al., 2011 for details). The probability $P^{h,j}$ depends on the current state of the system, behaviors of individuals in the other health classes, C^{-h}, and behavior of individuals in the own health class, $\bar{C}^h$, which the individual takes as given. I focus on the case when $P^{h,j} = 0$ for $h \neq j$, except for the basic epidemiological transitions of $P^{s,i}$ and $P^{i,z}$, and where $P^{i,z}(C^i) = \bar{P}^{i,z} \forall C^h$. This structure implies that the recovery rate is invariant to behaviors in the population, it is not possible to go from s to z directly, or z to s, and $P^{z,z} = 1$.

The solution to problem (5) depends on the individual's current state. Consider the problem for recovered, type-z, individuals. Recovered individuals' first order condition is $u_{C^z}(z, C^z) = 0$ because there is no dynamic effect of type-z's decision, where the subscript denotes the partial derivative. The concave and single-peaked nature of $u(h,C^z)$ implies that C^z is constant, finite, and positive valued. The assumption that $P^{i,z}(C^i) = \bar{P}^{i,z} \forall C^h$ implies that type-i individuals have a similar result, $u_{C^i}(i, C^i) = 0$, implying that that C^i is constant, finite, and positive valued. Under a decentralized decision making scheme, infected and recovered individuals do not behave strategically.

Susceptible individuals make forward-looking decisions, which for an interior solution are modeled by satisfying

$$u_{C^s}(s, C^s) - P^{si}_{C^s}(\cdot)[V(s) - V(i)] = 0. \quad (6)$$

Re-ordering Eq. (6), $\frac{u_{C^s}(s,C^s)}{P^{si}_{C^s}(\cdot)} + V(i) = V(s)$, shows that the value in the infected state must be less than the value in the susceptible state, which provides an incentive for behavioral change. Moreover, the infection itself must result in loss (or gain) of utility for the individual to consider the effect of present decisions on future health. Such incentives, or lack thereof, for behavioral adjustments may be particularly important when there are different classes of individuals with exogenous risk factors (e.g., sex or age).

Combining Eq. (6) with the version of Eq. (5) associated with $h = s$ and letting $u^{dc}(s,C^s)$ equal the optimized value of utility (5), and calling the derivative of utility evaluated at the optimal contact level $u^{dc}_{C^s}(s, C^s)$, yields

$$\rho V(s) = (1 - E_R^{-1}) u^{dc}(s, C^s) \quad \text{where} \quad E_R = \frac{u^{dc}(s, C^s) P^{si}_{C^s}(\cdot)}{u^{dc}_{C^s}(s, C^s) P^{s,i}} \geq 0. \quad (7)$$

Eq. (7) is an asset value equation (Shapiro and Stiglitz, 1984) related to human health capital held in the susceptible state. The term $-E_R$ in Eq. (7) is the elasticity of demand for infection risk (the percent change in the optimal probability of infection divided by the percent change in the optimal utility payoff), where prices are thought of in utility terms (Weitzman, 2001). Eq. (7) extends the notion of "prevalence-elasticity" (Philipson, 2000) to contact behavior, self-determined social distancing, and individual behavioral decisions. From the terms that make up E_R it must be that $E_R > 0$, and for Eq. (7) to remain positive $E_R > 1$. This implies that individuals will require disproportionally large amounts of risk reduction to give up utility.

The core message from Eq. (7) is that the individual only considers infection risk to himself and his own utility. The individual disregards how limiting contacts influences the common pool of public health and may benefit others over the course of the epidemic. This realization has led public health researchers to consider the role of “social distancing” policies. However an important question is how do behaviors of others affect the individual’s own incentives.

Proposition 1—Changes in the behavior of recovered individuals have greater effect on the susceptible individual’s value function than changes in the behavior of infected individual, all else equal.

It is possible to compare $\partial V(s)/\partial C^i$ to $\partial V(s)/\partial C^z$ by applying the dynamic envelop theorem (Caputo, 2005) to Eq. (5), where C^i and C^z are taken as parameters (as would be the case from the susceptible individual’s perspective), $\partial V(s)/\partial C^h = P^{si}_{C^h}(C^s; S, I, Z, \overline{C}^s, C^i, C^z)(V(i)-V(s))$, $h \in \{i, z\}$. The sign of $\partial V(s)/\partial C^h$ is determined by the marginal effect of C^h on Eq. (4). C^i enters both the numerator and the denominator of Eq. (4). A marginal increase in C^i has a positive effect on the numerator, which necessarily exceeds its negative effect on the denominator. C^z only has the negative effect associated with the denominator. All else equal $-\partial V(s)/\partial C^i < \partial V(s)/\partial C^z$ because C^i has offsetting effects on $V(s)$.

Proposition 1 suggests that, all else equal, policy effects on recovered individuals are important. Yet, this class of individuals is often given little consideration in health policy discussions.

2.3. The complete market and the social planner’s problem

A public social distancing policy may improve social welfare. The social planner’s problem, which is isomorphic to the case where all individuals consider the effects of their behavioral decisions on all other individuals – the case of a complete market for the state of public health, provides a benchmark for the maximum value that the system can deliver once the public health commons problem is resolved (Bell and Gersbach, 2009).[5] Individuals are assumed to be identical prior to the introduction of the disease. Therefore, I take the *ex ante* net present utility to be the social welfare function. This is consistent with the prior literature that has analyzed the economics of epidemics in a dynamic setting (Chakraborty et al., 2010; Francis, 1997, 2004; Geoffard and Philipson, 1996), and is consistent with the notion that economic decisions are based on *ex ante* calculations (Heckman, 2010). The time specific, socially optimal levels of contacts that complete the market for public health solves

$$\max_{C^s, C^i, C^z, T} \left\{ \int_{t=0}^{T} e^{-\rho t}[Zu(z, C^z) + S\,u(s, C^s) + Iu(i, C^i)]dt + \left(\frac{[Z(T)u(z, C^z) + S(T)u(s, C^s(T)) + I(T)u(i, C^i(T))]e^{-\rho T}}{\rho} \right) \right\} \quad (8)$$

subject to Eqs. (1)–(4), $I(0) > 0$, $S(0) < 1$, $0 \le Z(0) < 1$, where T is the time at which the epidemic fades out in an economic sense, which is defined explicitly in the next paragraph. Only two of the three Eqs. (1)–(3) are needed to solve the problem. It is useful to focus on Eqs. (1) and (3), where $I = 1 - (S + Z)$. A social planner manages the three groups of individuals and solves for the least restrictive program, choosing the contact levels for susceptible, infected, and recovered individuals. The social planner directly chooses behaviors, which is common in social planner problems for bioeconomic systems (Clark, 2005) and implies that the social planner has sufficient policy latitude to provide behavioral

[5] In the centenary issue of the American Economic Review, Stavins (2011) argues that such commons problems are fundamental to economics broadly.

incentives with no transaction costs and in a way that any income effects can be offset, hence the interpretation of a complete market for public health.

If susceptible individuals continue to become infected, then all individuals eventually enter the recovered compartment. Economic epidemic fadeout, which occurs at time T, is defined as a limit and is the first instance, following the epidemic, at which C^s is sufficiently similar to C^z so that susceptible and recovered individuals are nearly behaviorally equivalent.[6] If I is never sufficiently small, then all susceptible individuals become infected and eventually recovered, which would be a contradiction.[7]

The current value Hamiltonian for problem (8) is

$$H = Zu(z, C^z) + S\,u(s, C^s) + Iu(i, C^i) + \lambda^s\,\dot{S}\; + \lambda^z\,\dot{Z}$$

where λ^h is the co-state variable associated with the epidemic state expressed as a superscripted. The first order conditions for problem (8) are

$$\frac{\partial H}{\partial C^s} = u_{C^s}(s, C^s) + \frac{1}{s}\frac{\lambda^s C^i \beta S I}{S C^s + IC^i + ZC^z}\left(\frac{S C^s}{S C^s + IC^i + ZC^z} - 1\right) \le 0 \quad (9)$$

$$\frac{\partial H}{\partial C^i} = u_{C^i}(i, C^i) + \frac{1}{I}\frac{\lambda^s C^i \beta S I}{S C^s + IC^i + ZC^z}\left(\frac{IC^i}{S C^s + IC^i + ZC^z} - 1\right) \le 0 \quad (10)$$

$$\frac{\partial H}{\partial C^z} = Z\left(u_{C^z}(z, C^z) + \frac{\lambda^s C^i C^s \beta S I}{(S C^s + IC^i + ZC^z)^2}\right) \le 0 \quad (11)$$

Conditions (9)–(11) describe the socially optimal behavior of each group. The conditions will be negative if it is optimal to make zero contacts, given $C^h \ge 0$. Otherwise, contacts will be chosen to make the conditions hold as strict equalities.

