Plague and Prejudice: Disease Beliefs and Social Exclusion

Javier A. Birchenall
University of California at Santa Barbara

October 31, 2020

Abstract

I examine the social role of disease in an exchange economy where individual health is uncertain and imperfectly observed. Market exchange and the social control of disease involve a signal-extraction problem that leads to a “better safe than sorry” precautionary approach. In the model, disease and fear of contagion trigger selective social exclusion and frequent false alarms (i.e., over-diagnoses). The intrinsic characteristics of disease, organized here along danger and visibility lines, determine the social tolerance to disease. I interpret the social regulation of the most significant modern epidemics in the West (leprosy, plague, smallpox, cholera) along danger and visibility lines.

Keywords: Disease, uncertainty, social interactions and contagion, Western history
JEL classification: D80, I15, N10.
Communications to: Javier A. Birchenall
Department of Economics, 2127 North Hall
University of California, Santa Barbara CA 93106
Phone/fax: (805) 893-5275
Javier.Birchenall@ucsb.edu
1 Introduction

Disease is at once a biological event and a social actor that, throughout history, has served to enact boundaries of inclusion and exclusion drawn to minimize the risk of exposure to dangerous and contagious pathogens. In this paper, I study a model economy in which disease threats foster social exclusion and use the theory to offer an interpretive history for the social regulation of the most significant epidemic diseases in the West since early modern times. The social regulation of disease deals with the social construction of “diseased” labels and with disease-avoidance strategies based on individual beliefs that aim to limit sociality and consequently the extent of the market.

The model economy has two essential features. First, markets serve as social institutions with economic and non-economic functions. I consider decentralized trading as a social institution that promotes anonymous commodity exchange, but also facilitates disease transmission and the formation of beliefs about the extent of disease in the population. Second, all decision-making takes place in the presence of uncertainty about the prevalence of disease in the population. In the model, individuals are healthy or diseased, but types are unobservable so individuals rely on disease signs to infer the gains from trade and the risks associated with social interactions. Uncertainty on the health of a potential trading partner implies that contagion and fear of contagion limit the extent of the market. In equilibrium, unhealthy-looking individuals are socially labelled as “diseased” and selectively shunned from economic transactions. When faced with uncertain health risks, society confronts communicable diseases by enforcing a “better safe than sorry” exclusionary approach.

The model views social exclusion as a preventive strategy against communicable diseases. This strategy makes society liable to miss detections (i.e., labeling a diseased individual as “clean”) and issue false alarms (i.e., labeling a healthy individual as “diseased”). While false alarms may lead to over-diagnoses (i.e., ‘unnecessary’ social labeling and exclusion), a missed detection might lead to widespread disease or death. In theory, facing potential harm even when no “real risks” are present, social control sides with false alarms. The social regulation of disease therefore involves exclusionary actions for non-threatening individuals whose observed features may not be due to disease at all (e.g., limp amputations and skin
conditions). Not all diseases, however, should elicit the same social responses. I categorize the social tolerance to disease along *danger* and *visibility* lines. That is, I derive equilibrium comparative statics for diseases with intrinsic differences along danger (i.e., ex-post costs of social encounters) and visibility (i.e., stochastic properties of disease such as persistence in the disease state and the signal-to-noise ratio).

To study disease-specific responses, it would be ideal to examine social reactions in multiple but homogeneous populations with the same understanding of disease when each experiences one and only one epidemic of a different disease. This ideal is obviously not possible. It is even difficult to systematically quantify many concepts in the model, including the visibility of disease signs and the degree of social exclusion which can range from subtle avoidance to outright expulsion. To highlight the key tensions in the model, I interpret some social responses to the most important epidemic diseases that visited Western Europe and offshoots since early modern times using *danger* and *visibility* lines. A complete analysis with wider spatial and temporal reach for all relevant epidemics is not possible within the constraints of this paper, so my case studies do not consider the pandemic century that began in 1918, which also has a much better understanding of disease. I consider Hansen’s disease (leprosy), plague, smallpox, and cholera. These diseases are sufficiently different to validate the notions of danger and visibility in the model. They were also sequential, starting in the late Middle Ages, and all originated outside the area of study.

Hansen’s disease, for example, featured detailed rules for social labeling and exclusion since ancient times. *M. leprae* (Hansen’s disease causative agent), is a very “poor pathogen” because of its low infectivity and high natural immunity. Leprosy, however, is persistent (i.e., chronic) and highly visible because it causes lesions to the face, the focus of social encounters (Oaten et al. [42]). Smallpox, more infectious and deadly than leprosy, is also highly visible for similar reasons. Smallpox, however, was acute and granted prolonged immunity to survivors. The model associates the detailed exclusionary rules for leprosy to its high signal-to-noise ratio. The model associates the higher social tolerance to smallpox, including its deliberate exposure (i.e., variolation), to its transiency. On the other extreme, plague and cholera represent among the most dangerous epidemics in history. Both are acute diseases that feature high infectivity, high case fatality rates, and very low signal-to-noise
ratios in the early stages of the disease. The model associates the broad social response to these diseases to their danger and invisibility. The model also views some exclusionary policies along class, ethnicity, and race dimensions as part of false alarms emerging from a high sensitivity to disease signs.

**Related literature.** Economists often focus on the demographic and economic impacts of disease.\(^1\) My focus here is on the social response to disease. The model economy suggests that disease-avoidance strategies lead to a ‘disintegrative’ social order that limits the extent of the market. As in economic models of *discrimination* (i.e., Fang and Moro [20]), a dislike for trading with unhealthy individuals is exercised in noisy and uncertain environments. I consider a fixed-price exchange (i.e., indivisible commodities), so diseased-looking individuals are not charged higher prices for the same commodities, but they are all together excluded from the market. Disease, in other words, triggers a form of *statistical segregation*.\(^2\) It is no wonder, drawing insights from the model, that since early modern times quarantines for traders and travelers have applied even for those who do not appear to be sick.

Historians offer insights on the role of disease in areas where economists often pay relatively little attention to. Historians have long recognized the ‘disintegrative’ effect of epidemic diseases on social, economic, and political institutions, but have not offered a systematic analysis for the differential regulation of epidemic diseases in the past. There are countless accounts of fear of contagion and ‘unreasonable’ behaviors motivated by epidemic diseases, even after the improved understanding granted by the germ theory of disease. Important historical studies such as Cohn [11], Duffy [17], McNeill [37], Ranger and Slack [45], and Watts [53] emphasize the novelty of the disease and the scale of the outbreak, but not the intrinsic characteristics of the disease behind it. I rely on their detailed catalogues

---

1. For example, Fogel [21] has shown that epidemics and crises mortality ‘explain’ only ten percent of the secular decline in mortality since the mid-eighteenth century in England and Wales, while Weil [54] has shown that health disparities across countries ‘account’ for only ten percent of the current disparities in income per capita. I have, in a separate paper, reconsidered the significance of disease for economic development through sorting and segregation of healthy and diseased types in environments where health is fully observable; see Birchenall [8].

2. I allow for learning in a stationary equilibrium, but I do not consider human capital investments, as in Lundberg and Startz [33], or self-fulfilling outcomes, as in Coate and Loury [10]. I leave these aspects for future research. I have, in a separate paper, reconsidered the significance of disease for economic development through sorting and segregation of healthy and diseased types in environments where health is fully observable; see Birchenall [8].
and historical narratives to provide a synthesis that highlights differential responses along danger and visibility lines. As noted above, my focus on the most spectacular diseases that visited the West is due to the sequential nature of their epidemics and the fact that their main mode of transmission was trading and travel. My model-based interpretation of history is parsimonious but obviously risks oversimplification.

There is a separate literature that uses ecological and evolutionary principles to relate disease to the acquisition of culture and socially learned values and behaviors, including social exclusion and stigma; see, e.g., Kurzban and Leary [30] and Oaten et al. [42]. A particular strand, the parasite-stress theory of values and sociality (Thornhill and Fincher [51]), argues for a behavioral immune system parallel to a biological immune system that functionally selects values and preferences based on pathogen exposure and shared immunity. High pathogen environments, for example, favor collectivism (i.e., nepotism, xenophobia, and ethnocentrism) while low pathogen environments favor individualism (i.e., liberalism, outgroup tolerance and generosity). Their work has a much broader scope and examines an impressive range of correlations. An evolutionary view is complementary and not contradictory to the present paper. Their verbal theory, for example, has no role for uncertainty and detection principles (i.e., errors), it is not specific about differential mechanisms by which disease leads to social exclusion, and it does not consider a trade-off between the risk of exposure and the potential gains from trade.

2 Theory: Exchange and Exclusion

This section considers a highly stylized (random search) model of exchange subject to disease spillovers and informational frictions. Trade is a social activity, and trading partners are either healthy or diseased. Health types, however, are unobservable so the prevalence of disease in the economy must be inferred from a noisy signal, i.e., a disease sign. Interacting with an ill partner is costly so disease and fear of contagion are regulated through selective social exclusion. Unhealthy-looking traders, in particular, will be precautionarily excluded from economic transactions in a “better safe than sorry” strategy.

Environment. There is a continuum of individuals of two types, healthy and diseased.
Health is a binary random variable with mutually exclusive values $H = \{d, h\}$; $H = d$ denotes a diseased or unhealthy individual and $H = h$ a healthy one. Health evolves as a Markov process over imperfectly observed states (i.e., as a hidden Markov process). Only nature knows an individual’s health type. Potential trading partners observe a signal that conveys imperfect information about a trading partner’s type. Inferences about the prevalence of disease in the population are based on these disease signs. Errors in detecting a partner’s true health type are costly and will motivate social exclusion.

Individuals are buyers or sellers, and exchange is bilateral and anonymous. Meetings take place at random, in proportion to the number of potential trading partners. During a meeting, traders observe the signal of their counterpart, update their beliefs, and decide whether to trade or not. I proceed backwards and describe first the decisions buyers and sellers make once they are in a meeting. I focus on a buyer’s decisions since a seller’s decisions are symmetric. Details about the trading protocol are discussed later.

An individual’s health is disguised and governed by a first-order Markov process. A transition into a diseased state, from a healthy state, is given by $\Pr\{H' = d|H = h\} = \delta$. An individual remains in a diseased state according to $\Pr\{H' = d|H = d\} = 1 - \delta$. For the healthy state, $\Pr\{H' = h|H = d\} = \eta$ and $\Pr\{H' = h|H = h\} = 1 - \eta$. The transition matrix for health types is

$$Q = \begin{bmatrix} 1 - \delta & \delta \\ \eta & 1 - \eta \end{bmatrix}. \tag{1}$$

In the paper, $Q$ is exogenous, but I consider endogenous changes in the distribution of disease types in Appendix A.

