https://pubsonline.informs.org/journal/opre

OPERATIONS RESEARCH
Vol. 72, No. 2, March–April 2024, pp. 533–548
ISSN 0030-364X (print), ISSN 1526-5463 (online)

Crosscutting Areas

Testing, Voluntary Social Distancing, and the Spread of an Infection

Daron Acemoglu,[a] Ali Makhdoumi,[b,*] Azarakhsh Malekian,[c] Asuman Ozdaglar[d]

[a] Department of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142; [b] Fuqua School of Business, Duke University, Durham, North Carolina 27708; [c] Rotman School of Management, University of Toronto, Toronto, Ontario M5S 3E6, Canada; [d] Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
*Corresponding author

Contact: daron@mit.edu (DA); ali.makhdoumi@duke.edu, https://orcid.org/0000-0001-5422-0314 (AlM); azarakhsh.malekian@rotman.utoronto.ca, https://orcid.org/0000-0001-9464-746X (AzM); asuman@mit.edu, https://orcid.org/0000-0002-1827-1285 (AO)

Received: June 25, 2020
Revised: February 17, 2021; August 19, 2021
Accepted: October 14, 2021
Published Online in Articles in Advance: June 19, 2023

Area of Review: Policy Modeling and Public Sector OR

https://doi.org/10.1287/opre.2021.2220

Abstract. We study the effects of testing policy on voluntary social distancing and the spread of an infection. Agents decide their social activity level, which determines a social network over which the virus spreads. Testing enables the isolation of infected individuals, slowing down the infection. However, greater testing also reduces voluntary social distancing or increases social activity, exacerbating the spread of the virus. We show that the effect of testing on infections is nonmonotone. This nonmonotonicity also implies that the optimal testing policy may leave some of the testing capacity of society unused.

Funding: The authors acknowledge support from C3.DTI funding.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/opre.2021.2220.

Keywords: game theory • network formation • contagion process • voluntary social distancing • optimal testing policy

1. Introduction

The COVID-19 pandemic has reignited interest in models of epidemics and their control. A point of broad agreement among different approaches is that ramping up testing capacity is one of the most effective ways of combating the pandemic (Brumfiel 2020, BruSoe-Lin and Hecht 2020, Searchinger et al. 2020). One issue that has not received much attention, however, is whether and how different testing strategies impact voluntary social distancing decisions of individuals.

In this paper, we develop a simple model to investigate the effects of testing on infections and provide insights into optimal testing strategies. We model social activity and voluntary distancing as a network formation problem and use a simple percolation process to represent the spread of a virus over the endogenous social network. Motivated by tractability considerations, we choose a model of percolation rather than the SIR (susceptible, infected, and recovered) setup more commonly used in the analysis of COVID-19. As we explain here, our percolation model leads to the same behavior of cumulative infections as a discrete version of a standard SIR model, but enables a much more tractable and general analysis of the spread of the virus and its dependence on the endogenous social network.[1]

Our main results confirm the major benefit society can reap from testing and isolating infected individuals and leads to two new insights.[2] First, greater testing can lead to more social activity (less social distancing) and thus a denser social network, because when infected individuals are more likely to be identified and isolated, agents feel more secure to initiate contacts. We show that, for a nontrivial set of parameters, greater testing can increase infections. This happens, in particular, when the equilibrium involves some groups choosing an intermediate level of social activity because of their fear of infection. Second, we demonstrate that when the testing capacity of society is limited, optimal policy may involve leaving some of this capacity unused in order to avoid adverse effects on social distancing. This result also implies that testing should be combined with mandatory social distancing measures to avoid these adverse behavioral effects.

More formally, our model consists of n individuals of two types, high-type agents with greater value from social activity and low-type agents with lower values (more than two types are considered as an extension). Each individual i chooses a social activity level, $x_i \in [0,1]$, and the vector of social activities defines a contact (social) network among individuals. Specifically, there will be contact between agents i and j with probability $\eta x_i x_j$, where $\eta < 1$ captures the probability of a match conditional on activities. The utility of each individual is their utility gain from social activity minus their

infection probability. The virus spreads over the social network following a cascade process, whereby each infected and nonisolated individual transmits the infection to each one of its neighbors with probability $\beta \in (0,1]$. We assume that although an infected individual does not transmit the virus after isolation, transmission still occurs before testing detects and isolates an infected agent. More specifically, we assume that an infected individual who is detected transmits the virus with a smaller probability, βp, where $p \in [0,1)$.

Our analysis relies on a key lemma that establishes that the stochastic process for the spread of the virus satisfies natural monotonicity and concavity (submodularity) properties: More social activity leads to more infections but an individual's probability of infection increases less in their own social activity when they already have a high activity level.

We first prove that, as often maintained in prior analyses, more testing reduces infections when we take activity levels (and thus the social network) as given. Our key results concern the case in which the social activity levels are endogenous. We show that when testing probabilities are large, both types of individuals (high value and low value) choose maximal social activity (which can be interpreted as the same level of activity as the prepandemic period). In this region, the equilibrium behaves in an identical fashion to the exogenous social network case, and testing always reduces infections. More interestingly, however, for lower testing probabilities, either high-value or low-value individuals choose intermediate levels of social activity, and greater testing increases their infection probability.

We next turn to an analysis of optimal testing where the objective is to maximize the value of social activity minus the cost of infections. We assume that there is a fixed supply of tests and investigate how these should be allocated. Because high-value agents are socially more active, they are more likely to be infected and thus they should be tested first. If there is sufficient testing capacity, it is optimal to test all agents. However, most interestingly, we prove that for intermediate or low testing capacity, it is socially optimal not to use all available tests. The social planner should either test all high-value agents, but not the low-value agents, to discourage them from high levels of social activity. Or they should have no testing, even though there is capacity to test some of the high-value agents. The intuition can again be obtained by considering the impact of testing on voluntary social distancing: additional testing over this range reduces voluntary social distancing so much that it has no benefit in terms of containing the infection.

As already emphasized, these nonmonotonicity results are due to the impact of testing on equilibrium social activity (social distancing). If the social planner can mandate social distancing, then there is no nonmonotonicity and it is always optimal to use all of the available testing capacity.

In addition, we show that uniform testing policies, where tests are allocated without reference to the type of agents, are worse than targeted testing policies. In our baseline model, high-value agents should be tested first because they are more active and thus more likely to transmit the virus. In an extension where we allow individuals to choose type-specific social activity levels (e.g., how much to socialize with high-value and low-value agents), we show that not testing high-value agents may, however, have an additional strategic benefit—It discourages low-value agents from socializing with high-value agents, slowing down the spread of the virus. Although this is not our main goal, we hope that our analysis is also helpful for the study and design of mitigation policies for pandemics, including COVID-19.

1.1. Related Literature

Our paper is related to three distinct literatures. First, ours is a model of endogenous social network formation. Seminal papers in this area include Jackson and Wolinsky (1996), Bala and Goyal (2000), and Currarini et al. (2009) (see Jackson (2008) and Vega-Redondo (2007) for book-length treatments of issues of network formation and contagion in networks). Differently from the most common approach in this literature, which is to look at pairwise or coalition-wise stable outcomes, we adopt a noncooperative approach to network formation (Bollobás and Béla 2001, Newman et al. 2001, Cabrales et al. 2011). In particular, we model the probability of connection between two agents as being proportional to the product of their levels of social activity, which leads to a tractable but rich set of interactions, in part because activity levels are neither strategic substitutes nor complements. The probability of a link between two agents depends on the action of both agents, which is also different from Goyal (1993), Bala and Goyal (2000), and Acemoglu et al. (2017b), who assume that players can unilaterally create directed links to others.

Second, there is a large amount of literature spanning various areas of economics where precautionary tools increase risk taking, as first emphasized by Peltzman (1975). An iconic example is hydraulic breaks increasing driving speed, thus undoing some or all of its beneficial effects on accidents (Lindgren and Stuart 1980, Crandall and Graham 1984, Keeler 1994). Relatedly, Philipson (2000) explores the interplay between the spread of an epidemic, which is reduced by preventative measures, and the demand for prevention, which is higher when the epidemic is more widespread (Philipson and Posner 1993). Philipson points out that the initial containment of an epidemic may make its ultimate eradication harder because it discourages preventative measures. Lakdawalla et al. (2006) show that

advances in HIV treatment have raised risky sexual behaviors and have slowed down the containment of the HIV virus (Bauch and Galvani 2013). Differently from other works in this vein, we establish that countervailing effects can outweigh the direct impacts and thus more testing can lead to higher infections.

Third, our paper is related to the recent literature on the effects of pandemics, especially of COVID-19. See, for example, Kruse and Strack (2020), Atkeson (2020), Jones et al. (2020), Glover et al. (2020), Berger et al. (2020), and Birge et al. (2020). More closely related are several papers endogenizing behavior and social distancing in the context of SIR models, such as Leung et al. (2018), Toxvaerd (2020), Eichenbaum et al. (2020), Farboodi et al. (2021), Maloney and Taskin (2020), and Gans (2020b). In particular, Eichenbaum et al. (2020) extend the canonical SIR model to study the interaction between economic decisions and pandemics, whereas Leung et al. (2018), Toxvaerd (2020), Farboodi et al. (2021), and Maloney and Taskin (2020) incorporate simple social distancing behavior in an otherwise standard SIR model.[3]

A few recent papers in this literature, which study the effects of lockdown and testing strategies, are more closely related. For example, Alvarez et al. (2020), Acemoglu et al. (2021), Piguillem and Shi (2020), Brotherhood et al. (2020), and Eshragh et al. (2020) explore the effects of different testing and isolation policies on the dynamics of infections and optimal lockdowns, whereas Deb et al. (2020) consider targeted testing combined with transfers. Even more closely related to our paper are Drakopoulos and Randhawa (2020) and Ely et al. (2021), who study optimal testing policy when tests are inaccurate, (and Kasy and Teytelboym (2020), who investigate the implications of false quarantine). None of these papers analyzes the impact of testing on voluntary social distancing and behavior. In addition, different from the previous papers, we provide a full characterization of social activity and show how it depends on infection probabilities and testing strategies.

Various other issues, such as estimation, testing, and control related to COVID-19 and more broadly pandemics have also been studied. Dasaratha (2020), in an SIR model, studies theoretical conditions for risk compensation to overcome the direct benefits of safety-enhancing interventions; Kaplan (2020) considers a statistical model for estimating the effectiveness of isolation and quarantine; and Drakopoulos et al. (2017) study the open-loop control of epidemics on a network and show how the network structure affects the number of resources required to contain the spread of infection. Finally, Wang et al. (2009) study the strategic interaction among states in allocating their resources, whereas Sun et al. (2009) examine the strategic considerations in allocating drugs during a pandemic. To the best of our knowledge, no other work has pointed out or studied the nonmonotonic impacts of testing on infections. For a survey of recent advances see, for example, McAdams (2020) and Gans (2020).

The rest of the paper is organized as follows. In Section 2, we present our model, describing the formation of contact network and the stochastic process governing the spread of infection. We also show that for a given (exogenous) network, increasing the testing capability decreases the infection probability. In Section 3, we characterize the equilibrium outcome and show that increasing testing probabilities may adversely increase the (equilibrium) infection probability. In Section 4, we characterize the optimal testing policy and show it may be optimal to have underused testing capacity. Section 5 considers three extensions: multiple groups of agents, costly tests, and social activity differentially targeted toward different groups. Section 6 concludes. The appendix contains the proofs of the main results, and the online appendix contains the remaining proofs and additional results.

2. Environment

We study the spread of an infection among n individuals (also referred to as agents or nodes) represented by the set $\mathcal{V} = \{1, \ldots, n\}$. Each agent $i \in \mathcal{V}$ decides about its level of *social activity*, denoted by $x_i \in [0,1]$. Higher social activity provides greater utility to agents but also leads to faster spread of the infection. Agents are heterogeneous in terms of their value of social interaction. In our main model, we assume that the agent's type (value of interaction) is either $v_H \in [0,1]$ or $v_L \in [0,1]$, where $v_L < v_H$. We also use $\mathcal{H}$ and $\mathcal{L}$ to denote individuals with high and low values, respectively, and r_H and r_L to denote the population fractions of high- and low-value agents. Section 5 extends our results to a setting with $m > 2$ types of agents.

A virus infects a random individual and then spreads to others through a stochastic process described in Section 2.2. We analyze the implications of testing (for infection), represented by testing probabilities for two types of agents, $\alpha_L, \alpha_H \in [0,1]$, and study the optimal *testing policy*, (α_L, α_H), of a (benevolent) social planner. Each infected individual who tests positive will be isolated from the rest of society. If an individual is infected and not isolated, it will expose its neighbors to the infection. If it is isolated, it still has a (smaller) chance to infect its neighbors and expose them to the infection.

We let $\mathbf{x} = (x_1, \ldots, x_n)$ denote the social activity profile of all individuals. We also let $\mathbf{x}_{-i}$ represent the social activity profile of all individuals except agent i. In what follows, for any vector $\mathbf{x} \in \mathbb{R}^n$ and set $S \subseteq \mathcal{V}$, $\mathbf{x}_S$ denotes the elements of $\mathbf{x}$ for the indices in set S, and $\mathbf{x}_{-S}$ denotes the elements of $\mathbf{x}$ for the indices outside S.

We next describe how social activity levels determine the (endogenous) social network in this community and how the infection spreads over this social network.

2.1. Network of Contacts

The social activity profile $\mathbf{x} = (x_1, \ldots, x_n)$ of agents generates a social (contact) network in which agents i and j are connected with a probability that depends on x_i and x_j. Let $\mathbf{G} = (\mathcal{V}, \mathbf{E})$ be a random network where $\mathbf{E} \in \{0,1\}^{n \times n}$ denotes the random edges (thus, $\mathbf{E}_{ij} = \mathbf{E}_{ji}$), where $\mathbf{E}_{ij}$ for $i, j \in \mathcal{V}$ are independent binary random variables with

$$\mathbb{P}[\mathbf{E}_{ij} = 1] = \eta x_i x_j,$$

for some $\eta \in (0,1]$. This parameter η captures the probability of match between two individuals conditional on activity levels. We denote a *realized network* by $G = (\mathcal{V}, E)$, where E_{ij} is a realization of $\mathbf{E}_{ij}$, and thus $E_{ij} = 1$ means that there is a link between agents i and j.

We next describe the stochastic process governing the spread of infection and define the infection probability of agents.