First, consider the socially optimal contact level for recovered, z, individuals. These individuals are immune from infection so they have no personal incentive to reduce contacts as noted in type z's solution to problem (5). Assume $Z > 0$. Condition (11) implies that the socially optimal behavior of recovered individuals differs from the recovered individual's private incentive: $u_{C^z}(z, C^z) + (\lambda^s C^i C^s \beta SI/(SC^s + IC^i + ZC^z)^2) = 0$ versus $\mu^{dc}_{C^z}(z, C^z) = 0$, respectively. The second term exists because a social planner considers how recovered individuals' behaviors influence the number of contacts the average individual (averaged across all health classes) in the population makes, described by the denominator in Eq. (4). The second term in Eq. (11), within the parenthesis, is positive and strictly positive if S and I

[6] In a continuous state, continuous time model I can only reach zero as t goes to infinity. The reality of a discrete and finite N will lead to demographic stochasticity, a stochastic effect resulting from the fact that a fraction of an individual cannot be infected, and disease fades out in finite time (Nasell, 2002). Nevertheless, the continuous time system has a limiting behavior (Hethcote, 2000). In all continuous time models equilibria are only approached as $t \to \infty$. Therefore, our definition of economic epidemiological fadeout is an alternative way of defining the limit. Nevertheless, the behavior of the model is well approximated by our construction, which is similar to disease eradication models in Barrett and Hoel (2007) and Horan et al. (2011b), and is useful for providing information about optimal establishment of an infectious disease and the limiting behavior of the optimized dynamical system.

[7] That the disease optimally fades out is a result of the constant population assumption with no mortality and no new susceptible individuals.

both exist and types s and i make contacts. Given the assumption that u is concave and single-peaked, to satisfy condition (11) recovered individuals must make contacts in *excess* of what they would have made acting in their own self-interest in the decentralized problem such that $u_{C^z} < 0$. This occurs because recovered individuals increase the probability of a "safe" contact for a susceptible individual by making the average number of contacts greater, but not increasing the number of infectious people to contact. Recovered individuals can also be seen as "soaking up" contacts made by infected individuals thereby protecting susceptible individuals and enabling susceptible individuals to make more contacts while holding disease risk constant.

Condition (10) illustrates the incentives that the social planner faces for determining the contact level of infected individuals. Assume contacts are an essential good or factor in producing utility, otherwise condition (10) could be negative, and no level of contacts can be chosen to set condition (10) equal to zero, making it optimal to ban all contacts by infected individuals – a quarantine policy.[8] Under a quarantine policy the incentives for susceptible and recovered individuals become the myopic incentives and susceptible and recovered classes do not need to consider the future. More generally, contacts may be essential for all classes, and in the following analysis I assume that this is the case. Nevertheless, it will generally be socially optimal for infected individuals to reduce contacts relative to their self-interested behavior.

Self-interested susceptible individuals are forward looking, and their incentives are illustrated by condition (6). The socially optimal behavior of susceptible individuals is also forward-looking and illustrated by condition (9). The term $\mu^s = V(S) - V(I)$ in Eq. (6) can be thought of the shadow value for susceptible health from the decentralized decision making point of view. The term $(1/S)(C^i\beta SI/(SC^s + IC^i + ZC^z))(SC^s/(SC^s + IC^i + ZC^z) - 1) = -F_{C^s}$ in Eq. (9) is the marginal effect of susceptible contacts on the infection rate. For Eq. (6) and Eq. (9) to provide the same incentives for susceptible individuals

$$\mu^s P^{si}_{C^s}(\cdot)=\lambda^s F_{C^s} \quad (12)$$

Combining Eqs. (9)–(11) and evaluating all variables along their optimal paths implies that

$$\lambda^s=-\frac{u_{C^z}(z, C^z)w}{C^{si}\beta S I}=u_{C^z}(z, C^z)w\dot{S}^{-1} \quad (13)$$

$$u_{C^s}(s, C^s)=\frac{u_{C^z}(z, C^z)(S C^s-w)}{S C^s} \quad (14)$$

$$u_{C^i}(i, C^i)=\frac{u_{C^z}(z, C^z)(IC^i-w)}{IC^i} \quad (15)$$

where $w = SC^s + IC^i + ZC^z$ is the total number of contacts made or because S, I, and Z are fractions of the population, the number of contacts made by the average individual. Eqs. (13)–(15) frame the optimal shadow value and optimal marginal utilities for susceptible and infected individuals in terms of the optimal marginal utility of recovered individuals, which is motivated by the backwards recursive nature of the social planner's problem. Eq. (13)

[8] This follows from the Kuhn–Tucker–Karsh conditions associated with the problem.

makes clear that if recovered individuals optimally act myopically (i.e., $u_{C^z} = 0$), then so should susceptible individuals, which means that infected individuals must make zero contacts or disease is not present. Assuming disease is present and infected individuals optimal make contacts, then Eq. (13) illustrates that $u_{C^z}(z, C^z) < 0$ for there to be value in remaining susceptible. Moreover, Eq. (13) states that the shadow value of remaining susceptible is the total value if all individuals were recovered, conditional on the optimal behavior for the current state of the population, per the rate at which susceptible individuals are lost to infection. Eq. (12) says that the average individual's incentives must be the same as the susceptible population's incentive averaged over susceptible individuals. However, $P^{si}_{C^s}(\cdot) \neq F_c$. This is because a decentralized decision maker does not consider how his decision affects the pool of available contacts, the denominator terms, but takes them as given, while the social planner considers these aggregate effects. Considering aggregate effects leads to the first term in the parenthesis of Eq. (9), which is positive. If $\mu^s = \lambda^s$, then susceptible individuals would have to make more contacts to reduce the marginal utility of contacts in Eq. (9) relative to the level chosen by decentralized susceptible decision makers. Substituting Eq. (13) into Eq. (12) suggests that $\mu^s < \lambda^s$; when making decentralized decisions susceptible individuals observe the "wrong price" of remaining susceptible as they fail to account for limiting their ability to infect others, much like in the case of a vaccination externality (Gersovitz and Hammer, 2004).

Combining Eqs. (14) and (15) provides an expression that can be thought of as the social planner's marginal rate of substitution, MRS, between contacts made by susceptible and infected individuals along the optimal path, $\partial C^i/\partial C^s = (SC^s - w)IC^i/(IC^i - w)SC^s$. This MRS expression is positive, implying that if it is optimal for susceptible individuals to increase contacts along the optimal path it is also optimal for infected individuals to increase contacts along the optimal path. Eq. (10) states that it is optimal for infected individuals to reduce contacts relative to the case of decentralized decision making, and this MRS expression implies susceptible individuals must also optimally reduce contacts relative to the decentralized decision making case.