Upon meeting a seller, a buyer observes a series of features or disease signs. I consider a continuous signal $x$ supported on an interval $X \equiv (x_-, x_+)$. The signal $x$ is generated either from a (conditional) density $p(x|d)$, for a diseased individual, or from a (conditional) density $p(x|h)$, for a healthy individual. (Sellers independently observe a series of disease signs $y \in X$ on the health of potential buyers.) The densities $p : X \times \{d, h\} \to [0, 1]$ are given and known. They represent the likelihood that observed features are consistent with a particular type.
**Assumption 1.** The likelihood ratio $L(x) \equiv p(x|d)/p(x|h)$ is everywhere continuous and monotone, i.e., $L(x') > L(x)$ for all $x' > x$ in $X$, with $L(x_-) = 0$ and $L(x_+) = \infty$.

A monotone likelihood ratio implies that disease signs are informative about the underlying health types. Continuity helps rule out problems with invertibility, random detection decisions, and mixed strategies. Limiting properties are convenient for the uniqueness of interior solutions.

**Disease detection.** Before agreeing to trade, buyers assess the health of potential sellers. (Sellers also assess the health of potential buyers.) Disease detection is important because transactions with diseased partners are riskier and costlier than transactions with healthy ones in ways specified later on.

The prevalence of disease in the population is unknown. A buyer’s initial prior about the disease prevalence is $\pi$. Priors may result from past inferences or personal judgements. At a meeting, buyers (respectively sellers) observe $x$ (resp. $y$) and update their disease prevalence beliefs based on the observed seller’s disease sign. Belief updates use Bayes rule:

$$P'(d|x) = \frac{P(1 - \delta)p(x|d) + (1 - P)\eta p(x|d)}{p(x)},$$  \hspace{1cm} (2)

where $p(x) = P(1 - \delta)p(x|d) + (1 - P)\eta p(x|d) + P\delta p(x|h) + (1 - P)(1 - \eta)p(x|h)$ is the marginal density of a signal $x$. The numerator in (2) accounts for diseased individuals that remain in the diseased state or transition from a healthy state. (Sellers share the common prior $P$ and update it using Bayes rule.)

**Buyers.** Buyers classify sellers according to observed disease signs. Classifications are non-strategic. The classifier, $\mathcal{H}_b : X \rightarrow \{d, h\}$, is a (statistical) decision rule where $\mathcal{H}_b(x) = d$ denotes a “diseased” seller and $\mathcal{H}_b(x) = h$ denotes a “clean” seller.

Disease detection is binary with four possible outcomes: a correct detection of disease, $\Pr\{\mathcal{H}_b(x) = d|d\}$; a missed detection of disease, $\Pr\{\mathcal{H}_b(x) = h|d\}$; a correct rejection, $\Pr\{\mathcal{H}_b(x) = h|h\}$; and a false alarm, $\Pr\{\mathcal{H}_b(x) = d|h\}$. Each outcome carries a benefit. The ex-post utility of classifying a seller as $\mathcal{H}_b(x) = i$, when the true type is $j$, is $U(i, j)$. Therefore, the expected utility of a “disease” label is $\mathbb{E}_x[U(d, j)] = P'(d|x)U(d, d) + P'(h|x)U(d, h)$, while a “clean” label yields $\mathbb{E}_x[U(h, j)] = P'(h|x)U(h, h) + P'(d|x)U(h, d)$. 7
**Assumption 2.** Ex-post utilities satisfy \(U(h, h) > U(d, h) > U(d, d) > U(h, d)\).

For either of the health types, Assumption 2 implies that correct assessments are better than misclassifications. A healthy partner is also preferable to a diseased one, regardless of the classification. The middle inequality in Assumption 2, for example, says that exchange with a healthy seller is preferable to exchange with a diseased seller, even if detected correctly. It is important for social exclusion that a missed detection is the worst possible outcome.

Disease detection maximizes a buyer’s expected utility, i.e., \(\mathcal{H}_b(x) \equiv \arg \max_{\mathbb{E}_x[U(i, j)]} \), with the (indirect) utility \(u : X \to \mathbb{R}_+\) given by \(u(x) \equiv \max_{\mathbb{E}_x[U(i, j)]}\).

**Proposition 1** (i) A buyer labels a seller with a signal \(x\) as “diseased” if \(x \geq x_b\), where

\[
x_b \equiv \mathcal{L}^{-1} \left( \pi(P) \left[ \frac{U(h, h) - U(d, h)}{U(d, d) - U(h, d)} \right] \right), \quad \text{and} \quad \pi(P) \equiv \frac{(1 - P)(1 - \eta) + P\delta}{P(1 - \delta) + (1 - P)\eta}.
\]

(ii) A buyer’s detection threshold \(x_b\) increases with the utility of correct rejections and missed detections; and decreases with the utility of false alarms and correct detections.

(iii) A buyer’s (indirect) utility \(u(x)\) is everywhere continuous and strictly decreasing in \(x\), with a kink at \(x_b\); and it increases weakly with the ex-post utilities.

**Proof.** The proof is direct. For instance, \(\mathcal{H}_b(x) = d\) iff \(P'(d|x)/P'(h|x) \geq [U(h, h) - U(d, h)]/[U(d, d) - U(h, d)]\). Using (2) and \(\mathcal{L}(x)\) yields (3) and (i). By Assumption 1, \(x_b\) is unique and interior. As \(\mathcal{L}(x)\) is increasing in \(x\), its comparative statics in (ii) follow from (3) with

\[
u(x) = \begin{cases} U(h, h) - P'(d|x)[U(h, h) - U(h, d)] \quad &\text{if } x < x_b \\ U(d, h) - P'(d|x)[U(d, h) - U(d, d)] \quad &\text{if } x \geq x_b. \end{cases}
\]

Write \(P'(d|x) = \mathcal{L}(x)/[\pi(P) + \mathcal{L}(x)]\). Since \(P'(d|x)\) is increasing in \(x\) (Assumption 1), Assumption 2 implies that (4) is strictly decreasing in \(x\). (Assuming differentiability in \(\mathcal{L}(x)\) yields \(P'_x(d|x) = \pi(P)\mathcal{L}_x(x)/[\mathcal{L}(x) + \pi(P)]^2 > 0\).

Continuity in \(x\) follows since the maximum of two continuous function is continuous. Since \(u(x)\) depends positively on all utilities \(U(i, j)\), an increase in any of them either increases \(u(x)\) or leaves it unchanged.  

---

3 For ease of exposition, I sometimes use differentiability in \(\mathcal{L}(x)\), but differentiability is not essential.
Figure 1 illustrates the previous proposition. Buyers label sellers as “diseased” when their observed disease sign $x$ is high enough relative to the threshold $x_b$. The threshold depends on the stochastic process for health types, summarized by $\pi(P)$, and on the ex-post consequences of each decision. Proposition 1(ii) says, for example, that a lower utility for missed detections $U(h,d)$ leads to a lower detection threshold $x_b$, and a more common use of “diseased” labels. Comparative statics for $x_b$ can be seen directly from Figure 1. (The threshold $x_b$ also depends on the prior $P$, but priors here are arbitrary.)

Proposition 1(iii) says that buyers rank sellers according to their disease signs. A healthy-looking seller dominates unhealthy-looking ones. Suppose that a buyer receives a low signal with no scope for detection errors, i.e., $x = x_-$. Then, $u(x_-) = U(h,h)$ in (4) as there is certainty about healthy types. As $x$ increases, buyers face the prospect of detection errors. The probability $P'(d|x)$ determines the likelihood of errors and $U(h,h) - U(h,d)$ their utility cost; see (4). At $x = x_b$, buyers switch to a “diseased” label that carries detection errors that now depend on $U(d,h) - U(d,d)$, as seen in (4). A high signal with no scope for errors, i.e., $x = x_+$, yields $u(x_+) = U(d,d)$; see also Figure 1.

**Sellers.**— Sellers act in a symmetric way. After observing a disease sign $y \in X$, a seller’s non-strategic classification of buyers uses $\mathcal{H}_s(y)$ and the previous conventions. Costs reverse the order of Assumption 2. For a seller, a missed detection is also the worst possible outcome:
**Assumption 3.** Ex-post costs satisfy \( C(h, d) > C(d, d) > C(d, h) > C(h, h) \).

A seller’s disease detection decision yields \( \mathcal{H}_s(y) \equiv \arg \min_i \mathbb{E}_y[C(i, j)] \) with the (indirect) cost \( c : X \to \mathbb{R}_+ \) given by \( c(y) \equiv \min_i \mathbb{E}_y[C(i, j)] \). Thus,

**Proposition 2** (i) A seller labels a buyer with a signal \( y \) as “diseased” if \( y \geq y_s \), where

\[
y_s \equiv \mathcal{L}^{-1} \left( \pi(P) \left[ \frac{C(d, h) - C(h, h)}{C(h, d) - C(d, d)} \right] \right).
\]

(ii) A seller’s detection threshold \( y_s \) decreases with the cost of correct rejections and missed detections; and increases with the cost of false alarms and correct detections.

(iii) A seller’s (indirect) cost \( c(y) \) is everywhere continuous and strictly increasing in \( y \), with a kink at \( y_s \); and it increases weakly with the ex-post costs.

The intuition and proof repeat previous arguments, so I omit them. For future reference, the seller’s (indirect) cost,

\[
c(y) = \begin{cases} 
C(h, h) + P'(d|y)[C(h, d) - C(h, h)] & \text{if } y < y_s \\
C(d, h) + P'(d|y)[C(d, d) - C(d, h)] & \text{if } y \geq y_s,
\end{cases}
\]

varies with \( y \) inversely to the way \( u(x) \) varies with \( x \). I depict Proposition 2 in Figure 2, which is a symmetric version of Figure 1.

**Meetings, value functions, and trading strategies.** Exchange is pairwise and anonymous. Meetings take place through a random matching process. The fraction of sellers is \( s \in (0, 1) \). The fraction of buyers is \( b = 1 - s \). The rate at which a seller finds a buyer is proportional to the number of buyers, \( b \). A buyer finds a seller in proportion to the number of sellers, \( s \). It takes a period for a buyer (resp. seller) to find a seller (resp. buyer). Time is discounted at a rate \( r > 0 \).

Buyers and sellers are risk-neutral. There is a single non-storable and indivisible consumption good endowed to sellers. The *value functions* of a buyer and a seller are \( B(P) \) and \( S(P) \), respectively. Beliefs evolve according to (2) and are the only state variable. Updating depends on the observed signals \( x \) and \( y \), but I focus on a stationary equilibrium with common beliefs.
Figure 2: A seller’s (indirect) cost as a function of the signal $y \in X$ and threshold $y_s$. The direct costs $C(i, j)$ are ordered as in Assumption 3.

Trading decisions depend on the observed signals. A buyer finds trade agreeable with a seller featuring a disease sign $x$ according to a probability $\Pi_b(x)$. The probability that a seller finds trading agreeable with a buyer is given by $\Pi_s(y)$. Since trading decisions are strategic, buyers and sellers must also conjecture the probability that their trading partners will accept a trade. I assume that, after their beliefs are updated, traders observe their own disease signs and form trading conjectures $\pi_s(y)$ and $\pi_b(x)$. The probability $\pi_s(y)$ represents a buyer’s conjectured belief that a seller will undertake an exchange after seeing a signal $y$, and the probability $\pi_b(x)$ is a seller’s conjectured belief that a buyer will undertake an exchange after seeing a signal $x$. Trade requires correct conjectures and mutual agreement, i.e., $\pi_s(y) = \Pi_s(y)$ and $\pi_b(x) = \Pi_b(x)$ for correctness, and $\Pi(x, y) = \Pi_b(x)\Pi_s(y)$ for mutual agreement. Since signals are continuous, it is intuitive to consider a non-cooperative equilibrium in pure strategies.