2.2. Spread of the Infection

Let us denote the neighbors of node i by $N(i) = \{j \in \mathcal{V} : E_{ij} = 1\}$. For a given network $G = (\mathcal{V}, E)$, there is a stochastic process that governs the spread of infection as follows. One of the individuals uniformly at random becomes infected, and the infection then spreads to others via a percolation process over the social network that is a generalization of the independent cascade model (Kempe et al. 2015).[4]

We can describe the dynamics of infection as follows: At time 0, one of the agents (chosen uniformly at random) $s \in \mathcal{V}$ gets infected. At any round $t \geq 0$ for any agent $i \in \mathcal{V}$, we let $d_i^{(t)}$ denote the number of neighbors of agent i who are infected and tested at time t and $\hat{d}_i^{(t)}$ denote the number of neighbors of agent i that are infected at time t but not tested. For each agent i, these two variables are initially zero and then evolve over time as described next. At time 0, node s will be tested with probability α_i, in which case, at time 1 for each neighbor j of agent s, we have $d_j^{(1)} = 1$. Here, α_i is either α_H or α_L depending on whether agent s belongs to $\mathcal{H}$ or $\mathcal{L}$. With probability $1 - \alpha_i$, however, agent s will not be tested, in which case, at time 1 for all neighbors of agent s, such as j, we have $\hat{d}_j^{(1)} = 1$. If an infected node is not tested (and therefore not isolated), it will be *active* for one round and transmits infection to its neighbors with *transmission probability* $\beta \in (0,1]$. If an infected node is tested, it can be isolated and prevented from transmitting the virus. However, before isolation takes place, some social contacts may occur and spread infection. We represent this possibility with a smaller probability of infection, βp, where $p \in [0,1)$. From active agents, the infection simultaneously and independently transmits to each of their uninfected neighbors. If an uninfected agent is a neighbor to multiple infected individuals (i.e., there exists at least two i and i' such that $j \in N(i)$ and $j \in N(i')$), then the infection is transmitted to agent j in an order-independent fashion. This implies, for example, that if j is uninfected and a neighbor to two active agents who are not tested, then j becomes infected with probability $1 - (1 - \beta)^2$. If a neighbor of an active node does not become infected at time t, then it will never again become infected via that node. Given this process, for each node i, the probability of getting infected at time $t + 1$ becomes

$$1 - (1 - \beta)^{\hat{d}_i^t}(1 - \beta p)^{d_i^t},$$

where we recall that $\hat{d}_i^t$ is the number of i's neighbors who are infected and not tested at time t and d_i^t is the number of i's neighbors who are infected and tested at time t.

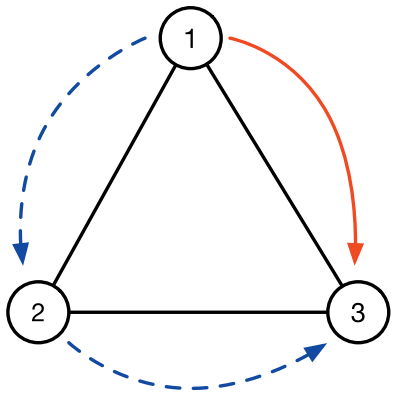
Definition 1. For any agent $i \in \mathcal{V}$, network $G = (\mathcal{V}, E)$, and testing policy (α_L, α_H) we let $\mathbb{P}_i^{\inf}(G, \alpha_L, \alpha_H)$ denote the probability of infection reaching agent i. This probability is over the randomness in the source of infection, the randomness in testing, and the randomness in the stochastic process described above.

We next illustrate the stochastic process and the infection probability by means of an example.

Example 1. We consider a setting with three agents {1, 2, 3} who are fully connected to each other with testing probabilities $\alpha_H = \alpha_L = \alpha$ and find the infection probability of agent 3. We list the three cases for the source of infection and find this probability:

- With probability 1/3, the infection hits agent 3. In this case, agent 3 gets infected.
- With probability 1/3, the infection hits agent 1. In this case, there are two possibilities:
  (i) Node 1 will get tested, with probability α. Here, there are two further cases. In the first, the infection reaches node 3 directly through the edge between nodes 1 and 3 with probability βp (i.e., through the solid path in Figure 1). In the second case (with

Figure 1. (Color online) From Agent 1, the Infection Can Reach Agent 3 Through Two Paths



Notes. (i) Directly via the edge between them whose probability depends on whether agent 1 will be tested or not. (ii) Through agent 2 that, again, depends on whether agents 1 and 2 will be tested.

probability $1-\beta p$), this edge is not active, and the infection reaches node 3 through node 2 with probability $(1-\alpha)\beta^2 p + \alpha\beta^2 p^2$ (i.e., through the dashed path in Figure 1). The first term is the probability of node 2 not being tested, in which case the infection reaches node 3 with probability $\beta p \times \beta$, and the second term is the probability of node 2 being tested, in which case the infection reaches node 3 with probability $\beta p \times \beta p$.
(ii) Node 1 will not get tested, with probability $(1-\alpha)$, and proceeding as in the previous case, we can find the infection probability to be $(\beta + (1-\beta)((1-\alpha)\beta^2 + \alpha\beta^2 p))$.

Therefore, starting from the infection of agent 1 the overall infection probability of node 3 is

$$\alpha\Big(\beta p + (1-\beta p)\Big((1-\alpha)\beta^2 p + \alpha\beta^2 p^2\Big)\Big) + (1-\alpha)\Big(\beta + (1-\beta)\Big((1-\alpha)\beta^2 + \alpha\beta^2 p\Big)\Big).$$

- With probability 1/3, the infection hits agent 2. This case is identical to the previous one by swapping the role of nodes 1 and 2.

Putting these three cases together, the infection probability of agent 3, $\mathbb{P}_3^{\inf}(G, \alpha_L = \alpha, \alpha_H = \alpha)$, is

$$\frac{1}{3} + \frac{2}{3}\Big(\alpha\Big(\beta p + (1-\beta p)\Big((1-\alpha)\beta^2 p + \alpha\beta^2 p^2\Big)\Big) + (1-\alpha)\Big(\beta + (1-\beta)\Big((1-\alpha)\beta^2 + \alpha\beta^2 p\Big)\Big)\Big).$$

Definition 2. For a given social activity profile $\mathbf{x}$ and testing policy (α_L, α_H), we denote the infection probability of individual i in the random network of contacts by $\mathbb{P}_i^{\inf}(\mathbf{x}, \alpha_L, \alpha_H)$, that is,

$$\mathbb{P}_i^{\inf}(\mathbf{x}, \alpha_L, \alpha_H) = \mathbb{E}_{\mathbf{G}=(\mathcal{V},E):\, E_{ij}\sim\text{Bernoulli}(\eta x_i x_j)}[\mathbb{P}_i^{\inf}(\mathbf{G}, \alpha_L, \alpha_H)].$$

This probability is over the randomness in the formed network, the randomness in the source of infection, the randomness in testing, and the randomness in the stochastic process governing the spread of infection.

2.3. Utility of Agents and Solution Concept

The utility function of agent I is given by

$$u_i(\mathbf{x}, \alpha_L, \alpha_H) = v_i x_i - \mathbb{P}_i^{\inf}(\mathbf{x}, \alpha_L, \alpha_H) - c\ (\alpha_L \mathbf{1}\{i \in \mathcal{L}\} + \alpha_H \mathbf{1}\{i \in \mathcal{H}\}), \quad (1)$$

where $v_i \in \{v_L, v_H\}$, and $c > 0$ is a (possibly small) cost of testing for an agent. The first term, $v_i x_i$, represents the utility gain from social activity. The second term, $-\mathbb{P}_i^{\inf}(\mathbf{x}, \alpha_L, \alpha_H)$, is the loss caused by getting infected. The third term is the expected cost of getting tested.[5]

As a solution concept, we use symmetric pure-strategy or mixed-strategy (Nash) equilibrium.

Definition 3. A pure-strategy social activity profile $\mathbf{x}^e$ is a (pure-strategy) Nash equilibrium if

$$u_i(\mathbf{x}^e, \alpha_L, \alpha_H) \geq u_i((x_i, \mathbf{x}^e_{-i}), \alpha_L, \alpha_H), \quad \text{for all } i \in \mathcal{V}, x_i \in [0,1].$$

A mixed-strategy social activity profile takes the form $\mu^e = \prod_{i=1}^n \mu_i^e$, where μ_i^e is a probability distribution over $[0,1]$. A mixed-strategy social activity profile is an equilibrium if

$$\mathbb{E}_{\mathbf{x}^e\sim\mu^e}[u_i(\mathbf{x}^e, \alpha_L, \alpha_H)] \geq \mathbb{E}_{\mathbf{x}^e_{-i}\sim\mu^e_{-i}}[u_i((x_i, \mathbf{x}^e_{-i}), \alpha_L, \alpha_H)], \quad \text{for all } i \in \mathcal{V}, x_i \in [0,1].$$

A *symmetric* pure (mixed) strategy equilibrium is a pure (mixed) strategy equilibrium in which the decision of each agent i only depends on its value v_i and its infection probability and not on its identity.

2.4. Monotonicity and Concavity of Infection Probability

As illustrated in Example 1, the infection probability depends on the graph structure and testing policy in a complex way. Nevertheless, the next lemma shows that the stochastic process of the spread of the infection satisfies natural monotonicity and concavity properties. In what follows for two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$, we write $\mathbf{a} \geq \mathbf{b}$ to denote $a_i \geq b_i$ for $i = 1, \ldots, m$.

Lemma 1. *The infection probability satisfies the following relations.*

(a) *For any agent $i \in \mathcal{V}$, we have*

$$\mathbb{P}_i^{\inf}(\hat{\mathbf{x}}, \alpha_L, \alpha_H) \geq \mathbb{P}_i^{\inf}(\mathbf{x}, \alpha_L, \alpha_H), \quad \text{for all } \hat{\mathbf{x}} \geq \mathbf{x}. \quad (2)$$

(b) *For any agent $i \in \mathcal{V}$, and any social activity profile $\mathbf{x}_{-i} \in [0,1]^{n-1}$, $\mathbb{P}_i^{\inf}((x_i, \mathbf{x}_{-i}), \alpha_L, \alpha_H)$ is concave in x_i.*

Part (a) of Lemma 1 shows that the infection probability for an agent is increasing in the social activity levels of all agents in society, because higher social activity leads to a denser social network over which the virus spreads. Part (b) shows that this probability is concave in the individual's own social activity level, because additional social activity brings the virus to the individual only if its existing links did not do so already.

Proof Sketch of Lemma 1. In the appendix, we prove a somewhat more general version of this lemma. Namely, we demonstrate that for any set of nodes $S \subseteq \mathcal{V}$ that are infected at time 0, the probability of infection reaching node i is increasing in $\mathbf{x}$ and concave in x_i. We denote this probability by $\mathbb{P}_i^{\inf}(\mathbf{x}, \alpha_L, \alpha_H | \text{source} = S)$, and define the following auxiliary infection probability: for any set of nodes S, $\tilde{\mathbb{P}}_i^{\inf}(\mathbf{x}, \alpha_L, \alpha_H | \text{source} = S)$ denotes the probability of infection reaching node i in one round (i.e., only through the nodes in S). We prove the lemma by induction on the size of S that for any set S and $i \in \mathcal{V}$, the probability $\tilde{\mathbb{P}}_i^{\inf}(\mathbf{x}, \alpha_L, \alpha_H | \text{source} = S)$ is increasing in $\mathbf{x}$ and concave in x_i. Next, we express

$\mathbb{P}_i^{\inf}(\mathbf{x}, \alpha_L, \alpha_H | \text{source} = S)$ recursively in terms of the auxiliary probabilities of infection and complete the proof of the lemma using induction on the number of nodes.

2.5. Exogenous Activities: Impact of Testing Policy on Infection Probability

We first analyze the case in which social activity levels are fixed and thus the social (contact) network G is given. The next proposition shows that more testing always reduces the spread of the infection in this case.

Proposition 1. *For any network $G = (\mathcal{V}, E)$, the infection probability of each agent is decreasing in (α_L, α_H). In particular, we have*

$$\mathbb{P}_i^{\inf}(G, \alpha_L', \alpha_H') \le \mathbb{P}_i^{\inf}(G, \alpha_L, \alpha_H),$$
$$\text{for all } i \in \mathcal{V}, (\alpha_L', \alpha_H') \ge (\alpha_L, \alpha_H).$$

The testing policy changes the dynamics of the independent cascade process, affecting the infection probability of agents in a nontrivial way. To prove Proposition 1, we first reformulate the independent cascade process governing the spread of infection in terms of a sequence of independent and identically distributed random variables and then use a coupling argument to relate the infection probabilities under testing policy (α_L, α_H) to the infection probabilities under (α_L', α_H'). This proposition establishes that when social activity levels are given and do not respond to testing, greater testing, that is, higher (α_L, α_H), slows down the spread of the virus and reduces all infection probabilities. This result is intuitive: greater testing enables the detection and isolation of infected individuals, preventing the rapid spread of the virus.

3. Endogenous Activities

With endogenous networks, each individual has a strategic decision: the social activity level (or conversely, her social distancing), which determines the expected number of neighbors this individual will have. Lemma 1 establishes that the utility of each agent i is convex in her social activity x_i. This implies that agents do not choose an intermediate level of activity because the convexity of utility makes mixing between zero and one (strictly) preferable. This observation is formally stated in the next lemma.

Lemma 2. *Given an agent $i \in \mathcal{V}$ and action profile $\mathbf{x}_{-i}$, let x_i^e denote the best response social activity of agent i, that is,*

$$x_i^e \in \arg\max_{x \in [0,1]} u_i(x, \mathbf{x}_{-i}, \alpha_L, \alpha_H).$$

Then, we have $x_i^e \in \{0, 1\}$.