The shadow value of susceptible individuals depends on whether decisions are made according to a decentralized or socially optimal program. A socially optimal program requires that the adjoint conditions

$$\dot{\lambda}^s = \rho\lambda^s - \frac{\partial H}{\partial S} = \rho\lambda^s - u(s, C^s) + u(i, C^i) - \lambda^s\beta(C^{si}(I-S) + C^{si}_s S I) + \lambda^z v \quad (16)$$

$$\dot{\lambda}^z = \rho\lambda^z - \frac{\partial H}{\partial S} = \rho\lambda^z - u(z, C^z) + u(i, C^i) - \lambda^s\beta S(C^{si}_Z I - C^{si}) + \lambda^z v \quad (17)$$

are satisfied. These may be re-written as golden rule equations

$$\rho = \frac{\dot{\lambda}^s}{\lambda^s} + \frac{u(s, C^s) - u(i, C^i)}{\lambda^s} + \beta C^{si}(I-S) - \beta C^{si}_s S I - \frac{\lambda^z}{\lambda^s} v \quad (18)$$

$$\rho = \frac{\dot{\lambda}^z}{\lambda^z} + \frac{u(z, C^z) - u(i, C^i)}{\lambda^z} + \frac{\lambda^s}{\lambda^z}\beta S(C^{si}_Z I - C^{si}) - v \quad (19)$$

The LHS of Eq. (18) is the discount rate. This is the opportunity cost of protecting susceptible individuals. The RHS of Eq. (18) is the rate of return from protecting susceptible individuals. It comprises five terms. The first is a net capital gains term associated with preserving susceptible human capital that goes to zero as the epidemic wanes. The second term is the relative marginal benefit of a larger stock of susceptible individuals for the representative agent's utility. The third RHS term is the marginal cost of having more susceptible individuals to protect, assuming that $S > I$. The fourth and fifth terms together are the forgone marginal utility associated with the opportunity to be recovered, which requires infection.

The LHS of Eq. (19) is the discount rate, which, as in Eq. (18), is the opportunity cost of protecting susceptible individuals. However, if the epidemic leads to full recovery, this is isomorphic to marginal cost of preventing infection that ultimately yields recovery and immunity. The RHS of Eq. (19) is the marginal benefit of recovery. The first term is a net capital gains term associated with immune human capital that goes to zero as the disease fades out. The second RHS term is the relative marginal value of recovered individuals compared to infected individuals. The third RHS term is the marginal cost of avoiding infection and ultimate immunity. The final RHS term is the recovery rate. An increase in the recovery rate effectively increases the discount rate in Eq. (19), because an increase in the recovery rate reduces the cost of infection and decreases the marginal cost of myopic behavior. In comparing Eq. (18) with Eq. (19) the factor λ^z/λ^s places a wedge between the opportunity cost of protecting susceptible individuals and the marginal benefits from increasing recovered individuals.

Combining Eq. (16), and Eq. (17) yields

$$\lambda^z = \frac{u(z, C^z) - u(s, C^s)}{\rho} + \lambda^s \left(1 - \frac{\beta I}{\rho}(C^{si} + S\,C^{si}_S - S\,C^{si}_Z)\right) - \frac{\dot{\lambda}^s - \dot{\lambda}^z}{\rho} \quad (20)$$

Eq. (20) provides the relationship between the optimal values of λ^s and λ^z, and the optimal value of λ^s can be taken from Eq. (13). The first RHS term in Eq. (20) is the cost associated with permanent effects from infection.[9] The second term is the shadow value of susceptible individuals adjusted downward by the positive vaccination externality associated with recovered individuals. The second term in the parenthesis is positive and could be greater than 1. Assuming the dynamic adjustment term is not too large, $\lambda^z \leq \lambda^s$ and λ^z can be potentially less than 1 at some point in the epidemic. $\lambda^z \leq \lambda^s$ must be the case otherwise it would be optimal to speed up the epidemic and generate more infection. An alternative interpretation is that if a social planner would discount the susceptible human capital at a lower adjusted discount rate, then the social planner would discount recovered human capital using a recovery-adjusted discount rate. This may in part explain why the role of recovered individuals' behaviors has received less attention.

Models with constant behavior yield an exponential decline in cases, which leads to a positive limit for S and Z as time goes to infinity (Hethcote, 2000). The limiting values of S and Z are sensitive to initial conditions. Fig. 1 shows sample trajectories for S and Z in a system where behavior is held constant at the contact rate in the absence of infectious disease. This would only be the case when $\rho = \infty$.[10] If the $\rho < \infty$, then specification of the utility function matters for determining the final epidemic size. The limiting values of $S(T)$ and $Z(T)$ must lie on the hypotenuse of the triangle connecting (0,1) and (1,0) in S-Z space

[9] A small, but non-zero, mortality rate associated with disease could be interpreted as a lasting utility effect for the representative recovered individual. This is reasonable so long as mortality has negligible effects on the total size of the population, N. This does not imply negligible utility effects.

[10] I assume the social planner and representative agent share a common discount rate and the discount rate is not affected by health status.

in the segment between no behavioral change trajectory ($\rho = \infty$) and a vertical line from the initial condition, which represents the case when Eq. (10) optimally holds as an inequality $\forall t$ (Fig. 2). To determine the optimal trajectory, the social planner must consider the transversality conditions to determine the optimal time T and the speed at which the epidemic should fade out.

Given the known limiting behavior of the dynamic system, the relevant transversality condition is approximated as

$$M(T)-(u(z,C^z(T))Z(T)+u(s,C^s(T))S(T))e^{-\rho T}=0 \quad (21)$$

where M indicates the maximized current value Hamiltonian, where the second term comes from taking the derivative of the scape value function with respect to T (Caputo, 2005).

Proposition 2—It is only optimal to allow an infected individual to keep mixing in the population if the infected individual can compensate all other individuals to incur utility losses associated with defensive behavior.

Given the limiting behavior of an epidemic $\lim_{t\to T} \dot{S} \to 0$ *and* $\lim_{t\to T} \dot{Z} \to 0$ (Hethcote, 2000). If ε is arbitrarily small, then, the maximized Hamiltonian, Eq. (20), is approximately

$$u(z,C^z(T-\varepsilon))Z(T-\varepsilon)+u(s,C^s(T-\varepsilon))S(T-\varepsilon)+u(i,C^i(T-\in))I(T-\varepsilon) \approx (u(s,C^s(T))S(T)+u(z,C^z(T))Z(T))e^{\rho\varepsilon} \quad (22)$$

Eq. (22) implies a compensation criterion among health classes. In the instant prior to the optimal fadeout, infected individuals must gain enough utility from the first contact to be able to compensate susceptible and recovered individuals, in utility terms, for engaging in defensive behaviors.[11] If infected individuals cannot compensate the individuals in the other two classes, then a corner solution, a quarantine policy, is optimal. Moreover, Eq. (22) also suggests that infected individuals should only make contacts up to the point at which they can compensate other health classes for the marginal increase in infection risk. The social planner chooses the rate at which fadeout is approached. Quarantine of the first infected individual optimally would result in fadeout if T were optimally $\leq 1/\nu$. If $1/\nu$ is a short enough duration, then by induction the limiting case applies to the first infected individual.

A longer infection period, or a greater opportunity cost associated with infected individuals forgoing contacts, leads to a greater chance for a disease to optimally enter a population. In practice, a non-trivial portion of individuals are likely to be infected before an intervention takes place. Yet, the same compensation principle holds. More generally, the compensation principle implied by Proposition 2 will inform the extent to which infected individuals reduce contacts to protect susceptible individuals. It is an empirical question that requires knowing the nature of individual preference for contacts and health as to whether quarantine would be socially optimal.