Upon trading, a buyer gains $u(x)$ in utility and transitions into a seller whose lifetime utility is $S(P')$, as in

$$B(P) = \frac{1}{1 + r}\{s\Pi(x, y)[u(x) + S(P')] + s[1 - \Pi(x, y)]B(P') + (1 - s)B(P)\}. \quad (7)$$

The first term in (7) is the likelihood of meeting a seller times the trading probability and the meeting surplus, i.e., the instant payoff $u(x)$ and the value of becoming a seller $S(P')$
with an updated belief $P'$. The second term is the expected value of meeting a potential seller, not trading, and updating beliefs to obtain $B(P')$. The third term accounts for the possibility of no meetings and no belief updating. The seller’s value function is symmetric. A seller finds a buyer with probability $b$, incurs an expected cost $c(y)$, and transitions in value to $B(P')$.

In terms of flow values, the relevant Bellman equations are given by

$$rB(P) = s\Pi(x, y)\{u(x) + S(P') - B(P)\} + s[B(P') - B(P)],$$

and

$$rS(P) = b\Pi(x, y)\{B(P') - S(P) - c(y)\} + b[S(P') - S(P)].$$

I consider a standard notion of equilibrium:

**Definition 1** A stationary Bayesian equilibrium is a disease prevalence belief $P^* : X \rightarrow [0, 1]$ and a trading strategy $\Pi^* : X^2 \rightarrow \{0, 1\}$ such that beliefs are time invariant, common among traders, and agree with the evolution of health types; and buyers and sellers agree to trade if there are mutually beneficial gains.

**Beliefs.** To construct the equilibrium, I start with the disease prevalence beliefs $P^*(d|x)$. To ensure uniqueness, I first restrict the transition matrix $Q$ in (1):

**Assumption 4.** Health types are transient, with $\eta$ and $\delta$ satisfying $\eta + \delta < 1$.

The first part of Assumption 4 rules out fully persistent health types. Full persistence leads to complete learning and a trivial detection problem in equilibrium. The second part ensures that the stochastic process driving beliefs is monotone. Monotonicity is needed for the existence of an invariant distribution; see, e.g., Stokey et al., ([49], Section 12.4).

**Proposition 3** The stationary prevalence of disease in the population is $q^* = \eta/(\delta + \eta)$, and disease prevalence beliefs have a unique invariant distribution approximated by a (logistic) distribution

$$P^*(d|x) \simeq \frac{\sqrt{\eta L(x)}}{\sqrt{\delta} + \sqrt{\eta L(x)}},$$

(8)
Proof. The stationary distribution of types, \((q^*, 1 - q^*)\), follows directly from \(Q\). Updated beliefs \(P^*\) map \(P\) through \(T: [0, 1] \rightarrow [0, 1]\) as in \(P' = T(P) \equiv \mathcal{L}(x)/[\pi(P) + \mathcal{L}(x)]\). Stationary beliefs are a fixed point of \(T(P)\).

Updated beliefs are \textit{bounded}, i.e.,

\[
T(0) = \frac{\eta \mathcal{L}(x)}{\eta \mathcal{L}(x) + (1 - \eta)}, \quad \text{and} \quad T(1) = \frac{(1 - \delta) \mathcal{L}(x)}{(1 - \delta) \mathcal{L}(x) + \delta}.
\]

(9)

Hence, even if there is certainty in prior beliefs, i.e., \(P = \{0, 1\}\), there will be uncertainty about next period’s disease prevalence due to the Markov nature of health. (Under \(\eta = \delta = 0\), uncertainty is dissipated as \(T(0) = 0\) and \(T(1) = 1\).)

The mapping \(T(P)\) is also \textit{continuous} and \textit{strictly increasing} whenever \(\delta + \eta < 1\), which is Assumption 4. Since \(T(P)\) satisfies the boundary conditions (9), there is a fixed point \(P^*\) that solves \(P^* = T(P^*)\) or \(P^*\pi(P^*) = (1 - P^*)\mathcal{L}(x)\). Using \(Z \equiv P^*/(1 - P^*) = \mathcal{L}(x)/\pi(P^*)\), the fixed-point is the (positive) solution to the quadratic equation \(\delta Z^2 + Z[(1 - \eta) - (1 - \delta)\mathcal{L}(x)] - \eta \mathcal{L}(x) = 0\), that can be approximated by \(Z \simeq \sqrt{\eta \mathcal{L}(x)/\delta}\), which yields (8).

To understand Proposition 3, consider first fully persistent types.\(^4\) With \(\eta = 0\), the prevalence of disease in the population is \(q^* = 0\) as the healthy state is absorbent. In equilibrium, buyers and sellers correctly view all trading partners as healthy since \(P^*(d|x) = 0\) for all \(x \in X\). Likewise, under \(\delta = 0\), the prevalence of disease is \(q^* = 1\) as the diseased state is absorbent, and all partners are correctly perceived to be diseased, i.e., \(P^*(d|x) = 1\) for all \(x \in X\) in (8). With fully persistent types, disease detection is trivial in equilibrium.

Under transient types (Assumption 4), \(q^*\) is interior and learning is incomplete with common disease prevalence beliefs that converge monotonically to \(P^*(d|x)\). The limiting values in (8) are \(P^*(d|x_-) = 0\) and \(P^*(d|x_+) = 1\).

\textit{Trading strategies.–} The stationary disease prevalence beliefs in Proposition 3 determine the equilibrium trading strategies. Given the signals \((x, y)\), the stationary ‘capital’ gains for buyers are:

\[
B(P^*) - S(P^*) = \frac{\{su(x) + bc(y)\} \Pi^*(x, y)}{\Pi^*(x, y) + r}.
\]

(10)

\(^4\)Under \(\eta = 0\), \(P^* = 0\). For \(\delta = 0\), the two solutions of \(P^*\pi(P^*) = (1 - P^*)\mathcal{L}(x)\) are 1 and \(\eta \mathcal{L}(x)/[1 - (1 - \eta)\mathcal{L}(x)]\). The second solution is inconsistent with the stationary distribution of types.
A choice of \( \Pi^*(x, y) = 1 \) requires \( u(x) \geq B(P^*) - S(P^*) \geq c(y) \). Evaluating (10) at \( \Pi^*(x, y) = 1 \) implies that, in equilibrium, (pure) trading strategies require

\[
\frac{(r + b)u(x) - bc(y)}{1 + r} \geq 0, \text{ and } \frac{su(x) - (r + s)c(y)}{1 + r} \geq 0.
\]

(11)

I require some additional assumptions on the ex-post utilities and costs:

**Assumption 5.** Regardless of the assigned disease classification, it is mutually beneficial for healthy types to trade and for diseased types not to trade.

In a frictionless world, gains from trade require \( U(i, i) - C(i, i) > 0 \). The time cost of search, captured by \( r \), and the randomness of meetings, captured by \( s \), tax these gains. A complete version of Assumption 5 orders all ex-post utilities and costs, and is available in the Appendix. A particular set of conditions in Assumption 5 is \( sU(h, h) - (r + s)C(h, h) > 0 > sU(d, d) - (r + s)C(d, d) \). If the first part of these conditions is not satisfied, frictional trade will not be beneficial under any circumstance. If the second part is not satisfied, trade will be beneficial under all circumstances.

**Proposition 4** Trading follows a threshold strategy. The acceptance threshold is determined by the seller and given by \( \bar{y}(x) \equiv \{ y \in X : su(x) - (r + s)c(y) = 0 \} \). The acceptance threshold is everywhere continuous and strictly decreasing in \( x \); and it weakly shifts outward as ex-post utilities increase or ex-post costs decrease.

**Proof.** The proof, in Appendix B, relies on the detection thresholds.

Since sellers incur a cost, while buyers enjoy a benefit, exchange is always limited by the seller’s side. In a threshold strategy, trading probabilities are \( \Pi^*(x, y) = I\{(x, y) : y \leq \bar{y}(x)\} \). Proposition 4 says that trading takes place when traders feel ‘safe.’ It is unsafe for a seller to carry out transactions with a buyer that exhibits severe signs of illness. The threshold \( \bar{y}(x) \) behaves like an indifference curve. Along the threshold, an unhealthy signal in a buyer is compensated by a healthier signal in a seller, say. This means that sellers will accept unhealthy trades. Their willingness to accept unhealthy partners is larger for healthy-looking sellers than for unhealthy-looking ones. Figure 3, for instance, depicts the
acceptance threshold and shades the acceptance set. As the figure shows, a seller who displays the lowest disease sign \( x = x_- \) has the largest acceptance set. Such seller is willing to accept any buyer who displays a disease sign \( y \in [x_-, \bar{y}(x_-)] \).

The acceptance set summarizes how disease and fear of contagion limit the extent of the market. The acceptance set is non-empty. A signal \( x_- \) yields \( sU(h, h) - (r + s)C(h, h) > 0 \) and \( \bar{y}(x_-) > x_- \). The acceptance set is also bounded so exclusion from economic transactions is a potential outcome. That is, there are equilibrium outcomes in which individuals are unwilling to trade. In particular, a signal \( x_+ \) yields \( sU(d, d) - (r + s)C(d, d) < 0 \) so sellers observing any disease signs such that \( y \in [\bar{y}(x_+), x_+] \) are unwilling to trade (Figure 3). Finally, the acceptance threshold is monotone. For instance, if \( y' \leq \bar{y}(x') \), then \( y'' \leq \bar{y}(x'') \) for any signals \( y'' < y' \) and \( x'' < x' \). Hence, if a buyer-seller pair with signals \( (x', y') \) are willing to trade, they would also be willing to trade with healthier-looking partners. If a buyer-seller pair is unwilling to trade, they would also be unwilling to trade with unhealthier-looking partners.

Since healthy-looking buyers are more valuable in the presence of healthy-looking sellers, the acceptance set features (weak) type complementarities. It is, however, not possible to sort healthy and diseased types, so mixing is always observed in equilibrium. The healthiest-looking sellers are willing to trade with “clean” and “diseased” buyers, i.e., \( \bar{y}(x_-) > y_s \) (see Figure 3): Optimal trading strategies for \((x, y)\), and a seller’s acceptance threshold \( \bar{y}(x) \).
Figure 3), but the unhealthiest-looking sellers can only trade with “clean” buyers, i.e.,
\[ \bar{y}(x_+) < y_s. \]

**Some remarks.** I next present some informal remarks on the model. Extensions that consider *stigma* and dynamic contagion externalities are available in Appendix A.