3.1. Equilibrium Characterization

We first introduce some additional notation that will be used in the rest of the paper. We let $\mathbf{1}$ and $\mathbf{0}$ denote the vectors of all 1s and all 0s, respectively, where their dimension will be clear from the context. We also let l and h be a low- and high-value agent, respectively, and define the following sets:

$$\mathcal{A}_1 = \Big\{(\alpha_L, \alpha_H) \in [0,1]^2 : \mathbb{P}_l^{\inf}(\mathbf{x}_{\mathcal{H}} = \mathbf{1}, \mathbf{x}_{\mathcal{L}} = \mathbf{1}, \alpha_L, \alpha_H) \le v_L + \frac{1}{n},$$
$$\mathbb{P}_h^{\inf}(\mathbf{x}_{\mathcal{H}} = \mathbf{1}, \mathbf{x}_{\mathcal{L}} = \mathbf{1}, \alpha_L, \alpha_H) \le v_H + \frac{1}{n}\Big\}$$

$$\mathcal{A}_2 = \Big\{(\alpha_L, \alpha_H) \in [0,1]^2 : \mathbb{P}_l^{\inf}(\mathbf{x}_{\mathcal{H}} = \mathbf{1}, \mathbf{x}_{\mathcal{L}} = \mathbf{1}, \alpha_L, \alpha_H) \ge v_L + \frac{1}{n},$$
$$\mathbb{P}_l^{\inf}(\mathbf{x}_{\mathcal{H}} = \mathbf{1}, x_l = 1, \mathbf{x}_{\mathcal{L}\setminus\{l\}} = \mathbf{0}, \alpha_L, \alpha_H) \le v_L + \frac{1}{n},$$
$$\mathbb{P}_h^{\inf}(\mathbf{x}_{\mathcal{H}} = \mathbf{1}, \mathbf{x}_{\mathcal{L}} = \mathbf{0}, \alpha_L, \alpha_H) \le v_H + \frac{1}{n}\Big\}$$

$$\mathcal{A}_3 = \Big\{(\alpha_L, \alpha_H) \in [0,1]^2 : \mathbb{P}_l^{\inf}(\mathbf{x}_{\mathcal{H}} = \mathbf{1}, \mathbf{x}_{\mathcal{L}} = \mathbf{1}, \alpha_L, \alpha_H) \ge v_L + \frac{1}{n},$$
$$\mathbb{P}_l^{\inf}(\mathbf{x}_{\mathcal{H}} = \mathbf{1}, x_l = 1, \mathbf{x}_{\mathcal{L}\setminus\{l\}} = \mathbf{0}, \alpha_L, \alpha_H) \ge v_L + \frac{1}{n},$$
$$\mathbb{P}_h^{\inf}(\mathbf{x}_{\mathcal{H}} = \mathbf{1}, \mathbf{x}_{\mathcal{L}} = \mathbf{0}, \alpha_L, \alpha_H) \le v_H + \frac{1}{n}\Big\}$$

$$\mathcal{A}_4 = \Big\{(\alpha_L, \alpha_H) \in [0,1]^2 : \mathbb{P}_l^{\inf}(\mathbf{x}_{\mathcal{H}} = \mathbf{1}, \mathbf{x}_{\mathcal{L}} = \mathbf{1}, \alpha_L, \alpha_H) \ge v_L + \frac{1}{n},$$
$$\mathbb{P}_h^{\inf}(\mathbf{x}_{\mathcal{H}} = \mathbf{1}, \mathbf{x}_{\mathcal{L}} = \mathbf{0}, \alpha_L, \alpha_H) \ge v_H + \frac{1}{n}\Big\}.$$

Recall that $\mathcal{H}$ denotes the set of agents with a high value v_H for social activity and $\mathcal{L}$ denotes the set of agents with a low value v_L for social activity.

Proposition 2. *There exist $M \in \mathbb{N}_+$ and functions $\gamma_L : [0,1]^2 \to [0,1]$ and $\gamma_H : [0,1]^2 \to [0,1]$ such that for $n \ge M$ and (α_L, α_H), there are four possibilities for the equilibrium:*

(a) *For $(\alpha_L, \alpha_H) \in \mathcal{A}_1$, there exists a unique symmetric equilibrium, where $x_i^e = 1$ for all $i \in \mathcal{V}$.*

(b) *For $(\alpha_L, \alpha_H) \in \mathcal{A}_2$, there exists a unique symmetric equilibrium, where $x_i^e = 1$ for all $i \in \mathcal{H}$ and a mixed-strategy μ_j^e for all $j \in \mathcal{L}$ that puts probability $\gamma_L(\alpha_L, \alpha_H)$ on one and probability $1 - \gamma_L(\alpha_L, \alpha_H)$ on zero.*

(c) *For $(\alpha_L, \alpha_H) \in \mathcal{A}_3$, there exists a unique symmetric equilibrium, where $x_i^e = 1$ for all $i \in \mathcal{H}$ and $x_j^e = 0$ for all $j \in \mathcal{L}$.*

(d) *For $(\alpha_L, \alpha_H) \in \mathcal{A}_4$, there exists a unique symmetric equilibrium, where $x_j^e = 0$ for all $j \in \mathcal{L}$ and a mixed-strategy μ_i^e for all $i \in \mathcal{H}$ that puts probability $\gamma_H(\alpha_L, \alpha_H)$ on one and probability $1 - \gamma_H(\alpha_L, \alpha_H)$ on zero.*

Proposition 2 characterizes the equilibrium outcome for different testing policies (α_L, α_H). For the case in which the number of agents in society is greater than a threshold, it divides testing policies into four regions, each leading to a different type of equilibrium behavior. For example, when testing policy is in region $\mathcal{A}_1$, the infected are sufficiently likely to be identified and

isolated, so that all individuals choose full social activity—as if social contacts did not increase their probability of infection. Outside this region, individuals take precautionary action by reducing their social activity.[6] In region $\mathcal{A}_2$, high-value agents (who receive greater utility from social activity) still choose full activity, but now low-value agents mix between no activity and full activity. The other regions are defined similarly. In the proof of Proposition 2, we show that the four regions $\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{A}_3$, and $\mathcal{A}_4$ cover all possible testing policies $(\alpha_L, \alpha_H) \in [0,1]^2$ (for sufficiently large n).

The regions (or sets) $\mathcal{A}_1$ to $\mathcal{A}_4$ are functions of the model primitives, but they depend on the infection probability in the stochastic process governing the spread of infection, which is a complex quantity. Nonetheless, the next lemma shows that as $n \to \infty$, these sets have a simple characterization. For this lemma, recall that r_H and r_L denote the fractions of the two types of individuals. This characterization will be used in Section 4 to study the optimal testing policy.

Lemma 3. *There exist $\alpha_H^{(1)} \geq \alpha_H^{(2)} \geq \alpha_H^{(3)}$ such that as $n \to \infty$ the sets $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_4$ converge to*[7]

$$\mathcal{A}_1^* = \{(\alpha_L, \alpha_H) \in [0,1]^2 : \alpha_L r_L + \alpha_H r_H \geq \alpha_H^{(1)} r_H\},$$

$$\mathcal{A}_2^* = \{(\alpha_L, \alpha_H) \in [0,1]^2 : \alpha_L r_L + \alpha_H r_H \leq \alpha_H^{(1)} r_H, \alpha_H \geq \alpha_H^{(2)}\},$$

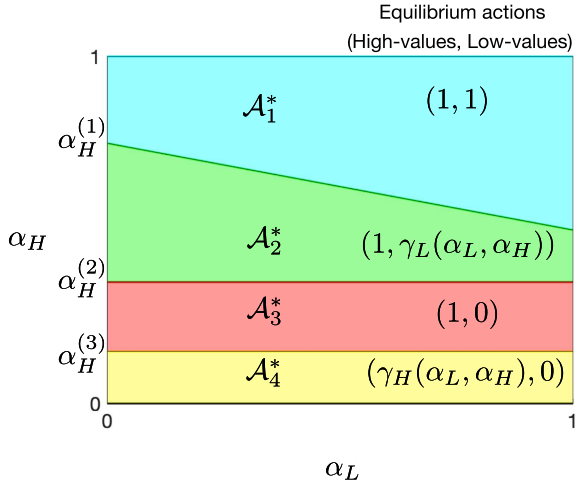
$$\mathcal{A}_3^* = \{(\alpha_L, \alpha_H) \in [0,1]^2 : \alpha_L r_L + \alpha_H r_H \leq \alpha_H^{(1)} r_H, \alpha_H^{(3)} \leq \alpha_H \leq \alpha_H^{(2)}\}, \text{ and}$$

$$\mathcal{A}_4^* = \{(\alpha_L, \alpha_H) \in [0,1]^2 : \alpha_L r_L + \alpha_H r_H \leq \alpha_H^{(1)} r_H, \alpha_H \leq \alpha_H^{(3)}\},$$

respectively.

The four sets in Lemma 3 are depicted in Figure 2. This figure additionally shows the equilibrium action profiles of high- and low-value agents as a function of the pair (α_L, α_H).

Figure 2. (Color online) Regions $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3$, and $\mathcal{A}_4$ (as $n \to \infty$) That Determine the Equilibrium in Proposition 2



To obtain an intuition of the proof of Lemma 3, let us consider set $\mathcal{A}_1$ for which both type of agents are playing one; therefore, their testing probability affects the infection probabilities of other agents. The convergence of the sets to these asymptotic objects follows from the fact that for large enough n, the number of tests will be concentrated around its mean, which is $\alpha_L r_L + \alpha_H r_H$ and the infection probability is decreasing in the number of tests. The proof does not readily follow from the law of large numbers, however, because the infection probability is a nonlinear function of the number of tests. To establish the lemma, we develop a "peeling argument" that uses the properties of the infection probability (such as submodularity) together with a concentration bound. More precisely, consider a society with n agents and let $Q(n,k)$ denote the infection probability of agents when $\mathbf{x} = \mathbf{1}$ and k out of n agents are tested. We show that

$$\lim_{n\to\infty} |\mathbb{P}_i^{\inf}(\mathbf{x} = \mathbf{1}, \alpha_L, \alpha_H) - Q(n, \lceil (\alpha_L r_L + \alpha_H r_H) n \rceil)| = 0, \quad \text{for all } i.$$

To establish this result, we first use the Chernoff-Hoeffding inequality, showing with a high probability the number of tested individuals is around $\lceil n(\alpha_L r_L + \alpha_H r_H) \rceil$. We then use the submodularity of the infection probability combined with the peeling argument to prove that for any small $\epsilon > 0$ and $k \in (\lfloor n(\alpha_L r_L + \alpha_H r_H - \epsilon) \rfloor, \lceil n(\alpha_L r_L + \alpha_H r_H + \epsilon) \rceil)$, we have

$$|Q(n,k) - Q(n, \lceil n(\alpha_L r_L + \alpha_H r_H - \epsilon) \rceil)| \leq \epsilon + \frac{\epsilon}{1 - (\alpha_L r_L + \alpha_H r_H)}.$$

In our baseline analysis, we focus on symmetric equilibria to simplify the notation (i.e., the strategy of an agent depends on the infection probability from its perspective and not its identity). In the online appendix, we show that the equilibrium characterization given in Proposition 2 is essentially unique. In particular, we show that the pure-strategy equilibrium in parts (a) and (c) are unique. We then characterize the asymmetric pure-strategy equilibrium (for parts (b) and (d)) and show that for large n, the expected number of infected individuals in any asymmetric pure-strategy equilibrium is the same as the expected number of infected individuals in the symmetric equilibrium characterized in Proposition 2.

3.2. Impact of Testing Policy on Infection Probability

The next theorem presents one of our main results: the nonmonotonic impact of greater testing on infections.

Theorem 1. *There exists $M \in \mathbb{N}_+$ such that for $n \geq M$, in the unique symmetric equilibrium we have:*

(a) *Higher (α_L, α_H) in the interior of $\mathcal{A}_1$ decreases the infection probabilities of both types of agents.*

(b) *Higher (α_L, α_H) in the interior of $\mathcal{A}_2$ increases the infection probability of low-value agents and does not change the infection probability of high-value agents.*

(c) *Higher (α_L, α_H) in the interior of $\mathcal{A}_3$ decreases the infection probabilities of both types of agents.*

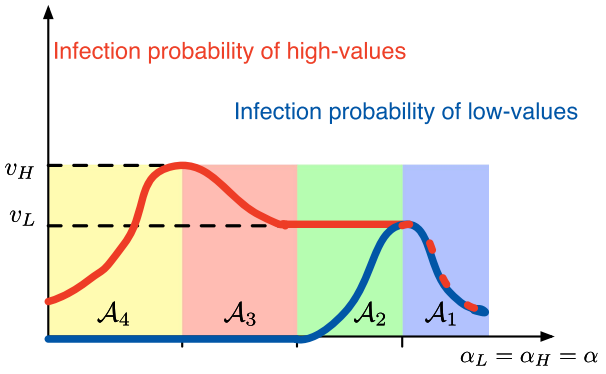
(d) *Higher (α_L, α_H) in the interior of $\mathcal{A}_4$ increases the infection probability of high-value agents and does not change the infection probability of low-value agents.*

Moreover, infection probabilities are continuous in (α_L, α_H) at the boundaries of the previous sets.

Theorem 1 establishes that whenever we have a mixed-strategy equilibrium, as in regions $\mathcal{A}_2$ and $\mathcal{A}_4$, the effects of greater testing on infections are nonmonotone. Intuitively, this is because greater testing and isolation makes agents who are mixing wish to go to full activity. However, when we are (and remain) in the interior of the sets $\mathcal{A}_2$ and $\mathcal{A}_4$, full activity is not an equilibrium (either for the low-value or the high-value agents). Hence, equilibrium is restored by some more of the relevant agents choosing high activity at the margin, which increases contacts and thus restores the incentives for mixing by increasing infection probabilities. In both cases, as Proposition 1 highlights, with given activity levels, greater testing would reduce the spread of the infection. The reason why the infection spreads more is because greater social activity levels make the social network denser. To see that the infection probability increases as testing rises, let us consider $(\alpha_L, \alpha_H) \in \mathcal{A}_4$. As we showed in Proposition 2, in this region, low-value agents play zero and high-value agents play a mixed strategy that puts probability $\gamma_H(\alpha_L, \alpha_H)$ on activity level 1 and probability $1 - \gamma_H(\alpha_L, \alpha_H)$ on activity level 0. Therefore, in this region the infection probability of high-value agents is $\mathbb{P}_h(\mathbf{x}_{\mathcal{H}} = \gamma_H(\alpha_L, \alpha_H)\mathbf{1}, \mathbf{x}_{\mathcal{L}} = \mathbf{0}, \alpha_L, \alpha_H)$. Changing the testing policy (α_L, α_H) affects this probability in two ways: (i) it alters the testing probability of agents in the governing stochastic process of the infection, and (ii) it changes the equilibrium social activity of high-value agents. Using the fact that mixing with probability $\gamma_H(\alpha_L, \alpha_H)$ is an equilibrium strategy for high-value agents, we show that this infection probability can be written in closed form as $\gamma_H(\alpha_L, \alpha_H)v_H + \frac{1}{n}$. The proof is completed by showing that $\gamma_H(\alpha_L, \alpha_H)$ is increasing in the testing policy (α_L, α_H).

In practice, it may not be possible to test different types of agents at different rates. If so, we would have to impose $\alpha_L = \alpha_H = \alpha$ for some $\alpha \in [0,1]$. Figure 3 depicts the infection probability of both types as a function of α, confirming that infection probabilities continue to be nonmonotonic in testing probabilities (in this case α).

Figure 3. (Color online) Schematic View of the Infection Probability of High- and Low-Value Agents in the Limit as $n \to \infty$ for Uniform Policy (i.e., $\alpha_H = \alpha_L = \alpha$) as α Increases



Theorem 1 reiterates the importance of the sets $\mathcal{A}_1, \ldots, \mathcal{A}_4$ in our analysis. We next provide a comparative statics for these sets as η (the probability of match conditional on activities), β (the transmission rate of the infection), and p (where βp is the transmission rate of tested individuals) vary.