2.4. The constrained social planner and non-targeted policies

Many proposed public social distancing policies (e.g., facility closures, increases in public transit rates) are not sufficiently flexible to provide targeted incentives across health classes. Rather, most practical public policies to encourage social distancing are blunt and provide

[11] The socially optimal defensive behavior for recovered and immune individuals is to increase the number of contacts.

incentives for all individuals to reduce contacts. An upper bound on the welfare that a blunt social distancing policy provides is

$$\max_{c,T}\int_{t=0}^{T} e^{-\rho t}[Zu(z,c)+S\,u(s,c)+Iu(i,c)]dt+\frac{[Z(T)u(z,c)+S\,(T)u(s,c(T))+I(T)u(i,c(T))]e^{-\rho T}}{\rho} \quad (23)$$

subject to disease dynamics Eqs. (1)–(3). Problem (23) assumes no enforcement costs or deadweight loss from government intervention. Problem (23) states that a constrained social planner chooses an untargeted, homogeneous level of contacts across all individual types to maximize the utility of the representative agent. Indeed, behavior in epidemics is commonly modeled as homogeneous (Begon et al., 2002; Hethcote, 2000). There are important issues with modeling behavior as homogeneous. First, the number of state variables exceeds the number of control variables and not all states are uniquely controllable. In practice, this may be because of institutional constraints associated with the feasible policy set of public health interventions do not fully control the system, because full controllability requires an equal number of control variables as binding constraints (Caputo, 2005; Horan et al., 2011a).[12] Second, in practice social distancing programs likely provide constraints to reduce contacts relative to the individuals' self-interested behaviors. It is not certain that such constraints will be binding on all classes of individual. This could be of particular concern if, under decentralized decision making, an infection itself makes contacts less desirable for infected individuals, in which case the program may work to reduce susceptible and recovered individuals' contacts. This may be the exact wrong incentive given that recovered individuals generate positive externalities through contacts and susceptible individuals have a strong private incentive to manage contacts to protect themselves.

The properties of the adjoint and transversality conditions associated with problem (23) remain the same as the social planner problem with full control, but the optimal value of the shadow value for the susceptible population under the optimal, constrained social planner program differs from its value under the unconstrained social planner program. The current value Hamiltonian for problem (23), indicated by H^{pc}, is the same as for problem (8) with λ replaced by φ to allow the shadow values to follow different time paths. There is only one first-order condition

$$\frac{\partial H^{pc}}{\partial c}=Zu_c(z,c)+S\,u_c(s,c)+Iu_c(i,c)-\varphi^s\beta S\,I=0 \quad (24)$$

if the social distancing program is binding on all individual types. Contact behavior does not enter the marginal effect on the infection rate; eliminating heterogeneity in contact behavior makes the marginal infection rate linear in contacts. Eq. (24) says the marginal utility to the representative agent per rate of infection per contact must equal the shadow value associated with the susceptible population.

When the policy maker is constrained, the shadow value associated with the susceptible population is

$$\varphi^s=\frac{c(Zu_c(z,c)+S\,u_c(s,c)+Iu_c(i,c))}{c\beta S\,I} \quad (25)$$

[12] A implication of the rank constraint condition (Caputo, 2005 pp. 150–151) is that "the number of binding constraints cannot be greater than the number of control variables."

In order for the marginal value of an increase in the susceptible fraction to be equal under management by the social planner an institutionally constrained social planner Eq. (25) would have to equal Eq. (13), which is generally not the case. Setting (13) equal to (25) and canceling like terms implies

$$-\frac{u_{C^z}(z,C^z)w}{C^{si}}=\frac{(Zu_c(z,c))+S\,u_c(s,c)+Iu_c(i,c))c}{c}.$$

If all individuals behave identically, then $C^{si} = w$. For λ^s to equal φ^s, the representative agent's marginal utility in the untargeted system would have to equal to the negative of the marginal utility of recovered individuals from the social planner problem. There is no reason that this should be true. Furthermore, socially optimal heterogeneous behavior is needed to complete the market in the unconstrained social planner problem, which implies that $(w/C^{si}) > 1$. If the utility function were symmetric in contacts, then the marginal utility forgone by the representative recovered agent in the complete market would have to be less than marginal utility gained by the population level representative agent in the constrained social planner setting. This implies that in the untargeted setting that all individuals would have to reduce contacts by more than the representative recovered agent increases contacts in the social planner problem (assuming a symmetrical utility function). Assume that $u(s, C^h) = u(z, C^h)$ for a given value of C^h. In this case, susceptible and recovered individuals would behave identically if they did not consider the future. The terms containing the λ^s in Eqs. (9) and (11) define how much the individuals should adjust behavior when they fully consider the future. This marginal user cost term, the term containing λ^s, is greater in absolute value in the case of recovered individuals than susceptible individuals, Eqs. (9) and (11). If the utility function is symmetrical in contacts, the recovered individuals optimally adjust behavior away from the myopic case more than susceptible individuals. Therefore, under the untargeted system susceptible individuals would be forced to make fewer contacts relative to the full control social planner program. The same is true for infected individuals under similar assumptions. This occurs because by requiring recovered individuals to make more contacts under the full control social planner program, infected and susceptible individuals can make more contacts without changing the infection rate. There is a loss of utility under the untargeted system and an increase in disease cases over the epidemic relative to the unconstrained social planner.

It is not surprising that if the social planner's choice set is constrained, then there is a loss of utility relative to a less constrained choice set, even though the constrained social planner's partially controlled solution maximizes utility conditional on the constrained choice set. Nevertheless, there is a need to consider policy settings when the policy maker faces such constrains, which may be thought of as institutional constraints (Dasgupta and Maler, 2003; Horan et al., 2011a). Such policies may be thought of as "second-best" in the sense that they seemingly achieve a socially optimal solution in the presences of existing distortions (Lipsey and Lancaster, 1956) resulting from institutional constraints. But, it is not clear that such "second-best" policy choices are better than decentralized decision making when the decentralized path of the choices is not nested within the constrained choice set.

Proposition 3—It is possible that the seemingly "second-best" program associated with a constrained social planner with partial controllability does not dominate the representative agent's utility that is associated with decision paths resulting from decentralized decision making.

In order for the constrained social planner program to clearly and always dominate decentralized decision making, the optimized version of the current value expression, i.e., H^{pc} must always be greater than the optimized current value expression for decentralized decision making.[13]

$$Zu(z,c)+S\,u(s,c)+Iu(i,c)+\varphi^{s}\,\dot{S}\,+\varphi^{z}\,\dot{Z}>Zu^{dc}(z,C^{z})+Iu^{dc}(i,C^{i})+S\,(u^{dc}(s,C^{s})-\mu^{s}P^{s,i}(C^{s},\cdot))\quad(26)$$

which implies

$$(S\,u(s,c)+\varphi^{s}\,\dot{S}\,+\varphi^{z}\,\dot{Z})-S\,(u^{dc}(s,C^{s})-\mu^{s}P^{s,i}(C^{s},\cdot)>Z(u^{dc}(z,C^{z})-u(z,c))+I(u^{dc}(i,C^{i})-u(i,c))\quad(26')$$

The LHS of inequality (26) is the current value Hamiltonian, H^{pc}, – a welfare measure (Brock and Xepapadeas, 2003; Weitzman, 1976) – for the constrained social planner problem. The RHS of inequality (26) is the equivalent current value expression for the society of decentralized decision makers. Assume that c and $C^{h}\forall h$ in Eqs. (26) and (26′) are chosen to optimize their respective decision problems conditional on an arbitrary starting value {S, I, Z}, at which the planner could consider intervening (e.g., once the planner is aware of the epidemic). The RHS of (26′) must be positive because the instantaneous utility payoffs for recovered and infected individuals must be greater under decentralized decision making than under any alternative program. Eq. (26) may not hold if the sign of the LHS of (26′) is ambiguous or negative, though this is not necessary for (26) to fail. The first LHS term of (26′), the current value Hamiltonian for the constrained social planner less the current period benefits to infected and recovered individuals, is ambiguous. Indeed, Eq. (21) suggests that this term could be negative; $\dot{S}<0$, making the second term in the first set of parenthesis negative. Moreover, as long as prevalence is increasing, which would be the case early in an epidemic, $\dot{Z}\le-\dot{S}$, and as already noted $\varphi^{z}<\varphi^{s}$ otherwise it would be optimal to accelerate the epidemic. So, the sum of the second two terms in the first set of parenthesis in (26′) must be negative over some periods of the epidemic. It is hardly certain that the current period payoff to susceptible individuals alone from contacts offsets the future value terms, the second two terms in the first set of parenthesis. The second LHS term in (26′) is the current value expression for the decentralized decision or the society of decentralized decision makers less the payoffs to infected and susceptible individuals. Generally, the terms in the second set of parenthesis on the LHS of (26′) could be positive or negative, though one would expect that given the maximizing behavior of susceptible individuals that the sum of these terms could be positive, making the whole term negative. The sign of the LHS of (26′) is ambiguous, and it is not possible to claim that a constrained social planner applying an optimized but non-target policy unambiguously yields a greater welfare than allowing for decentralized decision making.