**Social tolerance to disease.** Disease detection and exchange are discrete problems. Both rely on thresholds and a sense of *tolerance*. Potential trading partners are perceived as ill if their disease signs exceed a tolerable margin. Trade is also perceived as safe within the margins determined by the acceptance threshold. The thresholds serve the purpose of socially regulating disease. Social regulation is not exclusive to communicable diseases. Mental health and criminal justice systems are also in charge of identifying and isolating dangerous and threatening individuals. Social control often relies on classifications based on “deviance” in behavior from psychosocial, ethical, and legal norms.\(^5\)

Since detection errors depend on \( x_b \), the social tolerance to disease is manifested as a trade-off between detection errors,

\[
\Pr\{H_b(x) = d|h\} = \int_{x_b}^{x_*} p(x|h)dx, \text{ and } \Pr\{H_b(x) = h|d\} = \int_{x_-}^{x_b} p(x|d)dx .
\]

If the detection threshold \( x_b \) decreases, for example, false alarms become more prevalent at the expense of less prevalent missed detections.

False alarms are preferable and less costly than missed detections (Assumptions 2 and 3). With communicable diseases, the ex-post consequences of a false alarm are less severe than those of a missed detection. A false alarm may lead to unnecessary exclusion and labeling (i.e., over-diagnosis) but missing a detection might lead to widespread disease or death.\(^6\) As Oaten et al. ([42], p. 3435) note, the ex-post preference for false alarms implies “reacting to relatively scant evidence that someone is harboring a contagious disease, but requiring

\(^5\)A familiar form of social control is the confinement of mental health patients in mental institutions. Dangerousness to society is often a guiding principle for social control; see Pfohl [43] for studies on the sociology of deviance and social control.

\(^6\)A preference for false alarms can be rooted in evolutionary principles. Shermer [47] makes an informal but compelling case for the widespread use of false alarms in evolutionary contexts relevant for market transactions. Nesse [41] provides a formal evolutionary analysis.
much stronger evidence that someone is healthy, [...] especially if the threat of disease is highly salient or if we feel particularly vulnerable to disease.” I present what appears to be instances of false alarms in the social regulation of disease in the past later on.

**Discrimination.**— Traders prefer to look and be healthy. They also prefer healthy-looking partners. Social exclusion takes the simple form of sellers ‘refusing service’ to diseased-looking buyers. Social exclusion, however, is not only driven by a taste for discrimination. The (indirect) utilities and costs are endogenous and depend on the stochastic properties of disease, as in models of statistical discrimination (see; e.g., Fang and Moro [20] for a survey). The taste element that favors healthy types in Assumptions 2 and 3 is expressed as a detection problem. For example, disease tolerance and false alarms go together in theory and in the practical regulation of disease during past epidemics, as Section 4 hopes to show.

Buyers (sellers respectively) need to conjecture the probability that sellers (resp. buyers) will agree to trade. I have, however, ruled out strategic interactions in disease detection. For instance, I ignored concealment of disease signs. Concealment, if it operates to alter the given conditional densities associated with a signal, can be incorporated by making the densities \( p(x|d) \) and \( p(x|h) \) endogenous. Some disease signs, however, are very difficult to conceal because infection induces “expulsive defenses” (i.e., spitting, vomiting, diarrhea, coughing, sneezing, and rhinorrhea). Evolutionary pressures have also enhanced disease detection mechanisms from the coevolution between humans and parasites; see, e.g., Kurzban and Leary [30] and Thornhill and Fincher [51].

The model also ignored the possibility of a *self-fulfilling* equilibria where a social label induces a behavior consistent with the label, even in the absence of the conditions that lead to such a label. Buyers here do not directly react to a seller’s beliefs, and vice versa. Endogenous responses may lead to a self-fulfilling equilibrium where beliefs influence human capital accumulation, as in Coate and Loury [10] or Lundberg and Startz [33]. I emphasize learning about underlying health types as a way to curb incorrect beliefs.

---

7Hamermesh [23] discussed beauty in modern market settings. He framed an observed “beauty premium” as a consequence of ‘lookism’ and social productivity. These are individual or social preferences, but both are deterministic and not associated with detection problems.

8An example is the \( \varepsilon \)-contamination model with \( p(x|i) = (1 - \varepsilon)p^0(x|i) + \varepsilon p^\delta(x|i) \), where an original density is contaminated by a deviation according to \( \varepsilon \).

9The model centered on the gains from trade, not on the division of the surplus. Prices in pairwise models
I focused on the way beliefs induce social exclusion in random social encounters. Individuals here are not ignorant about health types, but the lack of perfect information makes social exclusion look like prejudice. Disease prevalence priors in Propositions 1 and 2 are exogenous and may be biased, but the disease prevalence beliefs in Proposition 3 are “correct.” One can consider overt prejudice in the stationary equilibrium as a distorted belief, but overt prejudice is not essential to socially regulate disease.10

3 Comparative Statics: Danger and Visibility

Equilibrium uniqueness allows for simple comparative statics on the characteristics of disease. I associate danger with the ex-post utilities and costs of disease detection and visibility with the stochastic properties of disease. Danger and visibility are central in disease avoidance (i.e., Kurzban and Leary [30]) and in many social contexts, some briefly noted above. I use these organizing principles and historical case studies to interpret the social regulation of disease in past epidemics in Section 4. For convenience, I focus on the detection and acceptance thresholds and relegate remaining proofs to Appendix B.11

**Danger.** Missing a detection is critical for communicable diseases:

**Corollary 1** A disease with a lower ex-post utility and a higher ex-post cost of missed detections has lower disease detection thresholds and an inward-shifted trading acceptance threshold.

Corollary 1 implies that when society faces a disease with costlier missed detections, social tolerance to disease declines. More individuals are labeled as “diseased,” false alarms become widespread, and more potential trades are blocked. Costlier missed detections are associated with dangerous diseases and more severe consequences of under-diagnoses. If the ex-

---

10Suppose that a buyer observing a signal $x$ in a seller believes with certainty that the actual signal is distorted, as in $x' = x/\phi$ with $\phi > 0$. The relevant likelihood ratio would be of the form $L(x') = L(x'/\phi)$, and so detection decisions would rely on a biased threshold of the form $x'_b = x_b/\phi$.

11Search frictions have no influence on the detection stage. As $r$ declines, the acceptance set expands. With $r = 0$, (11) becomes $\min\{s, b\}[u(x) - c(y)] \geq 0$, which depends on the short side of the market.
post disutility and cost from missed detections are extreme, \( \{U(h, d), C(h, d)\} \to \{-\infty, \infty\} \), the thresholds satisfy \( \{x_b, y_s\} \to \{x_-, x_-\} \) and \( \bar{y}(x) \leq x_- \) for all \( x \in X \). Tolerance and safety are completely eroded as society collapses: “diseased” labeling and false alarms become universal, and trade comes to a halt. Missing a detection is infinitely costly so no potential trading partner is deemed safe.\(^{12}\)

**Visibility.** Visibility depends on the evolution of disease over time and the informational value of the observed disease signs.

**Persistence.** I consider persistence in the disease state first:

**Corollary 2** A disease with a higher degree of persistence has lower disease detection thresholds and an inward-shifted trading acceptance threshold.

Persistence in the disease state varies by disease type. Diseases are often characterized as *acute* or *chronic*. Generally speaking, acute conditions are transient. They have a rapid onset and an overall short duration (i.e., common colds). Chronic conditions, on the other hand, are persistent. They are long-lasting and often worsen over time.\(^{13}\)

Corollary 2 recognizes that the social response to a chronic disease differs from that of an acute disease. When society faces a more persistent disease, the stationary prevalence of diseased types in the population and the disease prevalence beliefs \( P^*(d|x) \) increase; see Proposition 3. As a consequence, tolerance to disease signs declines. As \( \delta \to 0 \), all individuals are correctly labeled “diseased” and trade stops, i.e., \( \{x_b, y_s\} \to \{x_-, x_-\} \) and \( \bar{y}(x) \leq x_- \) for all \( x \in X \), leading also to societal collapse. (In contrast, under \( \eta \to 0 \) all individuals are correctly labeled as “clean” and trade is always beneficial.)

**Signal-to-noise.** Visibility also depends on the densities \( p(x|d) \) and \( p(x|h) \). To illustrate the role of the signal-to-noise content, I consider log-normal signals with opposite means and equal variance. That is, for the disease state, \( p(x|d) = (x\sigma\sqrt{2\pi})^{-1} \exp\{-[\ln x - \mu]^2/2\sigma^2\} \),

\(^{12}\)Figures 1 and 2 make it is easy to visualize the societal collapse. From the buyer’s perspective, as \( U(h, d) \to -\infty \), the value of a “clean” label \( \mathbb{E}_d[U(h, j)] \) becomes infinitely steep and a “diseased” label \( \mathcal{H}_b(x) = d \) becomes pervasive. A similar (but reversed) logic applies to Figure 2 for the seller.

\(^{13}\)Pathogens adapt by natural selection. There is no apparent evolutionary advantage associated with chronic or acute infections. Pathogens seem to trade-off transmission and virulence. As noted by Weiss [55], “[p]ersistent infections can afford to be less contagious than acute infections.” For instance, a persistent infection does not require a constant influx of new hosts. Alizon et al. [2], especially their Box 1, contains a recent overview of the many trade-offs in infectious disease dynamics.
and \( p(x|h) = (x\sigma\sqrt{2\pi})^{-1}\exp\{-[\ln x + \mu]^2/2\sigma^2\} \) for the healthy state. The spread between \( p(x|d) \) and \( p(x|h) \) is determined by the amplitude of the signal \( \mu \). The dispersion within each density is given by the amplitude of the noise \( \sigma \). The signal-to-noise ratio is \( (\mu/\sigma)^2 \), and the likelihood ratio is \( \mathcal{L}(x) = \exp\{2\alpha \ln x\} \). The parameter \( \alpha \equiv \mu/\sigma^2 \) determines the overlap in the previous densities.

**Corollary 3** A disease with less informative signals, i.e., a lower signal-to-noise ratio, has lower disease detection thresholds and an inward-shifted trading acceptance threshold.

Discriminating between the diseased and the healthy state is more difficult with weaker or noisier signals, i.e., lower \( \alpha \). To understand the social response to a disease with uninformative signals, consider first ‘perfect’ detection. As \( \alpha \to \infty \), traders are able to perfectly discriminate between healthy and diseased types. Since signals have opposite means, detection is ‘perfect’ when a positive sign (i.e., \( \ln(x) > 0 \) or \( x > 1 \)) receives a “diseased” label whereas a negative sign (i.e., \( \ln(x) < 0 \) or \( x < 1 \)) induces a “clean” label. ‘Perfect’ detection thus yields \( \ln(x_b) = 0 \) or \( x_b = 1 \).

Consider then Corollary 3. This result recognizes that when facing a disease with a low signal-to-noise ratio, society also becomes less tolerant to disease signs. As \( \alpha \) declines, detection thresholds decline toward \( x_- = 0 \), and the acceptance threshold (A19) shifts inward toward \( y(x) \leq x_- \) for all \( x \in X \). The limiting case of uninformative signals, i.e., \( \alpha \to 0 \), also leads to societal collapse because disease signs have the same predictive ability as randomization.