Lemma 4. *Let $\mathcal{A}_1(\eta, \beta, p)$ and $\mathcal{A}_4(\eta, \beta, p)$ denote the sets $\mathcal{A}_1$ and $\mathcal{A}_4$, respectively, as a function of the parameters η, β, and p. We have:*

(a) *Higher η shrinks the set $\mathcal{A}_1$ and expands the set $\mathcal{A}_4$. That is,*

$$\mathcal{A}_1(\eta', \beta, p) \subseteq \mathcal{A}_1(\eta, \beta, p) \text{ and } \mathcal{A}_4(\eta, \beta, p) \subseteq \mathcal{A}_4(\eta', \beta, p), \quad \text{for all } \eta' \geq \eta, \beta, p.$$

(b) *Higher β shrinks the set $\mathcal{A}_1$ and expands the set $\mathcal{A}_4$. That is,*

$$\mathcal{A}_1(\eta, \beta', p) \subseteq \mathcal{A}_1(\eta, \beta, p) \text{ and } \mathcal{A}_4(\eta, \beta, p) \subseteq \mathcal{A}_4(\eta, \beta', p), \quad \text{for all } \beta' \geq \beta, \eta, p.$$

(c) *Higher p shrinks the set $\mathcal{A}_1$ and expands the set $\mathcal{A}_4$. That is,*

$$\mathcal{A}_1(\eta, \beta, p') \subseteq \mathcal{A}_1(\eta, \beta, p) \text{ and } \mathcal{A}_4(\eta, \beta, p) \subseteq \mathcal{A}_4(\eta, \beta, p'), \quad \text{for all } p' \geq p, \beta, \eta.$$

This lemma directly follows from the definition of the sets $\mathcal{A}_1$ and $\mathcal{A}_4$ and the fact that the infection probabilities are increasing in η, β, and p. The boundary between sets $\mathcal{A}_2$ and $\mathcal{A}_3$ (for large enough n) also shift up as we increase either η, β, or p. Conversely, the sets $\mathcal{A}_2$ and $\mathcal{A}_3$ can either shrink or expand.

We conclude this section by noting that all the results presented thus far, as well as the broad outlines of the optimal testing results in the next section, hold for any contagion process that satisfies the properties stated in Lemma 1 and Proposition 1. Put differently, provided that for any $i \in \mathcal{V}$, $\mathbb{P}_i^{\inf}(\mathbf{x}, \alpha_L, \alpha_H)$ is increasing in $\mathbf{x}$,

concave in x_i, and decreasing in (α_L, α_H), the equilibrium characterization and the nonmonotonicity of infection probability in testing policy continue to apply.

4. Optimal Testing Policy

We now discuss the design of testing policy to maximize social welfare.

4.1. Fully Optimal Testing Policy

Throughout, we assume that there is a limited capacity to test, represented by θn, where $\theta \in [0,1]$ and n is the number of individuals. We refer to θ as *testing capacity*. Social welfare is

$$W(\mathbf{x}, \alpha_L, \alpha_H) = \sum_{i=1}^{n} u_i(\mathbf{x}, \alpha_L, \alpha_H).$$

The choice of testing policy induces a two-stage game between the planner and the agents, with the following timing:

1. The social planner chooses the testing policy (α_L, α_H).
2. Given the testing policy, the unique symmetric equilibrium from Proposition 2 is played by the agents.

With this timing and notation, the social planner's problem becomes

$$\max_{(\alpha_L, \alpha_H) \in [0,1]^2} \quad W(\mathbf{x}^e, \alpha_L, \alpha_H)$$
$$\text{s.t.} \quad \mathbf{x}^e \text{ is the unique (symmetric) equilibrium,}$$
$$\alpha_H |\mathcal{H}| + \alpha_L |\mathcal{L}| \le \theta n.$$

Our main result of this section, stated next, characterizes the optimal testing policy.

Theorem 2. *There exist $M \in \mathbb{N}_+$, $\overline{c} \in \mathbb{R}_+$, and $\theta^{(1)} \ge \theta^{(2)} \ge \theta^{(3)}$ such that for $c \le \overline{c}$ and $n \ge M$ we have:*

(a) *If $\theta > \theta^{(1)}$, then the optimal testing policy is to test all individuals with probability θ, and in the corresponding equilibrium, all agents are fully active; that is, $x_i^e = 1$ for all $i \in \mathcal{V}$.*

(b) *If $\theta^{(2)} < \theta < \theta^{(1)}$, then the optimal testing policy is to test only high-value individuals with probability $\frac{\theta^{(2)}}{r_H}$, and in the corresponding equilibrium, high-value agents are fully active, and low-value agents are inactive; that is, $x_i^e = 1$ for all $i \in \mathcal{H}$, and $x_j^e = 0$ for all $j \in \mathcal{L}$.*

(c) *If $\theta^{(3)} < \theta < \theta^{(2)}$, then the optimal testing policy is to test only high-value individuals with probability $\frac{\theta}{r_H}$, and in the corresponding equilibrium, high-value agents are fully active, and low-value agents are inactive; that is, $x_i^e = 1$ for all $i \in \mathcal{H}$, and $x_j^e = 0$ for all $j \in \mathcal{L}$.*

(d) *If $\theta < \theta^{(3)}$, then the optimal testing policy is to have zero tests, and in the corresponding equilibrium, all agents are inactive; that is, $x_i^e = 0$ for all $i \in \mathcal{V}$.*[8]

The most important result in Theorem 2 is that the optimal policy does not necessarily use all available tests. In particular, when there are enough tests so that all agents can be fully active with the appropriate testing in isolation, the social planner is happy to deploy all testing and allow all agents to be fully active (see Figure 4). This is the case when $\theta > \theta^{(1)}$. However, when $\theta^{(2)} < \theta < \theta^{(1)}$, the social planner prefers not to use all available tests. The intuition for this result is related to the nonmonotonicity of the comparative statics derived in Theorem 1: greater testing will encourage more social activity, in this case from low-value agents. The social planner, on the other hand, prefers zero activity from low-value agents so as to slow down the spread of the virus. Therefore, they opt for a policy that does not test low-value agents, thus discouraging their social activity and keeping the social network less dense. When $\theta < \theta^{(3)}$, the optimal policy is even more extreme. It does not test any agents. This is because just testing high-value agents would encourage sufficient social activity to lead to faster spread of the virus, which the social planner prefers to avoid.

One implication of our model, highlighted in Theorem 2, is that testing high-value agents ahead of low-value agents is optimal. In Section 5, we show that this is a consequence of individuals choosing a general social activity level. If, instead, the individuals could target their social activity to low-value and high-value agents separately, then the social planner might want to test low-value agents and refrain from testing high-value agents in order to discourage agents from interacting with these "super spreaders" (who are socially very active).

4.2. Optimal Uniform Testing Policy

It may be impossible for authorities to discriminate between or identify high-value and low-value agents, in which case, testing policy would have to be uniform, $\alpha_L = \alpha_H = \alpha$. With a uniform testing policy the social planner's problem becomes

$$\max_{\alpha \in [0,1]} \quad W(\mathbf{x}^e, \alpha_L = \alpha, \alpha_H = \alpha)$$
$$\text{s.t.} \quad \mathbf{x}^e \text{ is the unique (symmetric) equilibrium,}$$
$$\alpha \le \theta.$$

The following corollary follows from Proposition 2 and characterizes the optimal uniform policy.

Corollary 1. *Let $\theta^{(1)} \ge \theta^{(2)} \ge \theta^{(3)}$ be thresholds found in Theorem 2. There exist $\overline{c} \in \mathbb{R}_+$ and $M \in \mathbb{N}_+$ such that for $c \le \overline{c}$ and $n \ge M$, we have:*

(a) *If $\theta > \theta^{(1)}$, then the optimal policy is to test all individuals with probability θ, and in the corresponding equilibrium, both agent types have full social activity; that is, $x_i^e = 1$ for all $i \in \mathcal{V}$.*

(b) *If $\min\left\{\frac{\theta^{(2)}}{r_H}, \theta^{(1)}\right\} < \theta < \theta^{(1)}$, then the optimal policy is to test all individuals with probability $\min\left\{\frac{\theta^{(2)}}{r_H}, \theta^{(1)}\right\}$, and in the corresponding equilibrium, high-value agents*

have full social activity, whereas low-value agents have zero social activity; that is, $x_i^e = 1$ for all $i \in \mathcal{H}$, and $x_j^e = 0$ for all $j \in \mathcal{L}$.

(c) *If $\min\left\{\frac{\theta^{(3)}}{r_H}, \theta^{(1)}\right\} < \theta < \min\left\{\frac{\theta^{(2)}}{r_H}, \theta^{(1)}\right\}$, then the optimal policy is to test all agents with probability θ, and in the corresponding equilibrium, high-value agents have full social activity, whereas low-value agents have zero social activity; that is, $x_i^e = 1$ for all $i \in \mathcal{H}$, and $x_j^e = 0$ for all $j \in \mathcal{L}$.*

(d) *If $\theta < \min\left\{\frac{\theta^{(3)}}{r_H}, \theta^{(1)}\right\}$, then the optimal policy is to have zero tests in the corresponding equilibrium, and all agents have zero social activity; that is, $x_i^e = 0$ for all $i \in \mathcal{V}$.*

Comparing Corollary 1 with Theorem 2, we see that the top region, in which all agents are fully active, does not change. The second region shrinks (because $\frac{\theta^{(2)}}{r_H} > \theta^{(2)}$) and involves greater spread of the virus, because uniform policies are less effective at identifying and isolating the "super spreader" agents who are socially more active thus more likely to be infected and then spread the virus (because of their greater social activity). The third region may expand, but in this case, individuals continue to have the same infection probability as they did under targeted policies. Finally, the fourth region expands, and in this region, individuals have the same infection probability as under a targeted testing policy. Overall, uniform policies make testing less effective, but do not change our qualitative conclusions.

4.3. Optimal Testing Policy with Mandatory Social Distancing

The nonmonotonicity in our comparative statics and the unwillingness of the social planner to always use all testing capacity are related to the fact that greater testing reduces voluntary social distancing. This naturally suggests that testing should be combined with mandatory social distancing. The next proposition establishes that when this is the case, the social planner would always like to use all available testing capacity and would then deploy mandatory social distancing measures to limit the adverse behavioral effects of testing.

Formally, we suppose the social planner, in addition to the testing policy (α_L, α_H), can choose $\overline{x}_L$ and $\overline{x}_H$ indicating the maximum social activity levels of low- and high-value agents, respectively. We refer to such a policy, denoted by $(\alpha_L, \alpha_H, \overline{x}_L, \overline{x}_H)$, as a *testing policy with mandatory social distancing*. With this notation, the social planner's problem becomes

$$\begin{aligned} \max_{(\alpha_L, \alpha_H, \overline{x}_L, \overline{x}_H) \in [0,1]^4} \quad & W(\mathbf{x}^e, \alpha_L, \alpha_H) \\ \text{s.t.} \quad & \mathbf{x}^e \text{ is the unique (symmetric) equilibrium,} \\ & x_i \le \overline{x}_H \text{ for } i \in \mathcal{H}, x_j \le \overline{x}_L \text{ for } j \in \mathcal{L}, \\ & \alpha_H |\mathcal{H}| + \alpha_L |\mathcal{L}| \le \theta n. \end{aligned}$$

For a given testing capacity θ, we denote the first best by $(\alpha_L^{\mathrm{FB}}(\theta), \alpha_H^{\mathrm{FB}}(\theta), x_l^{\mathrm{FB}}(\theta), x_h^{\mathrm{FB}}(\theta))$, which is the solution of

$$\begin{aligned} \max_{(\alpha_L, \alpha_H, x_l, x_h) \in [0,1]^4} \quad & W((\mathbf{x}_{\mathcal{H}} = x_h \mathbf{1}, \mathbf{x}_{\mathcal{L}} = x_l \mathbf{1}), \alpha_L, \alpha_H) \\ \text{s.t.} \quad & \alpha_H |\mathcal{H}| + \alpha_L |\mathcal{L}| \le \theta n. \end{aligned}$$

Proposition 3. *For any testing capacity θ, a testing policy with mandatory social distancing with*

$$(\alpha_L, \alpha_H, \overline{x}_L, \overline{x}_H) = (\alpha_L^{\mathrm{FB}}(\theta), \alpha_H^{\mathrm{FB}}(\theta), x_l^{\mathrm{FB}}(\theta), x_h^{\mathrm{FB}}(\theta))$$

achieves the social welfare of the first best. Moreover, with this policy, the social planner uses all the testing capacity.

5. Extensions

In this section, we consider three extensions. First, we show that all of our results extend to an environment with multiple types, and nonmonotonicities become more likely in this case. Second, we allow for social activity levels directed to different types of agents (for example, individuals choosing how much to interact with more active/popular agents and how much to interact with other agents). We show that with such directed social activity behavior, optimal policy can try to discourage individuals from interacting with high-value agents who are more likely to spread the virus. Third, we characterize the optimal testing policy when there is no limit on the number of tests, but the tests are costly. We establish that the nonmonotonicity of the infection probability in testing implies that even when the cost of testing is small, the optimal policy may not involve testing all agents.

5.1. Multiple Type of Individuals

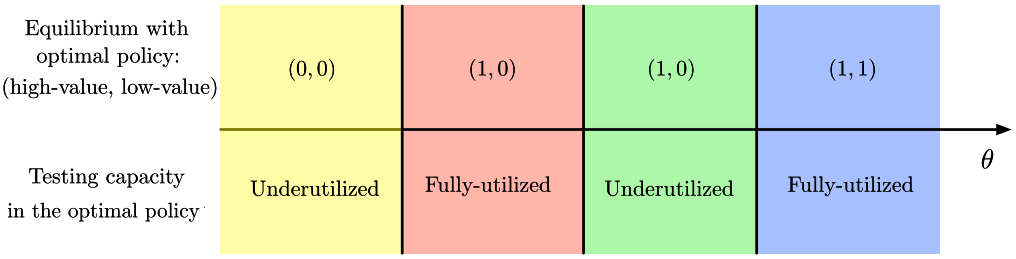
In our baseline model, we considered two types of individuals with different values for social activity. Here, we show that our main results and similar insights carry over to a more general setting with m different values for individuals. In particular, we let $v_1 < v_2 < \cdots < v_m$ denote the social activity value of different individual types and let $\mathcal{V}_k$ for $k = 1, \ldots, m$ denote the set of individuals of type k (therefore, $\bigcup_{k=1}^m \mathcal{V}_k = \mathcal{V}$). We also denote the testing probability of type k individuals by α_k for $k = 1, \ldots, m$ and let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m) \in [0,1]^m$ denote the testing policy.

In the online appendix, in Section B.2, we extend Proposition 2 and Theorem 1 to $m \in \mathbb{N}$ types of individuals. In particular, we establish that the nonmonotonicity of infection probability in testing policy continues to hold, and there will now be more regions where greater testing increases the spread of the infection.