It is possible, but not necessarily true, that the system managed through a non-targeted public policy lowers welfare relative to the decentralized management and no public intervention. The reason for this is that the non-targeted system may actually be more constrained than the decentralized system. In the decentralized system, individuals fail to account for how their actions affect the welfare of others: recovered individuals undersupply contacts and infected and susceptible individuals oversupply contacts relative to the socially optimal program. However, in the non-targeted program individuals are forced to disregard how their behavior affects their own wellbeing. By homogenizing behavior and not allowing individuals to behave in a way appropriate to their health status, recovered individuals

[13] The comparison of alternative forms of the current value Hamiltonian to assess alternative management programs follows Rondeau (2001).

drastically undersupply contacts relative to the decentralized case. Indeed, this result is only clear after considering the recovered individuals' behavior. An important difference between this constrained social planner problem and most dynamic problems, even those with a constrained social planner, is that a control function that mimics decentralized behavior is not nested within the constrained social planner's choice set.

3. Numerical illustration

In order to show how different decision frameworks lead to different epidemiological and welfare outcomes, I develop a numerical illustration based on decision making under decentralized decision making, a social planner, and a constrained social planner. Furthermore, I examine adaptive and rational expectations models for the case of decentralized decision making, and explore the case when a social planner is constrained to provide a minimum number of contacts but is otherwise free to target by health class.

To provide numerical examples it is necessary to specify a utility function and economic and epidemiological parameters. I specify the utility function based on the constant elasticity of substitution (CES) form as $u(h, C^h) = (a(C^h)^\gamma + (1 - a)(b - C^h)^\gamma)^{1/\gamma} - m^h$. The parameter b is a numeraire "good" that must be given up to make contacts. For example, if time in public requires contacts at a fixed rate and time in private is solitary and yields no contacts, then b is the time budget.[14] The parameter $\gamma = (\sigma - 1)/\sigma$, where σ is the elasticity of substitution between contacts and the numeraire good, and a is a share parameter. Finally, the parameter m^h is a daily lump sum cost of being in health class h. In our simulation $m^h = 0$ unless $h = i$, in which case $m > 0$.

Epidemiological parameters are taken from Fenichel et al. (2011), $\beta = 0.0925$ and $\nu = 0.1826$, to represent a flu-like epidemic. In order for this value of β to generate realistic flu-like dynamics the disease free contact level must be the same as in Fenichel et al., which is 5. For the base case, I assume an elasticity of substitution is less than unit elastic and equal to 0.6, implying $\gamma = -2/3$. Thinking of the numeraire as time, I set $b = 24$ and solve for $a = 0.098$ to preserve the optimality of 5 contacts in the absence of disease. The daily discount rate, $\rho = 1.37 \times 10^{-4}$, corresponds to an annual discount rate of 5%. Assuming no disease, 5 contacts yields a utility payoff of $\bar{u} = 15.610$, and we set $m^i = 2\bar{u}$, suggesting sick individuals lose two days of disease free utility for every infected day. I assume an initial prevalence of 1 in 1000 and no initial recovered population. These parameters and functional forms are chosen to illustrate the analytical results and to facilitate a numerical example.

Discrete approximations to the social planner, constrained social planner problems, and decentralized decision maker problem with rational expectations were solved by mathematical programming using the AD Model Builder (admb-project.org) algorithmic differentiation template and libraries for C++. Simulations were run sufficiently long so that the proportion of infected individuals was zero to numerical precision (at least 250 daily time steps). The decentralized decision making model with adaptive expectations was solved using a discrete approximation by dynamic programming with a 12-day planning horizon following Fenichel et al. (2011), implemented in Mathematica 8.0 (Wolfram Research).

The results of four simulated epidemics, associated with different decision making frameworks, are shown in Fig. 3. The top panel of Fig. 3 shows epidemic curves for the epidemic under decentralized rational (dotted curve) and adaptive (solid curve) expectations, the constrained social planner (dashed curve), and the social planner constrained to provide

[14]These are stylized assumptions to develop a specific utility specification with an interpretation. Real contact behavior is undoubtedly more complex, but this stylized model could serve as the base for future empirical work.

each health class with at least 3 contacts. The fully unconstrained social planner chooses contacts so that hardly any new infections occur (Table 1). Both decentralized decision models result in greater and earlier peak prevalence than when decisions are made by a constrained social planner or a social planner constrained to provide 3 contacts. The bottom panel of Fig. 3 is comparable to Fig. 2 and illustrates how the decision making context affects the limiting behavior of the dynamics and the final epidemic size. Table 1 shows that the final epidemic size, across a range of parameters, is similar across decentralized decision models and the constrained social planner model. Furthermore, in some cases decentralized decision making leads to smaller final epidemic size, i.e., the total fraction infected (Table 1).

To analyze the welfare impacts I calculated the net present value of utility for the representative agent in the population at the start of the epidemic for an infinite horizon and for the first 150 days of the epidemic (Table 2). Prevalence has gone to zero by day 150. Furthermore, welfare changes are easier to compare without adding a net present disease free value, but doing so does not change the rank ordering of welfare outcomes. The net present value of utility is greatest for the social planner, as is expected. Proposition 3 suggests that it is unclear if a constrained social planner will dominate decentralized decision making, and this lack of clarity is supported by numerical simulations. In some cases and under some assumptions the constrained social planner does make society better off than decentralized decision making, while in others society is better left to behave in a decentralized way. These results are sensitive to parameters related to utility (Table 2).[15] Numerical results reinforce that without detailed knowledge of the utility function a constrained social planner could make society worse off in utility and could induce a greater final epidemic size (Table 1). In other cases, the constrained social planner does make society better off. However, a social planner that is able to target across health classes, even if the social planner is limited in the ability to reduce contacts can still make society substantially better off in terms of utility (Table 2) and can reduce the final epidemic size (Table 1).

These different welfare and epidemiological results are driven by different patterns of behavior (Fig. 4). The social planner pursues a near quarantine strategy for infected individuals, which results in nearly no behavioral change for susceptible and recovered individuals. However, when the social planner is constrained to allow three contacts, the social planner optimally increases the number of contacts made by recovered individuals as suggested by the analytical model.

4. Discussion and conclusion

Reduced form models suggest that social distancing, defensive behavioral changes, appear to be important in epidemics (Caley et al., 2007). Increasingly, social distancing is not just discussed as decentralized behavioral change, but is also discussed in the context of policy to avert losses associated with epidemics (World Health Organization, 2006). To evaluate these policies and the behavioral incentives they generate requires an economic decision model (Heckman, 2010). The current paper provides an economic behavioral framework for understanding the economic incentives that infectious disease provides to individuals for engaging in social distancing. I also discuss the potential gains from public interventions to internalize dynamic externalities that occur because of the commons nature of public health

[15]Sensitivity analysis is conducted with respect to the parameters presented. However, if the elasticity of substitution is changed, then the share parameter must also be changed to preserve the optimality of 5 contacts. Furthermore, to preserve units I work in units of $\bar{u}$ for infection costs, which are recalculated based on the elasticity of substitution. Conducting sensitivity analysis this way enables me to separate the biological and economic parameters and focus on the economic parameters.

(Bell and Gersbach, 2009). This perspective is useful given the importance of commons problems in economics (Stavins, 2011).