## 4 Epidemics: Case Studies

This section interprets ways in which social exclusion and labeling were expressed in the past. I explore the social regulation of disease along danger and visibility lines, including instances where social exclusion took racial, ethnic, and religious turns because apparent “false alarms” lead to ‘unreasonable’ exclusionary policies.

\(^{14}\text{It is possible to consider differences in the variance, but in such case, the log-likelihood ratio is a quadratic function so it violates Assumption 1.}\)
The discussion of each disease is brief and it deals only with the aspects relevant for the model. (Kiple [28] provides an accessible history of disease tailored to social scientists.) I rely on the authoritative historical treatments of the social response to epidemics by Cohn [11], McNeill [37], Ranger and Slack [45], and Watts [53] (among many others), but I use danger and visibility as the central threads of my analysis. To be consistent with the model, my discussion concentrates on epidemic diseases introduced to Western Europe and its offshoots through trading and travel. The case studies obviously do not provide an exhaustive treatment or a complete explanation of social exclusion.

The discussion is organized by disease but there is an implicit chronological order as these epidemics were sequential. In Europe, leprosy became widespread in the twelfth and thirteenth centuries, plague visited Western ports and cities in a series of pandemics that started in the fourteenth century (and extended until the twentieth century), smallpox epidemics gained visibility during the seventeenth and eighteenth centuries, and cholera was one of the most feared diseases of the nineteenth century. The many diseases in the Pandemic Century (Honigsbaum [24]), the century starting in 1918, require at a minimum a dedicated paper.

**Medieval leprosy.** Known as Hansen’s disease, leprosy is a chronic infectious disease caused by the Mycobacterium leprae (\textit{M. leprae}) pathogen. Leprosy is likely original from India, but prevalent in ancient China and Egypt. Its introduction into western Asia and Europe is often credited to Alexander the Great’s troops.

Leprosy targets primarily the skin, the peripheral nervous system, the lining of the nose, the eyes, and the upper respiratory tract; see Kiple ([28], p. 837). Leprosy caused significant disabilities and severe disfigurement because of skin ulcers, nerve damage, and muscle weakness.\textsuperscript{15} As noted by Dols ([16], p. 891) “no disease is so fearsome and horrible as leprosy.” Leprosy conveyed “physical repulsiveness, moral perversion, and promiscuous

\textsuperscript{15}Leprosy is curable with multidrug therapy. Progress against Hansen’s disease has been limited because \textit{M. leprae} is unculturable in vitro, i.e., it cannot be grown in axenic culture medium where only one type of microorganism is growing. A review published in 2020 by Ploemacher et al. [44] note that “[s]olid evidence exists of an increased risk for individuals living in close contact with leprosy patients, most likely through infectious aerosols, created by coughing and sneezing.” Their systematic review also notes “that human-to-human transmission is not the only way leprosy can be acquired. The transmission of this disease is probably much more complicated than was thought before.”
infection; the leper is the archetypal outcast.” Leprosy, however, is not very contagious and
has low pathogenicity. Many people are infected but “only a few of the supposedly infected
persons [about 5 to 10 percent] develop clinical disease” as most of the population possess
natural immunity. Overall, in their review, De Beers et al. [4] note that “M. leprae is a poor pathogen.”

During the high Middle Ages, from around the First Crusade in 1099 to the Black Death
in the 1350s, Western Europe witnessed an epidemic outbreak of leprosy. “[T]owards the end
of antiquity,” according to Moore ([39], p. 69), leprosy “reached some parts of the Roman
world, including Egypt, France and England, but did not become widespread” until “a
great epidemic […] seized Europe in the twelfth and thirteenth centuries.” Accompanying
its spread, Moore ([39], p. 53) describes “undoubted signs of increasing nervousness of
the contagion of leprosy” in historical narratives. Besides Europe, medieval leprosy was
widely documented in the Near East, North Africa, and Andalusia; see, e.g., Dols [16].
The most virulent form of Hansen’s disease, lepromatous leprosy, gradually disappeared
from Europe after the Black Death; see, e.g., McNeill ([37], p. 128). As noted by Moore
([39], p. 69), lepromatous leprosy became a “rarity in most regions by the fifteenth century
and was virtually extinct by the seventeenth.” Leprosy continued to exist at large only in
Scandinavia.

An interpretation.– Leprosy is the prime example of a socially regulated disease. Social
regulation took place along religious lines, but according to Barber ([3], p. 14), “it was the
community which usually had to decide if a person had contracted the disease.” The “most
familiar and painful [image] that the medieval world has to offer” is that of “the wandering
leper” carrying a “begging bowl,” which could not be touched, and ringing a warning bell
or a clapper with “a yellow cross sewn to their cape or vestment;” see Moore ([39], p. 51)
and Kiple ([28], p. 838). Islam also expressed a fear of leprosy but less ferocity in enforcing

16 As Kilwein ([27], p. 186) notes, in Europe, “the Council of Lyons (AD 583) restricted the association
of lepers with sound persons,” and the Council of Orléans “in 549 ordered that lepers were to be given food
and clothing by bishops,” as noted by Moore ([39], p. 45). “In 1179 the Third Lateran Council reiterated
that lepers should be segregated, and were forbidden to go to church or to share churches and cemeteries
with the healthy.” These principles translated into “local and municipal regulations for the control and
isolation of lepers, such as those which forbade them to walk the streets of London in 1200, Paris and Sens
in 1202, Exeter in 1244;” see Moore ([39], pp. 45, 55).
social exclusion. Dols ([16], p. 906) noted that, in attribution to the Prophet, “a Muslim should flee from the leper as he would flee from the lion.” The fear associated with leprosy is likely attributable to the Old Testament; see, e.g., Kilwein [27].

The so-called Law of Leprosy in Leviticus regulates all matters regarding leprosy, including social labelling and exclusion:

“And the leper in whom the plague is, his clothes shall be rent, and his head bare, and he shall put a covering upon his upper lip, and shall cry, Unclean, unclean. All the days wherein the plague shall be in him he shall be defiled; he is unclean; he shall dwell alone; without the camp shall his habitation be.”

Leprosy’s central features are its high visibility, due to high signal-to-noise ratios and high persistence, and low danger. It is a chronic condition whose symptoms can take as long as two decades or more to develop. It also multiplies slowly and has a long incubation period to the point of being incorrectly viewed as hereditary in the past. In terms of the model’s predictions, leprosy can be seen as a highly persistent disease, i.e., a disease with a low value for $\delta$ (Corollary 2) and a high value of $\eta$ granted by a high rate of natural immunity. Second, its primary disease signs are highly visible, i.e., leprosy is a disease with a high signal-to-noise ratio $\alpha$ (Corollary 3). Leprosy is degenerative and it can lead to deformities in the hands and feet, and more importantly, in almost all parts of the face (i.e., destruction of the nasal septum, paralysis in the eyelids, and enlarged earlobes). As noted by Oaten et al. ([42], pp. 3435-3436), disease signs “directly displayed via the face” are highly visible because the face “represents the initial focus in social encounters.”

In light of the model’s predictions, the ancient fear and selective regulation of leprosy are ‘reasonable’ given its high visibility. The low tolerance for the disease signs associated with leprosy likely led to several false alarms. In the past, as noted McNeill ([37], p. 155) “leprosy

---

17 Leviticus provides purification rituals for those who had been “cured of the disease,” a highly unlikely outcome. Dols [16] also notes a “frequent confusion with other skin disorders” in the Islamic world. It has been long known (McEwen [36], p. 200) that the listed disease signs in the Leviticus, “viewed medically, [...] present simply a grouping of indefinite descriptions, applicable [...] to many forms of the skin diseases of the inflammatory type, including leprosy, but characteristic of none.” Kilwein ([27], p. 185) argues that “there must have been many misdiagnosed cases of leprosy.” As noted by Moore ([39], p. 43), “even today a standard medical textbook warns that a doctor who is not looking for leprosy may easily miss it, while one who is is liable to see it everywhere.”
was, of course, a generic term used to describe a number of different infections that affected the skin in conspicuous and horrible ways.” Even today, dermatological disorders are often associated with social exclusion even for non-contagious conditions such as birthmarks and psoriasis. Other examples include an almost universal “severe social stigma” to superficial skin conditions and physical distortions of the body (i.e., amputations). In Indian society, vitiligo (a common depigmentation condition) deems people “unmarriageable,” as noted by Oaten et al. ([42], p. 3443).

False alarms also manifested as an ‘unreasonable’ fear of contagion and contamination. In Medieval Europe, the social exclusion of lepers was accompanied by a much broader emergence of a “persecuting” society; see Moore [39]. Not only were secular and religious authorities concerned with lepers, other groups such as homosexuals and prostitutes, but especially heretics and Jews, were subject to exclusion and persecution. Social regulation took many forms. For example, as Moore ([39], p. 62) notes, “[Jews] too were required to advertise in their dress the segregation from society at large which was institutionalized by imprisoning heretics and confining lepers to lazaret houses or villages and Jews to their increasingly strictly defined residential quarters in the cities. In all three cases [lepers, heretics, and Jews], exclusion from the community extended to civil rights, denying access to public courts and office, security of property during life and disposal of it after death.”

The persecution and segregation of Jews was often based on “delusional” conspiracies with lepers to either kill or “make leprous all the Christians of France and Germany.” A particular instance involving drinking wells in 1321 in France has been discussed in great detail by Barber [3]. 18 Barber [3] also notes that “the Jews were of course the traditional victims of medieval prejudice, especially during the crusading era.” He further notes, however, that “[t]he association of the Jews with the lepers is more difficult to identify. It may have been the revival of an old prejudice which finds its written origins as long ago as the third century B.C., in which the Jews were chased out of Egypt as impure and leprous [...] or it may stem from a contrary idea based on the belief that the Jews rarely contracted leprosy, and could therefore have plotted with the lepers without fear of infection.”

18 The Inquisitor at Toulouse spoke of “an evil plan of the lepers against the healthy persons in the kingdom of France.” Often, the charges involved non-Christians: “the Jews [...] arranged for the actual poisoning to be done by the lepers, who continually mix with the Christians,” see Barber ([3], pp. 1, 10).
**Plague pandemics.** Plague is an acute bacterial disease caused by the Yersinia pestis (*Y. pestis*), typically found in small mammals and their fleas. Humans are an accidental host for the plague. *Y. pestis* attacks the lymphatic system. The most common type of plague is bubonic. Bubonic plague is caused by the bite of an infected flea. When untreated, its case fatality rate is close to 60 percent. Bubonic plague does not spread by contact. The deadliest type is the pneumonic plague, which is transmitted by airborne droplets. Its case fatality rate is virtually 100 percent. Septicemic plague, also deadly but less visible, occurs when infection spreads through the bloodstream; see, e.g., Yang and Anisimov [56].