5.2. Directed Social Interactions

Here, we show that if agents choose two levels of social activity, one directed to low-value and the other to high-value agents, then the optimal policy may involve testing low-value agents with a higher probability than high-value agents. In particular, we let the social

Figure 4. (Color online) Depending on θ, the Fraction of Society That Can Be Tested, There Are Four Possibilities for the Optimal Testing Policy and Corresponding Social Activity Profile in Equilibrium



activity of each agent i be a pair $x_i = (x_i^L, x_i^H)$, where x_i^L denotes agent i's social activity directed to low-value agents, and x_i^H denotes her social activity directed to high-value agents. In the network of contacts, the probability of an edge between agents i and j is therefore

$$\mathbb{P}[\mathbf{E}_{ij} = 1] = \begin{cases} \eta x_i^H x_j^H, & i,j \in \mathcal{H}, \\ \eta x_i^H x_j^L, & i \in \mathcal{H}, j \in \mathcal{L}, \\ \eta x_i^L x_j^H, & i \in \mathcal{L}, j \in \mathcal{H}, \\ \eta x_i^L x_j^L, & i,j \in \mathcal{L}. \end{cases}$$

The utility of agent i is similar to our baseline model and is given by

$$u_i(x_i, \mathbf{x}_{-i}, \alpha_L, \alpha_H) = v_i(x_i^L + x_i^H) - \mathbb{P}_i^{\inf}(x_i, \mathbf{x}_{-i}, \alpha_L, \alpha_H) - c\ (\alpha_L \mathbf{1}\{i \in \mathcal{L}\} + \alpha_H \mathbf{1}\{i \in \mathcal{H}\}).$$

The next proposition provides conditions under which it is optimal to test only low-value agents.

Proposition 4. *If* $r_H > \max\left\{\frac{1}{1+\beta}, \frac{v_H}{\beta}, 1 - v_L\right\}$ *and* $\theta < r_H - \frac{v_H}{\beta}$, *then there exist* $\overline{c}$ *and* $M \in \mathbb{N}_+$ *such that for* $n \geq M$ *and* $c \leq \overline{c}$, *the optimal policy is to only test low-value agents; in the corresponding symmetric unique equilibrium, all low-value agents play* $x_j = (1,0)$, *and all high-value agents play a mixed-strategy between* $(1,0)$ *and* $(1,1)$.

Proposition 4 proves that, for a sufficiently small testing capacity θ and a sufficiently large population fraction of high-value agents r_H, it is optimal to only test low-value agents; in the corresponding equilibrium, low-value agents will not interact with high-value agents. The intuition is that high-value agents, who are socially more active, act as "super spreaders," and the social planner would like to reduce their interactions with low-value agents. This was not possible in our baseline model because agents could not direct their social activity toward different groups. When such directed behavior is introduced, this encourages the social planner to reduce the testing of high-value agents to discourage low-value agents from interacting with them too much.

5.3. Optimal Testing Policy with Costly Tests

Our analysis thus far has considered a situation with a limited number of available tests. In this section, we study optimal testing policy (α_L, α_H) when the social planner has access to an unlimited supply of tests at a fixed marginal cost (i.e., the cost is linear in the (expected) number of agents who are tested). Social welfare in this case is simply

$$\mathrm{W}(\mathbf{x}, \alpha_L, \alpha_H) = \left(\sum_{i=1}^n u_i(\mathbf{x}, \alpha_L, \alpha_H)\right) - t\ (\alpha_H |\mathcal{H}| + \alpha_L |\mathcal{L}|),$$

where t is the per capita cost of testing.

The game among the planner and the agents is identical to before, and the social planner's problem is

$$\max_{(\alpha_L, \alpha_H) \in [0,1]^2} \mathrm{W}(\mathbf{x}^e, \alpha_L, \alpha_H)$$
$$\text{s.t.} \quad \mathbf{x}^e \text{ is the unique (symmetric) equilibrium,}$$

which only differs from our baseline social welfare maximization problem because the objective incorporates costs of testing, and the capacity constraint is removed.

The following proposition characterizes the optimal policy.

Proposition 5. *There exist* $M \in \mathbb{N}_+$, $\overline{c}$, $\hat{t}$, *and* $\hat{r} \in \mathbb{R}_+$ *such that for* $n \geq M$ *and* $c \leq \overline{c}$, *we have:*

(a) *If* $t > \hat{t}$, *then the optimal policy is to have zero tests.*

(b) *If* $t < \hat{t}$ *and* $r_H > \hat{r}$, *then the optimal testing is to test only high-value individuals, and in the corresponding equilibrium, high-value agents are fully active, and low-value agents are inactive; that is,* $x_i^e = 1$ *for all* $i \in \mathcal{H}$, *and* $x_j^e = 0$ *for all* $j \in \mathcal{L}$.

(c) *If* $t < \hat{t}$ *and* $r_H < \hat{r}$, *then the optimal policy is to test all individuals, and in the corresponding equilibrium, all agents are fully active; that is,* $x_i^e = 1$ *for all* $i \in \mathcal{V}$.

Proposition 5 establishes that when the cost of testing is not too large and testing occurs in the optimal policy, it is always optimal to ensure that high-value agents are fully active. If the fraction of high-value agents is large, then it is optimal to choose a testing policy

whereby only high-value agents can be fully active, while low-value agents remain inactive. Otherwise, if the fraction of high-value agents is small, then it is optimal to choose a testing policy that guarantees full social activity for all agents. There is no testing capacity left unused in this case, since producing this capacity is costly.

6. Concluding Remarks

This paper studied the effects of testing on social activity and voluntary social distancing in the context of an epidemic. Social activity levels determine the (endogenous) social network over which contacts take place and an infection spreads. Testing enables authorities to identify and isolate infected individuals who spread the virus and has been identified by the recent literature on COVID-19 and policymakers as a key tool for combating epidemics. Our analysis, however, shows that the impact of testing on the spread of an epidemic may be more complex because knowing that tests will lead to the isolation of infected individuals, agents can increase their social activity levels and refrain from voluntary social distancing. As a result, our analysis has established that the effects of testing on the spread of the infection can be nonmonotonic—greater testing can lead to higher infection probabilities.

Our analysis has also characterized the optimal testing policies. The same forces that lead to nonmonotonic comparative statics also imply that a benevolent social planner may prefer to leave their testing capacity partially or fully unused—because increasing testing can make the spread of the virus more likely. This implies that testing should often be combined with mandatory social distancing measures, which ensure that the adverse behavioral effects of testing can be countered by preventing excessively high social activity levels.

Our paper is part of a growing literature on the interaction between economic incentives and epidemiological dynamics. Two high-level contributions of our approach are to conceptualize the problem of endogenous behavior as one of social network formation and to use the percolation model rather than the SIR dynamic model. Both of these contributions can be useful beyond the confines of our specific question, but the robustness of our conclusions to relaxing both assumptions and adopting different modeling strategies needs to be investigated. Other interesting areas for research include the analysis of optimal testing and tracing when tests lead to type I and type II errors and policy is constrained by privacy considerations and nonobedience (both in acquiescing to testing and following mandatory social distancing guidelines). Another interesting direction is to consider different types of activities with different values for individuals. Finally, a fruitful avenue would be to enrich the setup to incorporate more heterogeneity and richer economic, social and epidemiological interactions so as to enable quantitative policy analysis.

Acknowledgments

The authors thank the entire review team for the many fruitful comments and suggestions that improved both the exposition and results of the paper.

Appendix A. Proofs

This appendix includes the proof of the main results. The rest of the proofs and additional results are presented in the online appendix.

Proof of Proposition 2. We first prove that the best response decision of each individual is either one or zero (i.e., the best response social activity of each agent is at the boundaries of $[0,1]$).

Lemma 2 also shows that in any mixed-strategy equilibrium each agent i must be mixing between playing one and playing zero. In what follows, we let $x=\gamma$ for some $\gamma\in[0,1]$ to denote a mixed strategy that puts probability γ on one and puts probability $1-\gamma$ on zero. We also let α_i denote the testing probability of agent I, which is equal to α_L if $i\in\mathcal{L}$ and is equal to α_H if $i\in\mathcal{H}$.

We now proceed with the proof of proposition.

Proof of Part (a). We first establish that $x_i=1$ for all $i\in\mathcal{V}$ is an equilibrium and then show that it is the unique symmetric equilibrium. Consider $i\in\mathcal{H}$. The utility of i with action profile $\mathbf{x}=\mathbf{1}$ is $v_H-\mathbb{P}_i^{\inf}(x_i=1,x_{-i}=1,\alpha_H,\alpha_L)-c\alpha_H$. If agent i deviates and plays $x_i=0$ (using Lemma 2, this is the only candidate for a profitable deviation), then its utility becomes $-\mathbb{P}_i^{\inf}(x_i=0,x_{-i}=1,\alpha_H,\alpha_L)-c\alpha_H=\frac{-1}{n}-c\alpha_H$, where we used the fact that if $x_i=0$, then the only way for agent i to get infected is be the source of infection. Therefore, given $(\alpha_H,\alpha_L)\in\mathcal{A}_1$ we have $u_i(x_i=1,\mathbf{x}_{-i},\alpha_L,\alpha_H)>u_i(x_i=0,\mathbf{x}_{-i},\alpha_L,\alpha_H)$. Similarly, any $j\in\mathcal{L}$ does not have a profitable deviation.

We next prove that this is the unique symmetric equilibrium. First, for a high-value agent $i\in\mathcal{H}$ no matter what the strategy of other agents are, the dominant strategy is to play $x_i=1$. This is because for any $\mathbf{x}_{-i}$, we can write $u_i(x_i=1,\mathbf{x}_{-i},\alpha_L,\alpha_H)=v_H-\mathbb{P}_i^{\inf}(x_i=1,\mathbf{x}_{-i},\alpha_L,\alpha_H)-c\alpha_H\geq v_H-\mathbb{P}_i^{\inf}(x_i=1,\mathbf{x}_{-i}=\mathbf{1},\alpha_L,\alpha_H)-c\alpha_H>-\frac{1}{n}-c\alpha_H=u_i(x_i=0,\mathbf{x}_{-i},\alpha_L,\alpha_H)$, where the first inequality follows from Part (a) of Lemma 1 and the second inequality follows from $(\alpha_L,\alpha_H)\in\mathcal{A}_1$ and $v_H>v_L$. Therefore, there are two other candidates for a symmetric equilibrium: (i) the action profile $x_i=1$ for all $i\in\mathcal{H}$ and $x_j=0$ for all $j\in\mathcal{L}$ and (ii) $x_i=1$ for all $i\in\mathcal{H}$ and $x_j=\gamma_L$ for all $j\in\mathcal{L}$ and some $\gamma_L\in(0,1)$. We next show that none of these can be an equilibrium because an agent $j\in\mathcal{L}$ has a profitable deviation to one. This is because $u_j(x_j=1,\mathbf{x}_{L\setminus\{j\}},\mathbf{x}_{\mathcal{H}}=\mathbf{1},\alpha_H,\alpha_L)=v_L-\mathbb{P}_j^{\inf}(x_j=1,\ \mathbf{x}_{L\setminus\{j\}},\mathbf{x}_{\mathcal{H}}=\mathbf{1},\alpha_H,\alpha_L)-c\alpha_L>v_L-\mathbb{P}_j^{\inf}(x_j=1,\mathbf{x}_{L\setminus\{j\}}=\mathbf{1},\mathbf{x}_{\mathcal{H}}=\mathbf{1},\alpha_H,\alpha_L)-c\alpha_L\geq\frac{-1}{n}-c\alpha_L=u_j(x_j=0,\mathbf{x}_{L\setminus\{j\}}=\mathbf{0},\mathbf{x}_{\mathcal{H}}=\mathbf{1},\alpha_H,\alpha_L)$, where the first inequality follows from Part (a) of Lemma 1 and the fact that we have either $\mathbf{x}_{L\setminus\{j\}}=\mathbf{1}$ or $\mathbf{x}_{L\setminus\{j\}}=\gamma_L\mathbf{1}$, which are both below $\mathbf{x}_{L\setminus\{j\}}=\mathbf{1}$ and the second inequality follows from $(\alpha_L,\alpha_H)\in\mathcal{A}_1$, completing the proof of Part (a).

Proof of Part (b). Before proceeding with the proof of this part, as we decrease (α_L, α_H), one of the constraints of $\mathcal{A}_1$ will be violated because the infection probability increases (using Proposition 1). For any $\epsilon > 0$, there exists $M \in \mathbb{N}_+$ such that for $n \geq M$, we have $|\mathbb{P}_l^{\inf}(\mathbf{x}_{\mathcal{H}} = \mathbf{1}, \mathbf{x}_{\mathcal{L}} = \mathbf{1}, \alpha_H, \alpha_L) - \mathbb{P}_h^{\inf}(\mathbf{x}_{\mathcal{H}} = \mathbf{1}, \mathbf{x}_{\mathcal{L}} = \mathbf{1}, \alpha_H, \alpha_L)| < \epsilon$. In what follows, we let $\epsilon < \frac{v_H - v_L}{2}$. Therefore, the constraint $\mathbb{P}_l^{\inf}(\mathbf{x}_{\mathcal{H}} = \mathbf{1}, \mathbf{x}_{\mathcal{L}} = \mathbf{1}, \alpha_H, \alpha_L) \leq v_L + \frac{1}{n}$ will be violated first, resulting in (α_L, α_H) that belongs to the set $\mathcal{A}_2$.

We now proceed with the proof of Part (b). Consider any symmetric mixed strategy for agents $i \in \mathcal{H}$. Using Lemma 2, this mixed strategy must have only two atoms {0, 1}. We let γ_H (and similarly γ_L) denote the probability of being 1 for high-value agents (and similarly for low-value agents). With the abuse of notation whenever we write $x_i = \gamma_H$ this means expectation over x_i, which is one with probability γ_H and zero with probability $1 - \gamma_H$.

We define $\gamma_L : [0,1]^2 \to [0,1]$ such that for any pair (α_L, α_H), we have $\mathbb{P}_l^{\inf}(\mathbf{x}_{\mathcal{H}} = \mathbf{1}, x_l = 1, \mathbf{x}_{\mathcal{L}\setminus\{l\}} = \gamma_L(\alpha_L, \alpha_H), \alpha_L, \alpha_H) = \frac{1}{n} + v_L$. For any $(\alpha_L, \alpha_H) \in \mathcal{A}_2$, there exists $\gamma_L(\alpha_L, \alpha_H)$ in $[0,1]$ that satisfies the previous equality. This is because by using Part (a) of Lemma 1, the function $f : [0,1] \to [0,1]$ where $f(y) = \mathbb{P}_l^{\inf}(\mathbf{x}_{\mathcal{H}} = \mathbf{1}, x_l = 1, \mathbf{x}_{\mathcal{L}\setminus\{l\}} = y\mathbf{1}, \alpha_L, \alpha_H)$ is increasing in y. For $y = 0$, we have $f(0) = \mathbb{P}_l^{\inf}(\mathbf{x}_{\mathcal{H}} = \mathbf{1}, x_l = 1, \mathbf{x}_{\mathcal{L}\setminus\{l\}} = \mathbf{0}, \alpha_L, \alpha_H) \leq \frac{1}{n} + v_L$ where the inequality follows from $(\alpha_L, \alpha_H) \in \mathcal{A}_2$. For $y = 1$, we have $f(1) = \mathbb{P}_l^{\inf}(\mathbf{x}_{\mathcal{H}} = \mathbf{1}, x_l = 1, \mathbf{x}_{\mathcal{L}\setminus\{l\}} = \mathbf{1}, \alpha_L, \alpha_H) \geq \frac{1}{n} + v_L$, where the inequality follows from $(\alpha_L, \alpha_H) \in \mathcal{A}_2$. Mean-value theorem shows that $\gamma_L(\alpha_L, \alpha_H) \in [0,1]$ exists.