The role of recovered individuals in protecting susceptible individuals is easily and often overlooked when considering public health interventions. "Common sense" suggests that it is safe to ignore recovered individuals because such individuals cannot get sick, are not suffering infection, and cannot infect others. However, the analytical and numerical analysis suggests that the behavior of recovered individuals is critical to behaviorally based disease prevention and management strategies. Kremer (1996) showed the importance of considering exogenous types of heterogeneity in the economics of infectious disease, and Auld (2003) extended this work illustrating the importance of a key source of heterogeneity, health state, that evolves with the epidemic. However, because these papers focused on HIV, they did not consider the role of immune individuals. Epidemiologists often discuss "herd immunity" when discussing vaccination strategies, and the positive externalities associated with vaccination have been considered by economists (Boulier et al., 2007; Gersovitz and Hammer, 2004). The herd immunity or positive externality associated with vaccination or acquired immunity is contradicted by non-targeted social distancing policies because untargeted social distancing policies induces recovered individuals to reduce contacts. Increasing the contacts made by recovered individuals lowers the probability of susceptible individuals contacting infected individuals or allows susceptible and infected individuals to increase contacts without changing infection probability. The effects of non-pharmaceutical behavior based policies on recovered individuals needs to be considered. It is however difficult to envision highly targeted policies aimed at immune individuals, perhaps a public transportation pass with a flu vaccination. In real systems immunity is uncertain and public health officials are unlikely to ask anyone, even those believed to be immune, to increase exposure to infectious agents. The policy-relevant insight is to recognize how policies may affect recovered individuals' behaviors, whether policies that elicit such behavioral changes provide net benefits, and the importance of considering heterogeneous behavior that is driven by health state.

Prior authors have emphasized the importance of targeting health interventions (Bell and Gersbach, 2009; Fenichel and Horan, 2007). It is not possible to analytically rule out that non-targeted behavioral policies yield lower social welfare, an economically undesirable outcome, and potentially worsen health outcomes, measured by final epidemic size. The numerical example provided reinforces this point because whether a constrained social planner enhances welfare relative to decentralized behavior depends on assumptions about expectations, cost of infection, and the utility of contacts. This analysis is particularly relevant for policy because it appears that most implemented social distancing strategies (e.g., Stern and Markel, 2009) or those promoted in the public health and epidemiology literature are not targeted by health status (Cauchemez et al., 2008). Moreover, such interventions are unlikely to be optimally chosen. Additionally, the models in this paper do not consider the administrative cost of enforcing a social distancing policy. The numerical results suggest that small administrative costs could potentially reverse the desirability of non-targeted interventions, in the cases when non-targeted interventions are desirable (though this is ultimately an open empirical question). In short, untargeted policies are not guaranteed to do no harm, and caution should be taken when adopting public social distancing policies.

Individual social distancing, conditional on one's health status, is certainly important in the spread of infectious disease and affects the welfare loss associated with a disease. Our results from the social planner problem suggest that targeted interventions, such as providing incentives for infectious individuals to socially distance or self-quarantine, are likely welfare enhancing. But, Kumar et al. (2012) provide evidence that in practice there may be strong dis-incentives to engage in social distancing, particularly for infected individuals. Yet, if there are few infectious individuals in the population, it seems reasonable that for infectious pathogens with characteristics that might be associated with a pandemic flu, e.g., a high conditional infectiveness, β, that susceptible individuals may be able to adequately compensate infected individuals to forgo contacts and self-quarantine or substantially reduce contacts.

For behavioral based epidemiological interventions for infectious disease it is important to consider how policies interact with individual's microeconomic incentives. While it is not always the case that non-targeted policies will make society worse off than decentralized decision making, it is clear that targeted policies would lead to greater benefits, and part of the reason that targeting is so important is the behavior of immune individuals. This analysis required thinking first about economic tradeoffs that lead to behavioral incentives and then about infection dynamics. Developing a structural understanding of economic behavioral response to infection risk is an open area of inquiry. Yet doing so is imperative for mechanism design that leverages private health incentives and yields efficient infectious disease policy. Moreover, it is important that this economic research is done in a way so that infectious disease epidemiologists, the people consulted on infectious disease policy, recognize the explicit integration of economics with epidemiology.

Acknowledgments

Thoughtful comments by members of the National Institute for Mathematical and Biological Synthesis SPIDER working group, Charles Perrings, and Nick Kuminoff contributed to previous versions of this manuscript. I thank two anonymous reviewers for their thoughtful suggestions. This publication was made possible by Grant number 1R01GM100471-01 from the National Institute of General Medical Sciences (NIGMS) at the National Institutes of Health. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS.

References

Almond D. Is the 1918 influenza pandemic over? Long-term effects of in utero influenza exposure in the post-1940 U.S. population. Journal of Political Economy. 2006; 114:672–712.

Almond D, Mazumder B. The 1918 influenza pandemic and subsequent health outcomes: an analysis of SIPP data. American Economic Review. 2005; 95:258–262.

Anderson RM, May RM. Population biology of infectious diseases: part I. Nature. 1979; 280:361–367. [PubMed: 460412]

Auld MC. Choice, beliefs, and infectious disease dynamics. Journal of Health Economics. 2003; 22:361–377. [PubMed: 12683957]

Barrett S, Hoel M. Optimal disease eradication. Environmental and Development Economics. 2007; 12:627–652.

Begon M, Bennett M, Bowers RG, French NP, Hazel SM, Turner J. A clarification of transmission terms in host-microparasite models: numbers, densities and areas. Epidemiological Infection. 2002; 129:147–153.

Bell C, Gersbach H. The macroeconomics of targeting: the case of an enduring epidemic. Journal of Health Economics. 2009; 28:54–72. [PubMed: 18977545]

Boulier BL, Satta TS, Goldfarb RS. Vaccination Externalities. The BE Journal of Economic Analysis and Policy. 2007; 7

Brock WA, Xepapadeas A. Valuing biodiversity from an economic perspective: a unified economic, ecological, and genetic approach. American Economic Review. 2003; 93:1597–1614.

Caley P, Philips DJ, McCracken K. Quantifying social distancing arising from pandemic influenza. Journal of the Royal Society Interface. 2007; 5:631–639.

Caputo, MR. Optimal Control Theory and Applicationss. Cambridge University Press; New York: 2005. Foundations of Dynamic Economic Analysis.

Cauchemez S, Valleron AJ, Boelle PY, Flahault A, Ferguson N. Estimating the impact of school closure on influenza transmission from Sentinel data. Nature. 2008; 452:750–755. [PubMed: 18401408]

Chakraborty S, Papageorgiou C, Sebastian FP. Disease, infection dynamics, and development. Journal of Monetary Economics. 2010; 57:859–872.

Chen F, Jiang M, Rabidoux S, Tobinson S. Public avoidance and epidemics: insights from an economic model. Journal of Theoretical Biology. 2011; 278:107–119. [PubMed: 21419135]

Clark, CW. Mathematical Bioeconomics Optimal Management of Renewable Resourcess. John Wiley & Sons; Hoboken: 2005.

Dasgupta P, Maler KG. The economics of non-convex ecosystems: introduction. Environmental and Resource Economics. 2003; 26:499–525.

Fenichel EP, Castillo-Chavez C, Ceddia MG, Chowell G, Gonzalez Parra PA, Hickling GJ, Holloway G, Horan R, Morin B, Perrings C, Springborn M, Velazquez L, Villalobos C. Adaptive human behavior in epidemiological models. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108:6306–6311. [PubMed: 21444809]

Fenichel EP, Horan RD. Gender-based harvesting in wildlife disease management. American Journal of Agricultural Economics. 2007; 89:904–920.

Ferguson NM, Cummings DAT, Cauchemez S, Fraser C, Riley S, Meeyai A, lamsirithaworn S, Burke DS. Strategies for containing an emerging influenza pandemic in Southeast Asia. Nature. 2005; 437:209–214. [PubMed: 16079797]

Francis PJ. Dynamic epidemiology and the market for vaccinations. Journal of Public Economics. 1997; 63:383–406.

Francis PJ. Optimal tax/subsidy combinations for the flu season. Journal of Economic Dynamics & Control. 2004; 28:2037–2054.