The incubation period for plague is between two to six days. After that time, individuals “display a sudden onset illness characterized by headache, chills, fever, malaise, and the appearance of a painful bubo” (i.e., swollen, blackened lymph nodes) that gives the disease its name. In the last stages, plague leads to septicemia, internal bleeding, and hemorrhagic pneumonia. Septicemic plague may occur “without prior evidence of a bubo. Occasionally, bubonic and septicemic infections progress to secondary pneumonic infections,” which are “nearly always fatal;” see Smiley ([48], p. 257).

Plague is not endemic to Europe. It was introduced by North African and Central and East Asian traders during three Old World pandemics: the Justinian plague (in 541), the Black Death (in 1347), and the Third Pandemic (in 1899). Plague outbreaks were explosive and lethal. It is not possible to even list all the aspects touched by bubonic plague in Europe here, but to roughly indicate, a few examples might help. McNeill ([37], pp. 109-113), for instance, views the first plague pandemic as key in “the failure of Justinian’s effort to restore imperial unity to the Mediterranean, [...] and the [success of the] first critical stages of the Moslem imperial expansion” during the seventh century. Ziegler [57] is a general introduction into the Black Death in Europe. He discussed in some detail the “considerable dislocation to its economy and its social structure,” including the increased spatial and social mobility of peasants; the changes in human capital accumulation, architecture, and agricultural technologies resulting from severe labor shortages; the movement away from Latin toward vernacular languages; and the weakening of the Church that paved the way for the Reformation.

Until the nineteenth century, there was an almost complete ignorance of the pathology
of plague. The Third Plague Pandemic took place when the germ theory was becoming established, making it possible to characterize the disease; see Dols ([15], Chapter 3). The Third Plague Pandemic originated in southwest China. It reached Guangzhou (Canton) and Hong Kong in 1894, where Alexandre Yersin isolated the *Y. pestis* bacterium in culture and identified it under the microscope. By ship, plague spread to India, and later to the rest of the world. Between 1899 and 1901, significant outbreaks of plague were recorded in Alexandria, Mumbai (Bombay), Buenos Aires, Cape Town, Glasgow, Guayaquil, Honolulu, Lima, Manila, Porto, Osaka, Rio de Janeiro, and San Francisco, and Veracruz among many others; see Echenberg ([18], p. 432). Except for western Europe and Australia, there are now reservoirs of *Y. pestis* in all regions of the world.

An interpretation.— Plague is a dangerous and deadly disease. It “is one of the world’s most virulent human pathogens;” see Smiley ([48], p. 255). Most infections led to a gruesome and rapid death. As Boccaccio remarked, victims often ate lunch with their friends, and ate dinner with their ancestors in paradise.

Plague’s exceptionally rapid course is the consequence of “the lack of the normal protective innate immune response to infection;” see Smiley ([48], p. 256). In simple terms, the initial stages of the plague disable the immune response. In more technical terms, *Y. pestis* uses “several stratagems to evade the innate and the adaptive immune responses. For example, infections with this organism are biphasic, involving an initial ‘noninflammatory’ phase where bacterial replication occurs initially with little inflammation and following by extensive [‘proinflammatory’ phase] with considerable tissue destruction;” see Yang and Anisimov ([56], p. 273).19 These technical details are important because they imply that there were no reliable disease signs to discriminate between healthy and infected individuals in the early stages of plague. The detailed account of the Justinian plague (in 541) by Procopius, the Byzantine court historian, confirms it:

“with the majority it came about that they were seized by the disease without

19The innate immune system is the first to respond to a bacterial invasion. By the time the second line of defense takes place, the adaptive immune response, the body is already overwhelmed with bacteria. The pathogen’s adaptations seem more remarkable considering that *Y. pestis* has to infect a carrier. “Evolution also may have selected for the exceptional virulence of *Y. pestis*, since the death of infected mammals presumably facilitates transmission by compelling infected fleas to seek new hosts;” see Smiley ([48], p. 257). Yang and Anisimov [56] and Li and Yang [31] contain technical treatments of the pathology of plague.
becoming aware of what was coming either through a waking vision or a dream. They had a sudden fever, some when just roused from sleep, others while walking about, and others while otherwise engaged, without any regard to what they were doing. And the body showed no change from its previous colour, nor was it hot as might be expected when attacked by a fever, nor indeed did any inflammation set in, but the fever was of such a languid sort from its commencement and up till evening that neither to the sick themselves nor to a physician who touched them would it afford any suspicion of danger. It was natural, therefore, that not one of those who had contracted the disease expected to die from it.”

Plague’s central features are its high danger and invisibility, i.e., a low persistence with high values of \( \delta \), and low signal-to-noise ratio \( \alpha \). The fear of plague is ‘reasonable’ given its high mortality, infectivity, costs of missed detections, and its low initial signal-to-noise ratio. Plague, inspired by the isolation of lepers, led to the “legal enforcement against the introduction of disease” in cities during the Black Death; see Kilwein ([27], p. 185). Travelers were prevented from entering cities, plague victims were isolated, and contact tracing was introduced. Even without an understanding of its pathology, plague showed predictable transmission along trade routes.\(^{20}\) Plague is the sole disease responsible for the use of quarantines as a public health measure to socially regulate disease.\(^{21}\) Quarantines, however, were not as selective as the regulation of leprosy. Social exclusion applied even to travelers and traders who did not display any disease signs. Corollaries 1 to 3 suggest that a widespread and indiscriminate exclusion policy is likely the result of the very high danger and low visibility of plague.

The social control of the plague also led to several ‘unreasonable’ responses. The Black Death, for example, intensified the persecution of Jews and other minorities. “There was

\(^{20}\)The Black Death spread along trade routes. Port cities in the Mediterranean institutionalized quarantines. Across Europe, “there was a depressing readiness to stress that flight was the only possible defence against the plague,” which was available to “the nobles and the rich” only. King Alfonso from Castille “caught the disease and died on Good Friday, 26 March, 1350. He was the only ruling monarch of Europe to perish during the Black Death;” see Ziegler ([57], pp. 72, 114).

\(^{21}\)The Black Death also had a devastating impact in the Middle East; see Dols [15]. The Muslim response was quite different. There was a “prohibition against leaving a land that has been stricken by plague or entering a plague-stricken land.” The notion of contagion was not accepted in the Muslim world; see Dols ([15], p. 119). In Spain, Ibn Al-Khatib developed the notion of contagion, but his ideas were not accepted. He was later accused of heresy; see Dols [15], p. 62).
really only one charge levied against the Jews; that, by poisoning the wells of Christian communities, they infected the inhabitants with the plague;” see, Ziegler ([57], pp. 97-110). According to Ziegler ([57], p. 100), “[a] partial explanation may be that many wells in built-up areas were polluted by seepage from nearby sewage pits. The Jews, with their greater understanding of elementary hygiene, preferred to draw their drinking water from open streams, even though these might often be farther from their homes. Such a habit, barely noticed in normal times, would seem intensely suspicious in the event of a plague. Why should the Jews shun the wells unless they knew them to be poisoned and how could they have such knowledge unless they had done the poisoning themselves?” Jews also specialized in the trade of second-hand clothing. Given the fleas’ ability to transmit the disease, unintended cases of plague could have been transmitted through clothing; see Cohn ([11], Chapter 2).

Despite an enhanced understanding of its pathology, the Third Pandemic still featured ‘unreasonable’ responses. Plague reached Cape Town in 1901. The first cases of plague appeared “among Cape coloured and African dock workers;” Swanson ([50], p. 386). The epidemic “spurred the separation of poor whites and blacks.” Government and medical authorities associated Africans with the plague and “sought no less than the mass removal of Cape Town’s African population, even though the number of Africans contracting the plague was less than either whites or coloureds. [...] It was the merest step of logic to proceed from the isolation of plague victims to the creation of a permanent location for the black labouring class;” Swanson ([50], pp. 385-390). In 1902, the Native Reserve Location Act was passed authorizing the mass removal of Africans from the city center toward peri-urban locations. Systematic urban segregation extended to other cities in South Africa (e.g., Port Elizabeth, Durban, and Johannesburg).22

San Francisco and Honolulu also saw ‘unreasonable’ responses to plague. In both cities,

22Swanson ([50], p. 387) labeled the social control of disease using concepts of public health along racial hierarchies a sanitation syndrome. He further argued that “[t]his ‘sanitation syndrome’ can be traced as a major strand in the creation of urban apartheid.” In 1901, years after the discovery that the anopheles mosquito was the vector for malaria, the British decided upon a policy of racial residential segregation in their African colonies. Segregation was considered as the most practical solution against malaria, their most serious threat in the tropics. Malaria and yellow fever were controlled in Panama by controlling their vectors, but such task became too onerous in tropical Africa. Curtin [14] contains a detailed discussion of the role of medical knowledge in urban planning in tropical Africa.
fear led to social exclusion along racial lines. On March 6, 1900, the death of a Chinese resident of San Francisco’s Chinatown, “signaled the slow beginning of an epidemic” of plague in America; see Craddock ([13], p. 127) and Trauner ([52], p. 76). The first reaction to his confirmed cause of death was “the implementation of a cordon sanitaire around the entire district of Chinatown. [...] All whites were removed from the district before it was cordoned off; thereafter only whites could leave Chinatown, but no one could enter it;” Craddock ([13], p. 127). The San Francisco board of health rescinded the measure, but the mistaken notion that “rice-eaters” were particularly vulnerable to the plague due to the lack of protein in their diets served to justify a further confinement. As noted by Edelson ([19], p. 2874) “President McKinley ordered a quarantine of all Chinese and Japanese persons in San Francisco. Railroads and other means of public transportation were forbidden from carrying Asians and other members of what McKinley called ‘races liable to the plague’ out of the city unless they held health certificates from the Marine Hospital Service, the predecessor of the US Public Health Service.”

In Honolulu’s Chinatown, two cases of bubonic plague were reported in 1899. The epidemic “affected almost exclusively the Chinese community;” Craddock ([13], p. 131). By orders from the board of health, about 4500 “Chinese were removed to a quarantine camp, and the Chinese quarter was totally burned;” see Trauner ([52], p. 77). Medical officers used “fire to destroy one of the condemned sites in the Asian quarter at a moment when strong winds typical of that time of year arose. What resulted was the Great Fire of January 20th, 1900, making 7,000 Chinese and Japanese homeless, and depriving many of their enterprises and livelihood;” see Echenberg ([18], p. 444).23 Mohr [38] provides a detailed account of the policies used to combat the plague in Honolulu.

These ‘unreasonable’ and more radical responses relative to leprosy, for instance, were prompted by very few instances of plague. In San Francisco, only “six plague victims had been found in the ensuing search in Chinatown” and no Japanese victims “were to be reported for another year;” see Craddock ([13], p. 131). The ‘unreasonable’ responses were not only disproportionate but perhaps ineffective. Rats and their fleas spread bubonic

23 According to Echenberg ([18], p. 445), Honolulu’s plague outbreak played an important role in securing Hawaii’s privileged status as a United States territory, which eventually led to statehood.
plague so the confinement, relocation, and mobility restrictions applied to human populations experiencing the disease seem highly ineffective.\textsuperscript{24} Cleansing and disinfecting also seem impractical. The ‘sanitary fires’ first tested in Honolulu, for example, require that rats locate only in ‘unsanitary’ areas easy to identify.