We first prove that $x_i = 1$ for all $i \in \mathcal{H}$ and $x_j = \gamma_L(\alpha_L, \alpha_H)$ for all $j \in \mathcal{L}$ is an equilibrium. For $i \in \mathcal{H}$, we have

$$\begin{aligned}
&u_i(x_i = 1, \mathbf{x}_{\mathcal{H}\setminus\{i\}} = \mathbf{1}, \mathbf{x}_{\mathcal{L}} = \mathbf{1}\gamma_L(\alpha_L, \alpha_H), \alpha_H, \alpha_L)\\
&= v_H - \mathbb{P}_i^{\inf}(x_i = 1, \mathbf{x}_{\mathcal{H}\setminus\{i\}} = \mathbf{1}, \mathbf{x}_{\mathcal{L}} = \mathbf{1}\gamma_L(\alpha_L, \alpha_H), \alpha_H, \alpha_L) - c\alpha_H\\
&\overset{(a)}{\geq} v_H - \mathbb{P}_l^{\inf}(x_l = 1, \mathbf{x}_{\mathcal{H}} = \mathbf{1}, \mathbf{x}_{\mathcal{L}\setminus\{l\}} = \mathbf{1}\gamma_L(\alpha_L, \alpha_H), \alpha_H, \alpha_L) - \epsilon - c\alpha_H\\
&\overset{(b)}{=} v_H - \left(v_L + \frac{1}{n}\right) - \epsilon - c\alpha_H \overset{(c)}{\geq} -\frac{1}{n} - c\alpha_H = u_i(x_i = 0, \mathbf{x}_{\mathcal{H}\setminus\{i\}}\\
&= \mathbf{1}, \mathbf{x}_{\mathcal{L}} = \mathbf{1}\gamma_L(\alpha_L, \alpha_H), \alpha_H, \alpha_L),
\end{aligned}$$

where (a) follows from $n \geq M$, (b) follows from the definition of $\gamma_L(\alpha_L, \alpha_H)$, and (c) follows from $\epsilon < \frac{v_H - v_L}{2}$. This shows that high-value agents do not have a profitable deviation. For $j \in \mathcal{L}$, we have $u_j(x_j = 1, \mathbf{x}_{\mathcal{L}\setminus\{j\}} = \mathbf{1}\gamma_L(\alpha_L, \alpha_H), \mathbf{x}_{\mathcal{H}} = \mathbf{1}, \alpha_H, \alpha_L) = v_L - \mathbb{P}^{\inf}(x_j = 1, \mathbf{x}_{\mathcal{L}\setminus\{j\}} = \mathbf{1}\gamma_L(\alpha_L, \alpha_H), \mathbf{x}_{\mathcal{H}} = \mathbf{1}, \alpha_H, \alpha_L) - c\alpha_L = \frac{-1}{n} - c\alpha_L = u_j(x_j = 0, \mathbf{x}_{\mathcal{L}\setminus\{j\}} = \mathbf{1}\gamma_L(\alpha_L, \alpha_H), \mathbf{x}_{\mathcal{H}} = \mathbf{1}, \alpha_H, \alpha_L)$, where the last equality follows from the definition of $\gamma_L(\alpha_L, \alpha_H)$. This proves that each low-value agent is indifferent between playing zero and one given the action profile of others in this equilibrium.

We next prove that there exists no other symmetric equilibrium by listing all possibilities:

1. $(x_l = 0, x_h = 0)$: This is not an equilibrium because any agent i can deviate and receive $1 - \frac{1}{n} - c\alpha_i$ instead of $\frac{-1}{n} - c\alpha_i$.

2. $(x_l = 0, x_h = 1)$: This is not an equilibrium because any agent $j \in \mathcal{L}$ has a profitable deviation. This is because $u_j(x_j = 1, \mathbf{x}_{\mathcal{L}\setminus\{j\}} = \mathbf{0}, \mathbf{x}_{\mathcal{H}} = \mathbf{1}, \alpha_L, \alpha_H) = v_L - \mathbb{P}_j^{\inf}(x_j = 1, \mathbf{x}_{\mathcal{L}\setminus\{j\}} = \mathbf{0}, \mathbf{x}_{\mathcal{H}} = \mathbf{1}, \alpha_L, \alpha_H) - c\alpha_L \geq \frac{-1}{n} - c\alpha_L = u_j(x_j = 0, \mathbf{x}_{\mathcal{L}\setminus\{j\}} = \mathbf{0}, \mathbf{x}_{\mathcal{H}} = \mathbf{1}, \alpha_L, \alpha_H)$, where the inequality follows from $(\alpha_L, \alpha_H) \in \mathcal{A}_2$.

3. $(x_l = 0, x_h = \gamma_H)$: This is not an equilibrium for a similar reason to case 2.

4. $(x_l = 1, x_h = 0)$: This is not an equilibrium because any agent $i \in \mathcal{H}$ has a profitable deviation for a similar reason because

$$\begin{aligned}
&u_i(x_i = 1, \mathbf{x}_{\mathcal{H}\setminus\{i\}} = \mathbf{0}, \mathbf{x}_{\mathcal{L}} = \mathbf{1}, \alpha_L, \alpha_H) = v_H - \mathbb{P}_h^{\inf}\\
&(x_h = 1, \mathbf{x}_{\mathcal{H}\setminus\{h\}} = \mathbf{0}, \mathbf{x}_{\mathcal{L}} = \mathbf{1}, \alpha_L, \alpha_H) - c\alpha_H\\
&\overset{(a)}{\geq} v_H - \mathbb{P}_l^{\inf}(x_l = 1, \mathbf{x}_{\mathcal{H}} = \mathbf{0}, \mathbf{x}_{\mathcal{L}\setminus\{l\}} = \mathbf{1}, \alpha_L, \alpha_H) - \epsilon - c\alpha_H \overset{(b)}{\geq}\\
&v_H - \epsilon - v_L - \frac{1}{n} - c\alpha_H\\
&\overset{(c)}{\geq} \frac{-1}{n} - c\alpha_H = u_i(x_i = 0, \mathbf{x}_{\mathcal{H}\setminus\{i\}} = \mathbf{0}, \mathbf{x}_{\mathcal{L}} = \mathbf{1}, \alpha_L, \alpha_H),
\end{aligned}$$

where (a) follows from n being large, (b) follows from $x_l = 1$ being equilibrium, and (c) follows from $v_H > v_L$ and $\epsilon < v_H - v_L$.

5. $(x_l = 1, x_h = 1)$: This is not an equilibrium because, given $(\alpha_H, \alpha_L) \in \mathcal{A}_2$, $u_j(x_j = 0, \mathbf{x}_{\mathcal{L}\setminus\{j\}} = \mathbf{1}, \mathbf{x}_{\mathcal{H}} = \mathbf{1}, \alpha_L, \alpha_H) = \frac{-1}{n} - c\alpha_L \geq v_L - \mathbb{P}_l^{\inf}(x_j = 1, \mathbf{x}_{\mathcal{L}\setminus\{j\}} = \mathbf{1}, \mathbf{x}_{\mathcal{H}} = \mathbf{1}, \alpha_L, \alpha_H) - c\alpha_L = u_j(x_j = 1, \mathbf{x}_{\mathcal{L}\setminus\{j\}} = \mathbf{1}, \mathbf{x}_{\mathcal{H}} = \mathbf{1}, \alpha_L, \alpha_H)$, implying that any $j \in \mathcal{L}$ has a profitable deviation.

6. $(x_l = 1, x_h = \gamma_H)$: This is not an equilibrium for a similar reason as case 5.

Proof of Part (c). By further decreasing the pair (α_L, α_H), using Proposition 1, the infection probabilities increase and therefore one of the inequalities $\mathbb{P}_l^{\inf}(\mathbf{x}_{\mathcal{H}} = \mathbf{1}, x_l = 1, \mathbf{x}_{\mathcal{L}\setminus\{l\}} = \mathbf{0}, \alpha_L, \alpha_H) \leq v_L + \frac{1}{n}, \mathbb{P}_h^{\inf}(x_h = 1, \mathbf{x}_{\mathcal{H}\setminus\{h\}} = \mathbf{1}, \mathbf{x}_{\mathcal{L}} = \mathbf{0}, \alpha_L, \alpha_H) \leq v_H + \frac{1}{n}$ will be violated first. For $n \geq M$, the first constraint that will be violated is $\mathbb{P}_l^{\inf}(\mathbf{x}_{\mathcal{H}} = \mathbf{1}, x_l = 1, \mathbf{x}_{\mathcal{L}\setminus\{l\}} = \mathbf{0}, \alpha_L, \alpha_H) \leq v_L + \frac{1}{n}$, resulting in the set $\mathcal{A}_3$. In region $\mathcal{A}_3$, we next list all the candidate symmetric equilibria and conclude that $x_i = 1$ for $i \in \mathcal{H}$ and $x_j = 0$ for all $j \in \mathcal{L}$ is the only symmetric equilibrium.

1. $(x_l = \gamma_L, x_h = 1)$: This is not equilibrium because if this was equilibrium we would have been in region $\mathcal{A}_2$. More precisely, because $\mathbb{P}_l^{\inf}(\mathbf{x}_{\mathcal{H}} = \mathbf{1}, x_l = 1, \mathbf{x}_{\mathcal{L}\setminus\{l\}} = \mathbf{0}, \alpha_L, \alpha_H) \geq v_L + \frac{1}{n}$.

2. $(x_l = \gamma_L, x_h = \gamma_H)$: This cannot be equilibrium for large enough n. This is because if it is equilibrium then we must have $\mathbb{P}_l^{\inf}(x_l = 1, \mathbf{x}_{\mathcal{L}\setminus\{l\}} = \mathbf{1}\gamma_L, \mathbf{x}_{\mathcal{H}} = \mathbf{1}\gamma_H, \alpha_L, \alpha_H) = v_L + \frac{1}{n}$ and $\mathbb{P}_h^{\inf}(x_h = 1, \mathbf{x}_{\mathcal{H}\setminus\{h\}} = \mathbf{1}\gamma_H, \mathbf{x}_{\mathcal{L}} = \mathbf{1}\gamma_L, \alpha_L, \alpha_H) = v_H + \frac{1}{n}$. For large enough n, the difference between the left-hand side of the previous equations becomes smaller than $v_H - v_L$, which is a contradiction.

3. $(x_l = \gamma_L, x_h = 0)$: This cannot be equilibrium for large enough n. This is because if it is equilibrium then we must have $\mathbb{P}_l^{\inf}(x_l = 1, \mathbf{x}_{\mathcal{L}\setminus\{l\}} = \mathbf{1}\gamma_L, \mathbf{x}_{\mathcal{H}} = \mathbf{0}, \alpha_L, \alpha_H) = v_L + \frac{1}{n}$ and $\mathbb{P}_h^{\inf}(x_h = 1, \mathbf{x}_{\mathcal{H}\setminus\{h\}} = \mathbf{0}, \mathbf{x}_{\mathcal{L}} = \mathbf{1}\gamma_L, \alpha_L, \alpha_H) \geq v_H + \frac{1}{n}$. For large enough n, the difference between the left-hand side of the previous equations becomes smaller than $v_H - v_L$, which again is a contradiction.

4. $(x_l = 0, x_h = 1)$: This is a symmetric equilibrium because a low-value agent has no profitable deviation. This follows from $u_l(x_l = 0, \mathbf{x}_{\mathcal{L}\setminus\{l\}}, \mathbf{x}_{\mathcal{H}}, \alpha_L, \alpha_H) = -\frac{1}{n} - c\alpha_L \geq v_L - \mathbb{P}_l^{\inf}(x_l = 1, \mathbf{x}_{\mathcal{L}\setminus\{l\}}\mathbf{x}_{\mathcal{H}}, \alpha_L, \alpha_H) - c\alpha_L = u_l(x_l = 1, \mathbf{x}_{\mathcal{L}\setminus\{l\}}, \mathbf{x}_{\mathcal{H}}, \alpha_L, \alpha_H)$, where the inequality follows from $(\alpha_L, \alpha_H) \in \mathcal{A}_3$. Also, a high-value agent has no profitable deviation. This is because $u_h(x_h = 1, \mathbf{x}_{\mathcal{H}\setminus\{h\}} = \mathbf{1}, \mathbf{x}_{\mathcal{L}} = \mathbf{0}, \alpha_L, \alpha_H) = v_H - \mathbb{P}_h^{\inf}(x_h = 1, \mathbf{x}_{\mathcal{H}\setminus\{h\}} = \mathbf{1}, \mathbf{x}_{\mathcal{L}} = \mathbf{0}, \alpha_L, \alpha_H) - c\alpha_H \geq \frac{-1}{n} - c\alpha_H = u_h(x_h = 1, \mathbf{x}_{\mathcal{H}\setminus\{h\}} = \mathbf{1}, \mathbf{x}_{\mathcal{L}} = \mathbf{0}, \alpha_L, \alpha_H)$, where the inequality follows from $(\alpha_L, \alpha_H) \in \mathcal{A}_3$.

5. $(x_l = 0, x_h = \gamma_H)$: This cannot be equilibrium for large enough n. This is because if it is equilibrium then we must have $\mathbb{P}_l^{\inf}(x_l = 1, \mathbf{x}_{\mathcal{L}\setminus\{l\}} = \mathbf{0}, \mathbf{x}_{\mathcal{H}} = \mathbf{1}\gamma_H, \alpha_L, \alpha_H) \geq v_L + \frac{1}{n}$ and $\mathbb{P}_h^{\inf}(x_h = 1, \mathbf{x}_{\mathcal{H}\setminus\{h\}} = \mathbf{1}\gamma_H, \mathbf{x}_{\mathcal{L}} = \mathbf{0}, \alpha_L, \alpha_H) = v_H + \frac{1}{n}$. For large enough n, the difference between the left-hand side of the previous equations becomes smaller than $v_H - v_L$, which again is a contradiction.

6. $(x_l=0,x_h=0)$: This cannot be equilibrium because any agent i can deviate and increase its utility from $\frac{-1}{n}-c\alpha_i$ to $v_i-\frac{1}{n}-c\alpha_i$.

7. $(x_l=1,x_h=1)$: This cannot be equilibrium because if it was, we would be in region $\mathcal{A}_1$.