Fraser C, Donnelly CA, Cauchemez S, Hanage WP, Van Kerkhove MD, Hollingsworth TD, Griffin J, Baggaley RF, Jenkins HE, Lyons E, Jombart J, Hinsley T, Grassly WR, Balloux NC, Ghani F, Ferguson AC, Rambaut NM, Pybus A, Lopez-Gatell OG, Alpuche-Aranda H, Chapela lBCM, Zavala EP, Espejo Guevara DM, Checchi F, Garcia E, Hugonnet S, Roth C. Pandemic potential of a strain of influenza A (H1NA): early finds. Science. 2009; 324:1557–1561. [PubMed: 19433588]

Funk S, Salathe M, Jansen VAA. Modelling the influence of human behaviour on the spread of infectious diseases: a review. Journal of the Royal Society Interface. 2010; 7:1247–1256.

Galvani AP, Reluga TC, Chapman GB. Long-standing influenza vaccination policy is in accord with individual self-interest but not with the utilitarian optimum. Proceedings of the National Academy of Sciences of the United States of America. 2007; 104:5692–5697. [PubMed: 17369367]

Geoffard PY, Philipson T. Rational epidemics and their public control. International Economic Review. 1996; 37:603–624.

Gersovitz M. The economics of infection control. Annual Review of Resource Economics. 2011; 3:277–296.

Gersovitz M, Hammer JS. Infectious diseases, public policy, and the marriage of economics and epidemiology. The World Bank Research Observer. 2003; 18:129–157.

Gersovitz M, Hammer JS. The economical control of infectious diseases. The Economic Journal. 2004; 114:1–27.

Gersovitz M, Hammer JS. Tax/subsidy policies toward vector-borne infectious diseases. Journal of Public Economics. 2005; 89:647–674.

Glass RJ, Glass LM, Beyeler WE, Min HJ. Targeted social distancing design for pandemic influenza. Emerging Infectious Diseases. 2006; 12:1671–1681. [PubMed: 17283616]

Halloran ME, Ferguson NM, Eubank S, Ira M, Longini J, Cummings DAT, Lewis B, Xu S, Fraser C, Vullikanti A, Germann TC, Wagener D, Beckman R, Kadau K, Barrett C, Macken CA, Burke DS, Cooley P. Modeling targeted layered containment of an influenza pandemic in the United States. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105:4639–4644. [PubMed: 18332436]

Heckman JJ. Building bridges between structural and program evaluation approaches to evaluating policy. Journal of Economic Literature. 2010; 48:356–398. [PubMed: 21743749]

Hethcote HW. The mathematics of infectious diseases. SIAM Review. 2000; 42:599–653.

Horan RD, Fenichel EP, Drury KLS, Lodge DM. Managing ecological thresholds in coupled environmental–human systems. Proceedings of the National Academy of Sciences of the United States of America. 2011a; 108:7333–7338. [PubMed: 21502517]

Horan RD, Fenichel EP, Melstrom RT. Wildlife disease bioeconomics. International Review of Environmental and Resource Economics. 2011b; 5:23–61.

Horan RD, Fenichel EP, Wolf CA, Gramig BM. Managing infectious animal disease systems. Annual Review of Resource Economics. 2010; 2:101–124.

Keogh-Brown MR, Wren-Lewis S, Edmunds WJ, Beutels P, Smith RD. The possible macroeconomic impact on the UK of an influenza pandemic. Health Economics. 2010; 19:1345–1360. [PubMed: 19816886]

Kermack, WO.; McKendrick, AG. Proceedings of the Royal Society, London Series A 115. 1929. Contributions to the mathematical theory of epidemics, part 1; p. 700-721.

Kremer M. Integrating behavioral choice into epidemiological models of AIDS. The Quarterly Journal of Economics. 1996; 111:549–573.

Kumar S, Crouse Quinn S, Kim KH, Daniel LH. Freimuth Vs. The impact of workplace policies and other social factors on self-reported influenza-like illness incidence during the 2009 H1N1 pandemic. American Journal of Public Health. 2012; 102:134–140. [PubMed: 22095353]

Lipsey RG, Lancaster K. The general theory of second best. Review of Economic Studies. 1956; 24:11–32.

McCallum HI, Barlow N, Hone J. How should pathogens transmission be modelled? TRENDS in Ecology and Evolution. 2001; 16:295–300. [PubMed: 11369107]

Mesnard A, Seabright P. Escaping epidemics through migration? Quarantine measures under incomplete information about infection risk. Journal of Public Economics. 2009; 93:931–938.

Nasell I. Stochastic models of some endemic infections. Mathematical Biosciences. 2002; 179:1–19. [PubMed: 12047919]

Philipson, T. Economic epidemiolgy and infectious diseases. In: Culyer, AJ.; Newhouse, JP., editors. Handbook of Health Economics. Vol. 1. Elsevier Science B.V; San Diego, CA: 2000.

Reluga TC. Game theory of social distancing in response to an epidemic. PLoS Computational Biology. 2010; 6:e1000793. [PubMed: 20523740]

Rondeau D. Along the way back from the brink. Journal of Environmental Economics and Management. 2001; 42:156–182.

Shapiro C, Stiglitz JE. Equilibrium unemployment as a worker discipline device. American Economic Review. 1984; 74:433–444.

Smith RD, Keogh-Brown MR, Barnett T, Tait J. The economy-wide impact of pandemic influenza and the UK: a computable general equilibrium modelling experiment. BMJ. 2009; 339:b4571. [PubMed: 19926697]

Stavins RN. The problem of the commons:still unsettled after 100 years. American Economic Review. 2011; 101:81–108.

Stern AM, Markel H. What Mexico taught the world about pandemic influenza preparedness and community mitigation strategies. Journal of the American Medical Association. 2009; 302:1221–1222. [PubMed: 19755702]

Veliov VM. On the effect of population heterogeneity on dynamics of epidemic diseases. Journal of Mathematical Biology. 2005; 51:123–143. [PubMed: 16012804]

Weitzman ML. On the welfare significance of national product in a dynamic economy. The Quarterly Journal of Economics. 1976; 90:156–162.

Weitzman ML. A contribution the theory of welfare accounting. Scandinavian Journal of Economics. 2001; 103:1–23.

World Health Organization. Nonpharmaceutical interventions for pandemic influenza, national and community measures. Emerging Infectious Diseases. 2006; 12:88–94. [PubMed: 16494723]

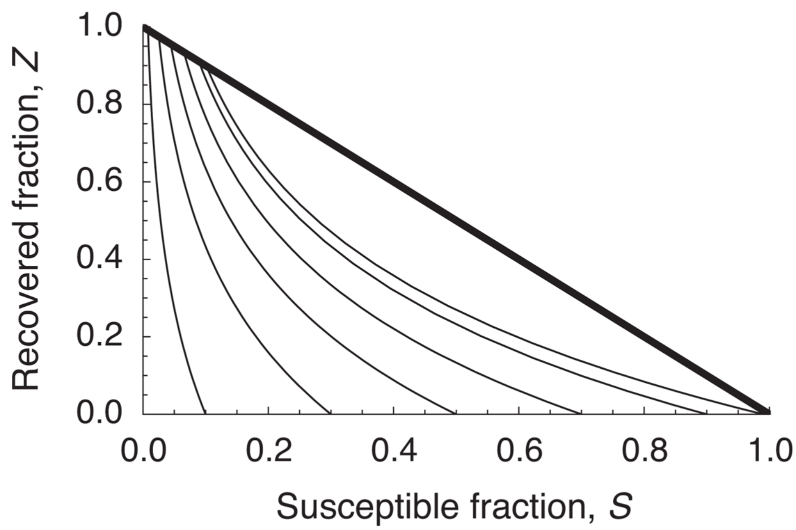

Fig. 1.
Sample trajectories when behavior is constant over the course of the epidemic.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