**Smallpox and childhood diseases.** Several viral infections including smallpox (variola), measles, rubella (German measles), mumps, and varicella (chickenpox) are often grouped as *childhood diseases.* Childhood diseases are transmitted by airborne droplets and have very high transmission rates. The number of new cases started from a single case of measles, the basic reproduction number, is around $R_0 = 15$. Mumps ($R_0 = 13$), varicella ($R_0 = 8.5$), rubella ($R_0 = 6$), and smallpox ($R_0 = 6$) are also highly contagious. (For comparison, $R_0$ is 2.5 for the regular cold.) Childhood diseases have been separately identified since ancient times because their disease signs are highly visible and sufficiently distinctive.\textsuperscript{25} Disease signs include rashes, blisters, and fever; see, e.g., Oaten et al. ([42], Table 2).

Childhood diseases differ in their severity and visibility. While most are nowadays ‘harmless,’ smallpox was one of the deadliest infectious diseases in the ancient world. Its most virulent form (variola major) had a case fatality rate of 50 percent.\textsuperscript{26} Smallpox was also notorious for leaving survivors with permanent disfiguring scars or *pockmarks.* Pockmarks are likely due to virus-mediated necrosis and destruction of sebaceous glands in the skin; see Regan and Norton [46] for a recent medical review of the scarring mechanism of smallpox. Pockmarks were widespread in the body, but they were more frequently found on the face.

\textsuperscript{24} During the Black Death, for example, local quarantines served to spread the plague by redirecting infected ships across Mediterranean ports; see Ziegler [57]. Vaccination against the plague was also practiced in Honolulu and San Francisco, but with an experimental vaccine that proved problematic; see Craddock [13]; Echenberg [18]; Mohr [38]; and Trauner [52] for further discussions of the Third Pandemic.

\textsuperscript{25} “Although confused with smallpox well into the seventeenth century, measles was identified clearly as early as 910 AD by Rhazes,” the Baghdad physician; see Duffy ([17], pp. 17, 165). The lesions produced by smallpox and syphilis are also similar. “In the early seventeenth century, smallpox was occasionally confused with syphilis,” but gradually the two diseases were differentiated. The pockets (pox) left in the skin after the subsidence of variola were smaller than those of syphilis, also known early on as the great-pocks or pox.

\textsuperscript{26} “The most convincing evidence of smallpox in the ancient world are three Egyptian mummies from the Eighteenth and Twentieth Dynasties (1570 to 1085 BC);” Hopkins ([25], p.14). Ancient Sanskrit medical texts and Brahmin traditions, including “the existence of temples for worship of a deity of smallpox,” suggest that smallpox “was of considerable significance in Hindu India from time immemorial,” according to McNeill ([37], p. 128). China appears to be a *virgin soil* for smallpox (and measles) with its first introduction “from the north around 250 BC;” Hopkins ([25], p.18).
due to its highest concentration of sebaceous glands. Occasionally blindness resulted from corneal ulcerations. The smallpox virus (variola) cannot sustain viability in non-human hosts. For rubella, humans are also the only known host.

Smallpox is an acute infection that grants prolonged immunity to survivors. Therefore, smallpox “could only exist in a community so long as susceptible persons were available to keep the disease going,” as Hopkins ([25], p. 8) noted. (The variola virus does not persist in the body after recovery, as it is the case of the chickenpox virus whose re-activation triggers shingles.) Smallpox and measles arrived to Western Europe during a large-scale epidemic, the “Antonine Plague” of 165 to 180 AD; see McNeill ([37], p. 104), although their introduction may be as early as 430 BC, as the mysterious “Plague of Athens.” In Europe, smallpox and childhood disease epidemics were frequent during the Middle Ages; see Hopkins ([25], Chapter 2). Smallpox became endemic in large European cities by the sixteenth century.

Smallpox visited port cities in the American colonies in epidemic waves with long intervals between waves. In the colonies, as noted by Duffy ([17], p. 16), “[smallpox] was regarded with terror.” Smallpox proved especially devastating for the native populations in the New World and the Australian aborigines; see Kiple [28], McNeill [37], Hopkins [25], and Hutchinson [26]. Quarantines and the isolation of individual cases were the main forms of social control in early colonial times. In Massachusetts, a law was passed in 1730 “to prevent persons concealing the small pox and requiring a red cloth to be hung out in all infected places.” Connecticut enacted “[a]n Act to prevent the small-pox being spread in this colony by pedlers, hawkers, [and] petty chapmen.” Pennsylvania enacted “many laws specifically covering ‘infected’ immigrants” of German and Scotch-Irish descent, besides those “sick with contagious diseases;” see Duffy ([17], p. 102).

For centuries, societies in India, China, and Africa deliberately infected individuals with smallpox. Variolation was introduced into Europe during the eighteenth century and refined by Edward Jenner and many others; see Hopkins ([25], p. 12). Jenner’s success was not based on laboratory methods or an understanding of germs. Variolation and inoculation against smallpox steadily increased in the American colonies since their first use in Boston
in 1721; see Duffy ([17], Chapter 1). The WHO declared smallpox eradicated on May 8th, 1980.

An interpretation. – Smallpox, in Jenner’s words, was “the most dreadful scourge of the human species.” According to English historian Thomas Macaulay’s 17th century depiction, taken from Duffy ([17], p. 19):

“Smallpox was always present, filling the churchyard with corpses, tormenting with constant fear all whom it had not yet stricken, leaving on those whose lives it spared the hideous traces of its power, turning the babe into a changeling at which the mother shuddered, and making the eyes and cheeks of the betrothed maiden objects of horror to the lover.”

Smallpox had high fatality rates and very high infectivity. It was also visible as its disease signs were directly displayed in the face. The extent of its social control, however, does not appear to be as selective as that of leprosy or as widespread as that of plague. Social labeling took the harmless form of the “red treatment,” based on the belief that red colored objects could combat smallpox. (This practice dated to antiquity and remained in use well into the 1930s; see Hopkins [25].) Smallpox “interrupted normal life” in American cities (Duffy [17]), but the disease was not as exclusionary as previous ones. After all, variolation and inoculation purposefully increased exposure to smallpox. In the US and Britain, as noted by Cohn ([11], pp. 296, 306), “direct action against the victims of smallpox seldom occurred.” Instead, “smallpox’s anxiety, fear, and hatred” targeted those who “refused the cure [inoculation].” Later on, quarantines did target newly arrived immigrants to the US, but exclusion was apparently not different from the ‘normal’ association of foreigners with germs and contagion in American society. In response to smallpox epidemics in 1868, 1876, 1881, and 1887, “the citizens of San Francisco resorted to an unfortunate but not uncommon tactic: they ascribed the disease to the community of Chinese living within their midst.” During the first epidemic, however, there “is the almost total lack of anti-Chinese rhetoric in descriptions of the 1868 incident.” Only the last epidemic was met with selective quarantines.

27 George Washington contracted smallpox at age 19, while visiting Barbados. He carried pockmark scars in his face. In 1777, his mandatory program of inoculation was credited with virtually eradicating smallpox from the Continental Army; see Duffy ([17], Chapter 1) and Hutchinson ([26], p. 93).
and the “desire to expel the Chinese on grounds of disease and sanitation;” see Craddock ([12], pp. 957, 961, 963). These responses were not uncommon in American cities; see, e.g., Markel and Stern [35].

Smallpox’s central features are its visibility and transiency. Smallpox had a high signal-to-noise ratio $\alpha$ (Corollary 3) and it was a transient disease that granted long-term immunity to survivors, i.e., it had a high value of $\delta$ (Corollary 2) and a low value of $\eta$. Missing a detection was costly, but smallpox, like many other childhood diseases, had a short infectious cycle and a minor demographic impact in endemic areas since very young children are relatively easy to replace. In eighteenth century Europe, for example, 80 percent of the victims were children under 10; Kiple ([28], p. 1010). As a first approximation, it is possible to view childhood diseases as featuring $\eta = 0$ because their acquired immunity means that health is an absorbent state. As Propositions 3 and 4 show, the stationary disease prevalence is $q^* = 0$ and all potential gains from trade are exercised in equilibrium. The model interprets the high tolerance to smallpox, relative to another highly visible disease like leprosy, as a difference in the stochastic properties of these diseases. Smallpox was a dangerous and visible disease that should not foster social exclusion because of its transitory effects and its prolonged immunity.

**Times of cholera.** Cholera is an acute dehydrating disease typically spread through water contaminated with the bacterium *Vibrio cholerae*. Cholera is not typically spread by direct contact with an infected person, but it is highly contagious through the fecal-oral route. Its basic reproduction number is around $R_0 = 2.1$. As noted by Nelson et al. ([40], p. 694), *V. cholerae* colonizes the small intestine for “12 to 72 hours before symptoms appear,” leading to vomiting, diarrhea and a very rapid loss of fluids, “up to 1 litre per hour [5 gallons a day]. These losses result in severe fluid volume depletion and metabolic acidosis, which may lead to circulatory collapse and death.” Its case fatality in the past was high, as the 15 percent rate in Cairo’s first visit in 1831; see McNeill ([37], p. 231). Oral rehydration therapy reduced its fatality rate to less than 1 percent.

Cholera can kill infected individuals within hours as “one of the most ghastly experiences a disease could inflict on a human being;” see Watts ([53], p. 172). As noted by McNeill ([37], p. 231), cholera’s symptoms “were peculiarly horrible: radical dehydration mean that
a victim shrank into a wizened caricature of his former self within a few hours, while ruptured capillaries discolored the skin, turning it black and blue. The effect was to make mortality uniquely visible: patterns of bodily decay were exacerbated and accelerated, as in a time-lapse motion picture, to remind all who saw it of death's ugly horror and utter inevitability.” Perhaps even more dramatic than the rapid decay was the fact that “physical degradation did not cease with death. For an hour or so after the spirit of life was extinguished, the legs and arms of the body continued to thrash about, leading those hovering nearby to hope that the corpse was not really dead,” see Watts ([53], p. 173).

The ancestral reservoir of cholera is the Ganges and Brahmaputra deltas, from where it spread across the world in seven pandemic waves. The first pandemic began in 1817 but it did not afflict Europe. The second pandemic lasted from 1826 to 1837 and spread across Europe, North Africa, and the eastern seaboard of North America. The third pandemic, 1841-1859 also reached South and Central America. The most widely spread pandemic was the fourth wave, 1863-1875. Subsequent waves took place well into the twentieth century, with the last one (1961-1975) based on a less virulent strand. Cholera remains a global threat.