8. $(x_l=1,x_h=\gamma_H)$: This is not equilibrium because if it was, then $\mathbb{P}_h^{\inf}(x_h=1,\mathbf{x}_{\mathcal{H}\setminus\{h\}}=\mathbf{1}\gamma_H,\mathbf{x}_{\mathcal{L}}=\mathbf{1},\alpha_L,\alpha_H)=\frac{1}{n}+v_H$ and $\mathbb{P}_l^{\inf}(\mathbf{x}_{\mathcal{H}}=\mathbf{1}\gamma_H,\mathbf{x}_{\mathcal{L}}=\mathbf{1},\alpha_L,\alpha_H)\le\frac{1}{n}+v_L$. For large n, the difference between the left-hand side of the previous inequalities is below ϵ which is a contradiction.

9. $(x_l=1,x_h=0)$: This is equilibrium because if it was, then $\mathbb{P}_h^{\inf}(x_h=1,\mathbf{x}_{\mathcal{H}\setminus\{h\}}=\mathbf{1}\gamma_H,\mathbf{x}_{\mathcal{L}}=\mathbf{1},\alpha_L,\alpha_H)\ge\frac{1}{n}+v_H$ and $\mathbb{P}_l^{\inf}(\mathbf{x}_{\mathcal{H}}=\mathbf{1}\gamma_H,\mathbf{x}_{\mathcal{L}}=\mathbf{1},\alpha_L,\alpha_H)\le\frac{1}{n}+v_L$. For large n, the difference between the left-hand side of the previous inequalities is below ϵ which is a contradiction.

Proof of Part (d). By further decreasing the pair (α_L,α_H), the infection probabilities increase and therefore the constraint $\mathbb{P}_h^{\inf}(x_h=1,\mathbf{x}_{\mathcal{H}\setminus\{h\}}=\mathbf{1},\mathbf{x}_{\mathcal{L}}=\mathbf{0},\alpha_L,\alpha_H)\le v_H+\frac{1}{n}$ will be violated and we get to region $\mathcal{A}_4$. We define $\gamma_H:[0,1]^2\to[0,1]$ such that for any pair (α_L,α_H), we have $\mathbb{P}_h^{\inf}(x_h=1,\mathbf{x}_{\mathcal{H}\setminus\{h\}}=\gamma_H(\alpha_L,\alpha_H)\mathbf{1},x_l=0,\mathbf{x}_{\mathcal{L}\setminus\{l\}}=\mathbf{0},\alpha_L,\alpha_H)=\frac{1}{n}+v_H$. For any $(\alpha_L,\ \alpha_H)\in\mathcal{A}_4$, there exists $\gamma_H(\alpha_L,\alpha_H)$ in $[0,1]$ that satisfies the previous equality. This is because by using Part (a) of Lemma 1, the function $f:[0,1]\to[0,1]$, where $f(y)=\mathbb{P}_h^{\inf}(x_h=1,\mathbf{x}_{\mathcal{H}\setminus\{h\}}=y\mathbf{1},\mathbf{x}_{\mathcal{L}}=\mathbf{0},\alpha_L,\alpha_H)$ is increasing in y. For $y=0$, we have $f(0)=\mathbb{P}_h^{\inf}(x_h=1,\mathbf{x}_{\mathcal{H}\setminus\{h\}}=\mathbf{0},\mathbf{x}_{\mathcal{L}}=\mathbf{0},\alpha_L,\alpha_H)=\frac{1}{n}\le\frac{1}{n}+v_H$. For $y=1$, we have $f(1)=\mathbb{P}_h^{\inf}(x_h=1,\mathbf{x}_{\mathcal{H}\setminus\{h\}}=\mathbf{1},\mathbf{x}_{\mathcal{L}}=\mathbf{0},\alpha_L,\alpha_H)\ge\frac{1}{n}+v_H$, where the inequality follows from $(\alpha_L,\alpha_H)\in\mathcal{A}_4$. Using mean-value theorem shows that $\gamma_H(\alpha_L,\alpha_H)\in[0,1]$ exist.

We first establish that $x_j=0$ for all $j\in\mathcal{L}$ and $x_i=\gamma_H(\alpha_L,\alpha_H)$ for all $i\in\mathcal{H}$ is an equilibrium. For $j\in\mathcal{L}$, we have

$$\begin{aligned}
&u_j(x_j=1,\mathbf{x}_{\mathcal{H}}=\mathbf{1}\gamma_H(\alpha_L,\alpha_H),\mathbf{x}_{\mathcal{L}\setminus\{j\}}=\mathbf{0},\alpha_H,\alpha_L)\\
&=v_L-\mathbb{P}_j^{\inf}(\mathbf{x}_{\mathcal{H}}=\mathbf{1}\gamma_H(\alpha_L,\alpha_H),x_j=1,\mathbf{x}_{\mathcal{L}\setminus\{j\}}=\mathbf{0},\alpha_H,\alpha_L)-c\alpha_L\\
&\overset{(a)}{\le}v_L-\mathbb{P}_h^{\inf}(x_h=1,\mathbf{x}_{\mathcal{H}\setminus\{h\}}=\mathbf{1}\gamma_H(\alpha_L,\alpha_H),\mathbf{x}_{\mathcal{L}}=\mathbf{0},\alpha_H,\alpha_L)+\epsilon\\
&\quad-c\alpha_L\overset{(b)}{=}v_L-\frac{1}{n}-v_H+\epsilon-c\alpha_L\\
&\overset{(c)}{\le}\frac{-1}{n}-c\alpha_L=u_j(x_j=0,\mathbf{x}_{\mathcal{H}}=\mathbf{1}\gamma_H(\alpha_L,\alpha_H),\mathbf{x}_{\mathcal{L}}=\mathbf{0},\alpha_H,\alpha_L),
\end{aligned}$$

where (a) follows from $n\ge M$, (b) follows from the definition of $\gamma_H(\alpha_L,\alpha_H)$, and (c) follows from $\epsilon<\frac{v_H-v_L}{2}$. This shows that low-value agents do not have a profitable deviation. For $i\in\mathcal{H}$, we have $u_i(x_i=1,\mathbf{x}_{\mathcal{H}\setminus\{i\}}=\mathbf{1}\gamma_H(\alpha_L,\alpha_H),\mathbf{x}_{\mathcal{L}}=\mathbf{0},\alpha_H,\alpha_L)=v_H-\mathbb{P}_i^{\inf}(x_i=1,\mathbf{x}_{\mathcal{H}\setminus\{i\}}=\mathbf{1}\gamma_H(\alpha_L,\alpha_H),\mathbf{x}_{\mathcal{L}}=\mathbf{0},\alpha_H,\alpha_L)-c\alpha_H=\frac{-1}{n}-c\alpha_H=u_i(x_i=0,\mathbf{x}_{\mathcal{H}\setminus\{i\}}=\mathbf{1}\gamma_H(\alpha_L,\alpha_H),\mathbf{x}_{\mathcal{L}}=\mathbf{0},\alpha_H,\alpha_L)$, showing that high-value agents are indifferent between playing one and zero with this activity profile, where the last equality follows from the definition of $\gamma_H(\alpha_L,\alpha_H)$. Similar to the proof of previous parts, listing all possible symmetric equilibria shows this is the unique symmetric equilibrium. □

Proof of Theorem 1

Proof of Part (a). Using Proposition 2, the unique symmetric equilibrium for $(\alpha_L,\alpha_H)\in\mathcal{A}_1$ is $x_i=1$ for all $i\in\mathcal{V}$ and therefore the infection probability of agent i becomes $\mathbb{P}_i^{\inf}(\mathbf{x}=\mathbf{1},\alpha_L,\alpha_H)$, which is decreasing in (α_L,α_H) as shown in Proposition 1.

Proof of Part (b). Using Proposition 2, the unique symmetric equilibrium for $(\alpha_L,\alpha_H)\in\mathcal{A}_2$ is $x_i=1$ for all $i\in\mathcal{H}$ and mixed action for all $j\in\mathcal{L}$ where $x_j=1$ with probability $\gamma_L(\alpha_L,\alpha_H)$ and $x_j=0$ with probability $1-\gamma_L(\alpha_L,\alpha_H)$. Here, $\gamma_L(\alpha_L,\alpha_H)$ is such that low-value agent are indifferent between playing $x=1$ and $x=0$, which implies

$$\mathbb{P}_l^{\inf}(x_l=1,\mathbf{x}_{\mathcal{L}\setminus\{l\}}=\mathbf{1}\gamma_L(\alpha_L,\alpha_H),\mathbf{x}_{\mathcal{H}}=\mathbf{1},\alpha_L,\alpha_H)=\frac{1}{n}+v_L. \tag{A.1}$$

The infection probability of low value agents is

$$\begin{aligned}
&\gamma_L(\alpha_L,\alpha_H)\mathbb{P}_l^{\inf}(x_l=1,\mathbf{x}_{\mathcal{L}\setminus\{l\}}=\mathbf{1}\gamma_L(\alpha_L,\alpha_H),\mathbf{x}_{\mathcal{H}}=\mathbf{1},\alpha_L,\alpha_H)\\
&+(1-\gamma_L(\alpha_L,\alpha_H))\mathbb{P}_l^{\inf}(x_l=0,\mathbf{x}_{\mathcal{L}\setminus\{l\}}=\mathbf{1}\gamma_L(\alpha_L,\alpha_H),\mathbf{x}_{\mathcal{H}}=\mathbf{1},\alpha_L,\alpha_H)\\
&=\gamma_L(\alpha_L,\alpha_H)\left(\frac{1}{n}+v_L\right)+(1-\gamma_L(\alpha_L,\alpha_H))\frac{1}{n}=\gamma_L(\alpha_L,\alpha_H)v_L-\frac{1}{n}.
\end{aligned}$$

We next prove that the $\gamma_L(\alpha_L,\alpha_H)$, which is the solution to (A.1) is increasing in (α_L,α_H). We prove this by assuming the contrary and reaching a contradiction. In particular, suppose that $(\alpha_L',\alpha_H')>(\alpha_L,\alpha_H)$ and $\gamma_L(\alpha_L',\alpha_H')<\gamma_L(\alpha_L,\alpha_H)$. We can write

$$\begin{aligned}
v_L+\frac{1}{n}&\overset{(a)}{=}\mathbb{P}_l^{\inf}(x_l=1,\mathbf{x}_{\mathcal{L}\setminus\{l\}}=\mathbf{1}\gamma_L(\alpha_L,\alpha_H),\mathbf{x}_{\mathcal{H}}=\mathbf{1},\alpha_L,\alpha_H)\\
&\overset{(b)}{>}\mathbb{P}_l^{\inf}(x_l=1,\mathbf{x}_{\mathcal{L}\setminus\{l\}}=\mathbf{1}\gamma_L(\alpha_L',\alpha_H'),\mathbf{x}_{\mathcal{H}}=\mathbf{1},\alpha_L,\alpha_H)\\
&\overset{(c)}{>}\mathbb{P}_l^{\inf}(x_l=1,\mathbf{x}_{\mathcal{L}\setminus\{l\}}=\mathbf{1}\gamma_L(\alpha_L',\alpha_H'),\mathbf{x}_{\mathcal{H}}=\mathbf{1},\alpha_L',\alpha_H')\overset{(d)}{=}v_L+\frac{1}{n},
\end{aligned}$$

which is a contradiction. In the previous derivation, (a) and (d) follow by invoking (A.1), (b) follows from the assumption that $\gamma_L(\alpha_L',\alpha_H')<\gamma_L(\alpha_L,\alpha_H)$ and part (a) of Lemma 1, and (c) follows from $(\alpha_L',\alpha_H')>(\alpha_L,\alpha_H)$ and Proposition 1. The infection probability of high-value agents is given in (A.1), which remains equal to $\frac{1}{n}+v_L$.

Proof of Part (c). Using Proposition 2, the unique symmetric equilibrium for $(\alpha_L,\alpha_H)\in\mathcal{A}_3$ is $x_j=0$ for all $j\in\mathcal{L}$ and $x_i=1$ for all $i\in\mathcal{H}$, and therefore the infection probability of an agent k becomes $\mathbb{P}_k^{\inf}(\mathbf{x}_{\mathcal{H}}=\mathbf{1},\mathbf{x}_{\mathcal{L}}=\mathbf{0},\alpha_L,\alpha_H)$, which is decreasing in (α_L,α_H) as shown in Proposition 1.

Proof of Part (d). Using Proposition 2, the unique symmetric equilibrium for $(\alpha_L,\alpha_H)\in\mathcal{A}_4$ is $x_j=0$ for all $j\in\mathcal{L}$ and mixed action for all $i\in\mathcal{H}$ where $x_i=1$ with probability $\gamma_H(\alpha_L,\alpha_H)$ and $x_i=0$ with probability $1-\gamma_H(\alpha_L,\alpha_H)$. Here, $\gamma_H(\alpha_L,\alpha_H)$ is such that high-value agents are indifferent between playing $x=1$ and $x=0$, which implies

$$\mathbb{P}_h^{\inf}(x_h=1,\mathbf{x}_{\mathcal{H}\setminus\{h\}}=\mathbf{1}\gamma_H(\alpha_L,\alpha_H),\mathbf{x}_{\mathcal{L}}=\mathbf{0},\alpha_L,\alpha_H)=\frac{1}{n}+v_H. \tag{A.2}$$

The infection probability of high value agents is

$$\begin{aligned}
&\gamma_H(\alpha_L,\alpha_H)\mathbb{P}_l^{\inf}(x_h=1,\mathbf{x}_{\mathcal{H}\setminus\{l\}}=\mathbf{1}\gamma_H(\alpha_L,\alpha_H),\mathbf{x}_{\mathcal{L}}=\mathbf{0},\alpha_L,\alpha_H)\\
&+(1-\gamma_H(\alpha_L,\alpha_H))\mathbb{P}_h^{\inf}(x_h=0,\mathbf{x}_{\mathcal{H}\setminus\{h\}}=\mathbf{1}\gamma_H(\alpha_L,\alpha_H),\mathbf{x}_{\mathcal{L}}=\mathbf{0},\alpha_L,\alpha_H)\\
&=\gamma_H(\alpha_L,\alpha_H)\left(\frac{1}{n}+v_H\right)+(1-\gamma_H(\alpha_L,\alpha_H))\frac{1}{n}=\gamma_H(\alpha_L,\alpha_H)v_H-\frac{1}{n}.
\end{aligned}$$

We next prove that the $\gamma_H(\alpha_L,\alpha_H)$, which is the solution to (A.2) is increasing in (α_L,α_H). We establish this by assuming

the contrary and reaching a contradiction. In particular, suppose that $(\alpha'_L, \alpha'_H) > (\alpha_L, \alpha_H)$ and $\gamma_H(\alpha'_L, \alpha'_H) < \gamma_H(\alpha_L, \alpha_H)$. We can write

$$v_H + \frac{1}{n} \overset{(a)}{=} \mathbb{P}_h^{\inf}(x_h = 1, \mathbf{x}_{\mathcal{H}\setminus\{h\}} = \mathbf{1}\gamma_H(\alpha_L, \alpha_H), \mathbf{x}_{\mathcal{L}} = \mathbf{0}, \alpha_L, \alpha_H)$$
$$\overset{(b)}{>} \mathbb{P}_h^{\inf}(x_h = 1, \mathbf{x}_{\mathcal{H}\setminus\{h\}} = \mathbf{1}\gamma_H(\alpha'_L, \alpha'_H), \mathbf{x}_{\mathcal{L}} = \mathbf{0}, \alpha_L, \alpha_H)$$
$$\overset{(c)}{>} \mathbb{P}_h^{\inf}(x_h = 1, \mathbf{x}_{\mathcal{H}\setminus\{h\}} = \mathbf{1}\gamma_H(\alpha'_L, \alpha'_H), \mathbf{x}_{\mathcal{L}} = \mathbf{0}, \alpha'_L, \alpha'_H) \overset{(d)}{=} v_H + \frac{1}{n},$$

which is a contradiction. In the previous derivation, (a) and (d) follow by invoking (A.2), (b) follows from the assumption that $\gamma_H(\alpha'_L, \alpha'_H) < \gamma_H(\alpha_L, \alpha_H)$ and part (a) of Lemma 1, and (c) follows from Proposition 1. The infection probability of low-value agents remains $\frac{1}{n}$, completing the proof. □

Endnotes

[1] In the online appendix, we show that our model can be viewed as an extended version (to account for testing and network structure) of a discrete version of an SIR setup, namely the Reed-Frost model.