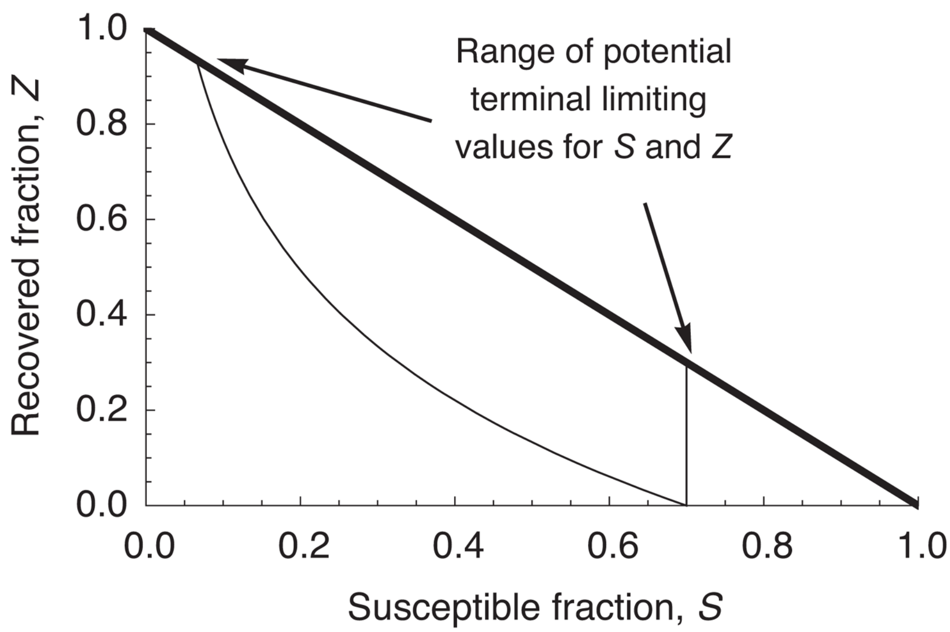

Fig. 2.

The range of limiting values for the susceptible and recovered population when the epidemic fades out given an initial infected population of 0.3. The arrows represent the extent of the range of limiting values on the triangle.

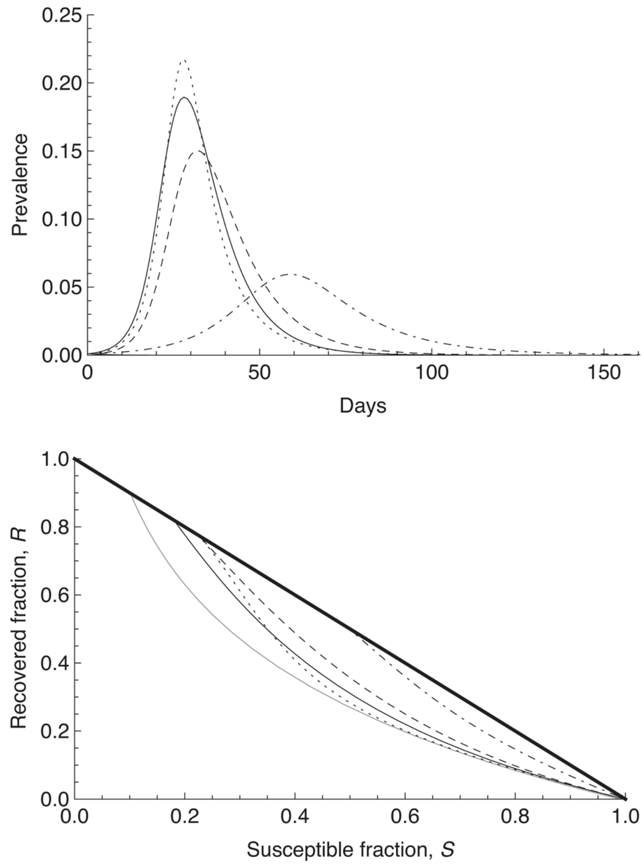

Fig. 3.
Epidemiological dynamics. The top panel shows the prevalence over time for decentralized decision making with adaptive expectations (solid curve), with rational expectations (dotted curve), a social planner constrained to provide at least 3 contacts to each health class (dot-dash curve), and the constrained social planner (dashed curve). The bottom panel shows the limiting behavior for these four cases and the no behavioral change case (gray curve). The social planner program leads the epidemic to die out to quickly to see in the figure.

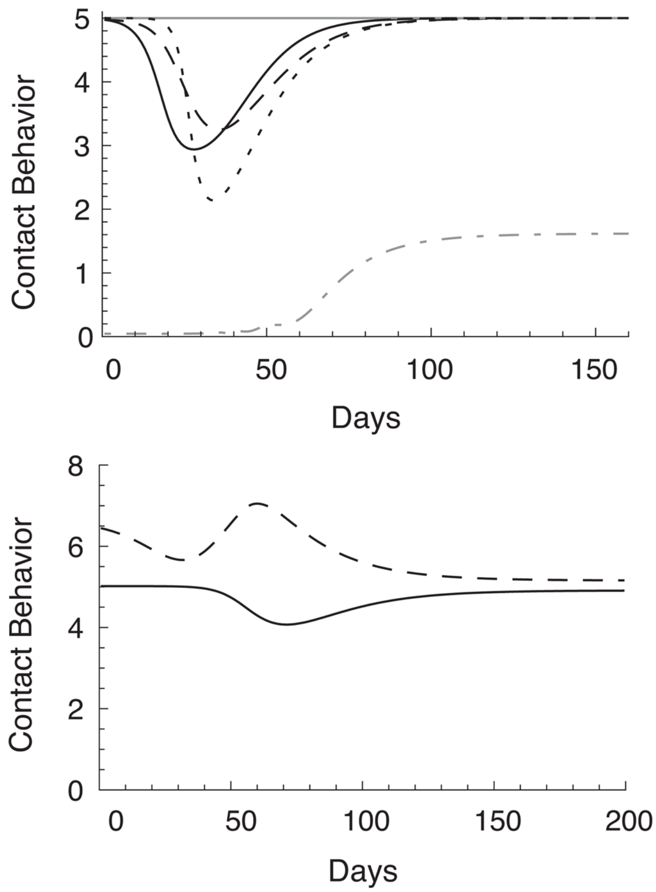

Fig. 4.
The time paths of behavior. The top panel shows behavior as controlled by the social planner (gray curves) for the infected class (dashed), the susceptible (solid), and the recovered (dotted); the constrained social planner (black, dashed curve), and susceptible individual behavior under decentralized decision making with rational expectations (black dotted curve) and adaptive expectations (black, solid curve). The bottom panel shows the behavior as controlled by the social planner that must provide 3 contacts for susceptible individuals (solid curve) and recovered individuals (dashed curve). Infected individuals are not shown, but always make 3 contacts.

Table 1

Final epidemic size for different economic parameter assumptions.

		σ		m	
	Baseline σ=0.6, $m = 2\bar{u}$	**Low $\sigma = 0.4$**	**High $\sigma = 0.8$**	**Low $m = 0.5\bar{u}$**	**High $m = 6\bar{u}$**
Social planner	0.001	0.001	0.001	0.001	0.001
Constrained social planner	0.775	0.832	0.724	0.884	0.643
Decentralized decision making (adaptive expectations)	0.820	0.853	0.785	0.819	0.823
Decentralized decision making (rational expectations)	0.776	0.823	0.733	0.873	0.652
Social planner, must allow at least 3 contacts	0.506	0.526	0.489	0.575	0.418

Table 2

Net present utility on day 0 of the epidemic for the representative agent as a percent of the net present utility achievable by the unconstrained social planner. The utility score for the unconstrained social planner is provided.

		σ		m	
	Baseline $\sigma = 0.6$, $m = 2\bar{u}$	**Low $\sigma = 0.4$**	**High $\sigma = 0.8$**	**Low $m = 0.5\bar{u}$**	**High $m = 6\bar{u}$**
Social planner	2317.58	2470.82	2210.63	2317.72	2317.2
Constrained social planner	93.78%	93.59%	93.98%	98.35%	83.9%
Decentralized decision making (adaptive expectations)	93.60%	93.50%	93.69%	98.11%	81.5%
Decentralized decision making (rational expectations)	93.83%	93.65%	94.02%	98.37%	83.6%
Social planner, must allow at least 3 contacts	95.94%	95.74%	96.08%	98.81%	89.7%