[2] In our model, testing is used for identifying and isolating infected individuals in order to reduce the spread of the infection to others. We also analyze the complementary effects of "social distancing" policies that lead to costly reductions in the quality and level of an individual's social interactions and reduce the likelihood that he or she becomes infected and, when infected, transmits it to others.

[3] Other related work includes Acemoglu et al. (2016), who introduce precautionary behavior in the context of a virus spreading over a network; Morris (2000), Tardos and Wexler (2007), Blume et al. (2011), Acemoglu et al. (2017a, b), Elliott et al. (2014), Capponi (2016), Bernard et al. (2019), and Capponi (2016), who study contagion over financial networks; and Manshadi et al. (2020), who study diffusion in random networks. In particular, Bernard et al. (2019) consider a dynamic game between a benevolent social planner and banks in which the social planner chooses an intervention policy (bail-in and bail-out), and banks decide whether to contribute financially to the rescue. This is reminiscent of our model in which a benevolent social planner chooses the testing probability of agents, and the agents decide on the amount of social activity. A crucial distinction is that in Bernard et al. (2019), the network of contractual relations is exogenously specified.

[4] Our percolation process is closely connected to the SIR model. In particular, as we discuss in the online companion, the relevant comparison is to the discrete-time version of the SIR model, often referred to as the Reed-Frost model. On a complete network, our percolation process generates the same behavior of cumulative infections as in the Reed-Frost model, but is particularly tractable when the social network is not complete, which is the main case of interest for us.

[5] In this formulation, the lower utility from reduced social activity due to isolation is assumed to be the same for all agents and incorporated into the second term. The results are identical if this cost is allowed to be type dependent.

[6] Because n is sufficiently large, when all agents are playing one, a change in one agent's testing probability does not alter the infection probability of other agents by more than $v_H - v_L$.

[7] We say a sequence of sets $\{A\}_{n=1}^{\infty}$ converges to set A if for any $\epsilon > 0$, there exists $M \in \mathbb{N}_+$ such that for $n \geq M$, we have $A_n \subseteq A^{(\epsilon)}$ and $A \subseteq A_n^{(\epsilon)}$, where for any set B, $B^{(\epsilon)}$ denotes $\bigcup_{b\in B}\{x : \|x - b\|_2 \leq \epsilon\}$.

[8] The thresholds on θ relate to the ones found in Lemma 3, and in particular, we have $\theta^{(1)} = r_H \alpha_H^{(1)}$, $\theta^{(2)} = \alpha_H^{(2)} r_H$, and $\theta^{(3)} = \alpha_H^{(3)} r_H$. These thresholds are such that $\frac{\theta^{(2)}}{r_H} \in [0,1]$ and for $\theta \in [\theta^{(3)}, \theta^{(2)})$, we have $\frac{\theta}{r_H} \in [0,1]$.

References

Acemoglu D, Malekian A, Ozdaglar A (2016) Network security and contagion. *J. Econom. Theory* 166:536–585.

Acemoglu D, Ozdaglar A, Tahbaz-Salehi A (2017a) Microeconomic origins of macroeconomic tail risks. *Amer. Econom. Rev.* 107(1):54–108.

Acemoglu D, Chernozhukov V, Werning I, Whinston MD (2021) Optimal targeted lockdowns in a multigroup SIR model. *Amer. Econom. Rev.: Insights* 3(4):487–502.

Acemoglu D, Makhdoumi A, Malekian A, Ozdaglar A (2017b) Privacy-constrained network formation. *Games Econom. Behav.* 105:255–275.

Alvarez FE, Argente D, Lippi F (2021) A simple planning problem for COVID-19 lock-down, testing, and tracing. *Amer. Econom. Rev.: Insights* 3(3):367–382.

Atkeson A (2020) What will be the economic impact of covid-19 in the us? Rough estimates of disease scenarios. NBER Working Paper No. 26867, https://www.nber.org/papers/w26867.

Bala V, Goyal S (2000) A noncooperative model of network formation. *Econometrica* 68(5):1181–1229.

Bauch CT, Galvani AP (2013) Social factors in epidemiology. *Science* 342(6154):47–49.

Berger DW, Herkenhoff KF, Mongey S (2020) An seir infectious disease model with testing and conditional quarantine. Preprint, submitted March 26, https://dx.doi.org/10.2139/ssrn.3561142.

Bernard B, Capponi A, Stiglitz JE (2022) Bail-ins and bail-outs: Incentives, connectivity, and systemic stability. *J. Political Econom.* 130(7):1805–1859.

Birge JR, Candogan O, Feng Y (2022) Controlling epidemic spread: Reducing economic losses with targeted closures. *Management Sci.* 68(5):3175–3195.

Blume L, Easley D, Kleinberg J, Kleinberg R, Tardos E (2013) Network formation in the presence of contagious risk. *ACM Trans. Econom. Comput.* 1(2):1–20.

Bollobás B, Béla B (2001) *Random Graphs* (Cambridge University Press, Cambridge, UK).

Brotherhood L, Kircher P, Santos C, Tertilt M (2020) An economic model of the covid-19 epidemic: The importance of testing and age-specific policies. Preprint, submitted June 5, https://dx.doi.org/10.2139/ssrn.3618840.

Brumfiel G (2020) To end the coronavirus crisis we need widespread testing, experts say. Accessed March 24, 2020, https://www.npr.org/sections/health-shots/2020/03/24/820157519/to-end-the-coronavirus-crisis-we-need-widespread-testing-experts-say.

BruSoe-Lin S, Hecht R (2020) To suppress Covid-19, we need to test those with no symptoms. Accessed April 26, 2020, https://www.nytimes.com/2020/04/26/opinion/coronavirus-test-asymptomatic.html.

Cabrales A, Calvó-Armengol A, Zenou Y (2011) Social interactions and spillovers. *Games Econom. Behav.* 72(2):339–360.

Capponi A (2016) Systemic risk, policies, and data needs. *Optimization Challenges in Complex, Networked and Risky Systems*, 185–206.

Crandall RW, Graham JD (1984) Automobile safety regulation and offsetting behavior: Some new empirical estimates. *Amer. Econom. Rev.* 74(2):328–331.

Currarini S, Jackson MO, Pin P (2009) An economic model of friendship: Homophily, minorities, and segregation. *Econometrica* 77(4):1003–1045.

Dasaratha K (2020) Virus dynamics with behavioral responses. Preprint, submitted. Accessed April 30, 2020, https://arxiv.org/abs/2004.14533.

Deb R, Pai M, Vohra A, Vohra R (2020) Testing alone is insufficient. Preprint, submitted. Accessed April 30, 2020, https://dx.doi.org/10.2139/ssrn.3593974.

Draief M, Massouli L (2010) *Epidemics and Rumours in Complex Networks* (Cambridge University Press, Cambridge, UK).

Drakopoulos K, Randhawa RS (2021) Demand control of information products: Why perfect tests may not be worth waiting for. *Management Sci.* 67(11):6678–6693.

Drakopoulos K, Ozdaglar A, Tsitsiklis JN (2017) When is a network epidemic hard to eliminate? *Math. Oper. Res.* 42(1):1–14.

Eichenbaum MS, Rebelo S, Trabandt M (2021) The macroeconomics of epidemics. *Rev. Financial Stud.* 34(11):5149–5187.

Elliott M, Golub B, Jackson MO (2014) Financial networks and contagion. *Amer. Econom. Rev.* 104(10):3115–3153.

Ely J, Galeotti A, Jann O, Steiner J (2021) Optimal test allocation. *J. Econom. Theory* 193:105236.

Eshragh A, Alizamir S, Howley P, Stojanovski E (2020) Modeling the dynamics of the covid-19 population in Australia: A probabilistic analysis. *PLoS One* 15(10):e0240153.

Farboodi M, Jarosch G, Shimer R (2021) Internal and external effects of social distancing in a pandemic. *J. Econom. Theory* 196:105293.

Gans J (2020) *Economics in the Age of COVID-19* (MIT Press, Cambridge, MA).

Gans JS (2022) The economic consequences of r= 1: Toward a workable behavioural epidemiological model of pandemics. *Rev. Econom. Anal.* 14(1):3–25.

Glover A, Heathcote J, Krueger D, Ríos-Rull J-V (2020) Health vs. wealth: On the distributional effects of controlling a pandemic. Preprint, submitted August 18, https://dx.doi.org/10.2139/ssrn.3583489.

Goyal S (1993) *Sustainable Communication Networks* (Econometric Institute, Erasmus University, Rotterdam).

Jackson M (2008) *Social and Economic Networks* (Princeton University Press, Princeton, NJ).

Jackson MO, Wolinsky A (1996) A strategic model of social and economic networks. *J. Econom. Theory* 71(1):44–74.

Jones CJ, Philippon T, Venkateswaran V (2021) Optimal mitigation policies in a pandemic: Social distancing and working from home. *Rev. Financial Stud.* 34(11):5188–5223.

Kaplan EH (2020) Containing 2019-ncov (Wuhan) coronavirus. *Health Care Management Sci.* 23(3):311–314.

Kasy M, Teytelboym A (2020) Adaptive targeted infectious disease testing. *Oxford Rev. Econom. Policy* 36:S77–S93.

Keeler TE (1994) Highway safety, economic behavior, and driving environment. *Amer. Econom. Rev.* 84(3):684–693.

Kempe D, Kleinberg J, Tardos É (2015) Maximizing the spread of influence through a social network. *Theory Comput.* 11(4):105–147.

Kruse T, Strack P (2020) Optimal control of an epidemic through social distancing. Preprint, submitted June 4, https://dx.doi.org/10.2139/ssrn.3583186.

Lakdawalla D, Sood N, Goldman D (2006) HIV breakthroughs and risky sexual behavior. *Quart. J. Econom.* 121(3):1063–1102.

Leung KY, Ball F, Sirl D, Britton T (2018) Individual preventive social distancing during an epidemic may have negative population-level outcomes. *J. Royal Soc. Interface* 15(145):20180296.

Lindgren B, Stuart C (1980) The effects of traffic safety regulation in sweden. *J. Political Econom.* 88(2):412–427.

Maloney W, Taskin T (2020) Determinants of social distancing and economic activity during covid-19: A global view. Preprint, submitted May 14, https://dx.doi.org/10.2139/ssrn.3599572.

Manshadi V, Misra S, Rodilitz S (2020) Diffusion in random networks: Impact of degree distribution. *Oper. Res.* 68(6):1722–1741.

McAdams D (2020) Economic epidemiology in the wake of Covid-19. *Economics* 82120:122900.

Morris S (2000) Contagion. *Rev. Econom. Stud.* 67(1):57–78.

Newman MEJ, Strogatz SH, Watts DJ (2001) Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E* 64(2):026118.

Peltzman S (1975) The effects of automobile safety regulation. *J. Political Econom.* 83(4):677–725.

Philipson T (2000) Economic epidemiology and infectious diseases. *Handbook Health Econom.* 1:1761–1799.

Philipson TJ, Posner RA (1993) *Private Choices and Public Health: The AIDS Epidemic in an Economic Perspective* (Harvard University Press, Cambridge, MA).

Piguillem F, Shi L (2022) Optimal covid-19 quarantine and testing policies. *Econom. J.* 132(647):2534–2562.

Searchinger T, LaMantia A, Douglas G (2020) Mass testing is the only sustainable solution to the Coronavirus. https://www.washingtonpost.com/opinions/2020/03/23/mass-testing-is-only-sustainable-solution/.

Sun P, Yang L, De Véricourt F (2009) Selfish drug allocation for containing an international influenza pandemic at the onset. *Oper. Res.* 57(6):1320–1332.

Tardos E, Wexler T (2007) Network formation games and the potential function method. Nisan N, Roughgarden T, Tardos E, Vazirani V, eds. *Algorithmic Game Theory*, chapter 19 (Cambridge University Press, Cambridge, UK), 487–516.

Toxvaerd F (2020) *Equilibrium Social Distancing* (Cambridge University Press, Cambridge, UK).

Vega-Redondo F (2007) *Complex Social Networks* (Cambridge University Press, Cambridge, UK).

Wang S, de Véricourt F, Sun P (2009) Decentralized resource allocation to control an epidemic: A game theoretic approach. *Math. Biosci.* 222(1):1–12.

Daron Acemoglu is an institute professor at Massachusetts Institute of Technology. He is the author of several books. His academic work covers a wide range of areas, including political economy, economic development, economic growth, technological change, inequality, labor economics, and economics of networks. He has received many awards including the John Bates Clark Medal in 2005, the Erwin Plein Nemmers Prize in 2012, and the 2016 BBVA Frontiers of Knowledge Award.

Ali Makhdoumi is an associate professor in the decision sciences area at Fuqua School of Business, Duke University. His research interests include optimization, game theory, network economics, and learning theory with applications to social and technological systems.

Azarakhsh Malekian is an associate professor of operations management and statistics at Rotman School of Management, University of Toronto. Her research interests include network economics, algorithmic game theory, market design, data markets, and privacy.

Asuman Ozdaglar is the MathWorks Professor of Electrical Engineering and Computer Science (EECS) at the Massachusetts Instituteof Technology. She is the department head of EECS and the deputy dean of academics in the Schwarzman College of Computing. Her research expertise includes optimization theory, game theory, distributed optimization and control, and network analysis. She is the recipient of several awards including National Science Foundation Career award and Donald P. Eckman award of the American Automatic Control Council.