An interpretation.— Cholera targets young adults but “roughly half of all cases,” according Nelson et al. ([40], p. 695), are asymptomatic. It was even difficult to determine if an infected individual with cholera had died. The “violent muscular contractions at the time of death” produced a rise in body temperature (known as post-mortem caloricity) for “1 to 2 hours after death” and victims suffered an apparent death condition in which the vital signs “are maintained at a very low pitch [so] the person is not actually dead but appears

---

28 The pathogen _V. cholerae_ traveled well in water tanks “and up two weeks in the warm water within the hump of a camel,” which is how cholera first left India in the 1820s; see Watts ([53], p. 171). _V. cholerae_ is delicate, but the transport revolution reduced travel times between India and the rest of the world.

29 During this wave, John Snow argued against the miasmas theory of disease using the cholera outbreaks along the Broad Street’ pump in London in 1849, just a few yards away from Karl Marx’s residence. Snow’s analysis was pure empiricism. He did not identify the cause of cholera, “and everyone continued to believe that cholera was an airborne disease;” see Lippi and Gotuzzo ([32], p. 192). Had he known the mechanism behind cholera, he would have known that boiling water contaminated with cholera would make the water safe to drink. Robert Koch discovered the causative factor for cholera in Calcutta in 1884.

30 Besides cholera, many communicable diseases are asymptomatic including influenza, HIV, COVID, and tuberculosis, one of the main causes of death in the past. A familiar example of asymptomatic transmission is _typhoid Mary_ who, as a cook, spread the disease in New York City in the late 1880s. Typhoid fever is also a gastrointestinal disease associated with poor sanitary conditions, but typhoid fever (a bacterial infection) develops at a much slower pace than cholera.
apparently dead […]. A person may live in this state for a few seconds to ≥ 30 minutes;” see Chand ([9], pp. 111, 128). As Lippi and Gotuzzo ([32], p. 192) note, “many people died, often being buried when they were still alive, as collapse and apparent death were not uncommon in the algid phase.”

Cholera’s mail features are its high danger and its invisibility. A rapid progression and its asymptomatic cases imply a limited value for disease signs and critical difficulties to categorize healthy and diseased individuals. In the model’s context, its transiency (i.e., a high δ in Corollary 2) suggests high tolerance to disease signs. Its especially very low signal-to-noise ratio (i.e., a low α in Corollary 3), however, suggest a low social tolerance to cholera. For instance, disease prevalence beliefs for uninformative signals satisfy $\lim_{\alpha \to 0} P^*(d|x) \simeq \sqrt{\pi}/(\sqrt{\delta} + \sqrt{\pi})$, which have essentially the same predictive power as randomization, i.e., flipping a coin.

The sudden appearance and unpredictability of cholera had a unique demoralizing impact because traditional public health efforts could not contain the epidemic. As a water-borne disease, cholera “seemed capable of penetrating any quarantine, of bypassing any man-made obstacle.” It appeared to “chose its victims erratically” according to McNeill ([37] p. 231). Cholera followed a “crazy-quilt patterning, striking every third or fourth house along a street, skipping half a mile to hit another street, then alighting on villages usually seen as out of the way, remote, and in the old days of bubonic plague safe;” Watts ([53], p. 172). Cholera’s unpredictability meant that there were no noticeable signs or symptoms to identify it. Cholera was not even accepted as a new disease early on during the first epidemic waves. (European observers, including Karl Marx, associated the epidemic with the return of plague.)

During the first European tour of cholera, ‘unreasonable’ but not groundless fears were directed toward providers of medical advice. “The coincidence of inexplicable outbreaks of mass mortality with the sudden appearance on the scene of government officials, troops and medical officers readily aroused popular suspicious;” see Ranger and Slack ([45], p. 163). Panic and popular mistrust were widespread, leading to riots and social violence in virtually every country visited by cholera in the 1830s. Cohn ([11], Chapters 7-10) provides a detailed historical account of the cholera riots during the seven cholera pandemics. Even
though cholera is strongly associated with lack of hygiene and poor sanitary conditions, social violence targeted public health officers and physicians. As Cohn [11] notes, no other disease has ever awaken such violent antipathy to the medical profession as cholera.31

Social violence was less prevalent in the US, but later disease waves restored to traditional exclusionary policies. The late nineteenth century epidemics of cholera and typhus in New York City were met with detention, segregation, and district quarantines for immigrants (especially Russian Jews) arriving from Germany and other disease-infected European ports; see Markel [34] and Markel and Stern [35]. Cholera played a unique role in fostering medical science (i.e., oral rehydration therapy) and urban sanitation (i.e., clean-water and sewage-disposal systems) to ultimately reverse the urban mortality penalty. Colonial science still relied to its usual social labelling practices of naming a disease by its place of origin with names such as ‘Indian cholera,’ ‘Asiatic cholera,’ ‘Asiatic plague.’ Social labelling, as popularized by Gabriel Garcia Marquez, also associates cholera to the adoption of the *yellow flag* as the universal symbol of quarantine.

5 Conclusions

After touring the most spectacular epidemics that ravaged the West, it is perhaps useful to first summarize the main lessons one can draw about the exclusionary role of disease in the past. The etiology of disease was misunderstood, but social responses were guided by reduced-form contagion principles. Social control is typically imprecise because it is not possible, even today, to detect all instances of disease. In the statistical sense outlined here, false alarms were key in the social regulation of disease as a defensive strategy consistent with a “better safe than sorry” approach. Social responses also differed according to the intrinsic characteristics of disease. At the risk of oversimplification, leprosy had detailed rules for social regulation, whereas the social regulation of plague applied broader rules (i.e.,

31 The usual charges levied were poisoning the water (since acute arsenic poisoning shares the same clinical features as cholera; see Chand ([9], p. 484)), burying victims alive (since public health officers promoted a rapid burial for cholera victims), and snatching bodies (since cadavers began their use for educational purposes). Mistrust in the medical profession was present during the early twentieth century outbreaks in Italy and Russia, and the recent outbreaks in Peru and Venezuela during the 1990 and in Haiti after the natural disaster in 2010. Cholera shares some pathological features with Ebola. West Africa experienced similar riots against health workers during Ebola’s 2014-2016 outbreak; see, e.g., Cohn ([11], p. 261).
quarantines). In the proposed model, these differences can be organized along danger and visibility lines. Leprosy and plague are at opposite extremes of danger, because of differences in infectivity and fatality rates, and visibility, because of differences in the noisiness and persistence of disease signs and symptoms. The invisibility of plague and cholera implies a lower social tolerance to disease relative to leprosy. The proposed model also suggests that the deliberate exposure to smallpox and its higher social tolerance are a consequence of its transiency and long-term immunity.

Epidemic diseases have declined in their importance for morbidity and mortality in the world, but with a few exceptions (e.g., smallpox and polio) they have not yet been eradicated. While there is now a solid scientific base to prevent, control, and treat most communicable diseases, there is also a higher potential for global transmission because of improved transportation technology, and increased interconnections and interdependence. The causative agents of disease are constantly evolving at much faster rates than humans. In the past one hundred years alone, there have been countless emerging health threats including four influenza pandemics and several emerging and reemerging diseases leading to what Honigsbaum [24] labeled as the Pandemic Century. As this paper is completed (in the mid 2020), the world is in the middle of the COVID pandemic. It not possible to foresee the path that COVID and future pandemics will follow, but the lessons from the past few centuries should be helpful for understanding the fear and social anxiety triggered by disease.

References


6 Appendix A: Some Extensions

Stigma. The “disease” label is a social construction based on observations noted by others (signs), and not on self-perceptions (symptoms).\(^{32}\) Disease signs are a stigmatized attribute for individuals labeled or designated as “diseased,” but the label does not induce explicit individual or social reactions.\(^{33}\)

It is possible to allow for additional values from a social label. Suppose that a buyer’s utility is given by two components, as in \(u'(x) \equiv u(x) + \phi d\), where \(u(x)\) is the extrinsic utility associated with disease detection, and \(\phi > 0\) is an intrinsic utility gain associated with assigning a “diseased” label, regardless of disease detection. The detection threshold (3) is now

\[
x'_b \equiv L^{-1} \left( \pi(P) \left[ \frac{U(h, h) - U(d, h) - \phi d}{U(d, d) + \phi d - U(h, d)} \right] \right),
\]

with \(x'_b < x_b\) due to monotonicity (Assumption 1). The intrinsic valuation \(\phi\) would distort ex-post utilities and costs making traders less tolerant to disease signs. The term \(\phi\) is reduced-form, but there is a large sociological literature on social labeling and stigma; see Becker [6] and Goffman [22] for foundational works. An equilibrium analysis of stigma is beyond the scope of this paper but Akerlof [1], Bénabou and Tirole [5], and Bernheim [7] study stigma, identity, and conformity in social equilibria.

Contagion externalities. Meeting a diseased partner yields lower utilities and higher costs because of contagion and fear of contagion (Assumptions 2 and 3). Health transitions, however, are not influenced by an encounter with an ill trader.

I next embed the health dynamics into a Susceptible-Infected-Susceptible disease transmission framework. Consider a communicable disease whose transmission is external to the traders. The prevalence of disease in the population is \(Pr\{H = d\}\). The number of healthy

\(^{32}\) A symptom is a manifestation of disease apparent to the patient (i.e., a self-perception) whereas a sign is a manifestation of disease apparent to a practitioner. A self-perception problem would introduce a separate state variable for each trader. Concealment of disease signs would introduce private information. I leave these possibilities for future research.

\(^{33}\) The detection thresholds introduce kinks in utilities and costs, and differences in the marginal valuation of a potential trader as signals cross the thresholds, i.e., they serve as reference points. Koszegi and Rabin [29] build on prospect theory and examine reference dependent preferences in a personal equilibrium.
but Susceptible buyers is \( b \Pr\{H = h\} \). The number of Infected sellers is \( s \Pr\{H = d\} \). If a pairwise disease transmission rate is \( \beta/2 \), by symmetry, the total number of new infections in a single period is \( \beta sb \Pr\{H = h\} \Pr\{H = d\} \). The prevalence of disease evolves as

\[
\Pr\{H' = d\} = (1 - \delta') \Pr\{H = d\} + \eta \Pr\{H = h\},
\]

with the persistence in the disease state as \( \delta' \equiv \delta - \beta sb \Pr\{H = h\} \), which depends on the random mixing of traders in the population.

The equilibrium remains unchanged as contagion is external. The disease prevalence \( q^* \) is the positive solution to the quadratic equation \( \beta q^2 + (\delta + \eta - \beta)q - \eta = 0 \), from the stationary system. The stationary prevalence of disease with contagion exceeds the one in the baseline model. That is, \( q^* \geq \eta/(\delta + \eta) \) with equality under \( \beta = 0 \), which is the solution in the baseline model (Proposition 3).

It is also possible to refine the transmission rate further to allow for contagion externalities associated with successful pairwise exchanges, not just meetings. This would mean that transmission takes place only once traders carry out an actual exchange. I assumed disease signs only have informational value. Many of the pathological signs of disease also increase the probability of onward transmission. Examples include skin lesions, sneezing and coughing, and diarrhea and vomiting. Not all pathological signs, however, aid disease transmission. Rashes, paralysis, dementia, and congenital defects (among others) do not aid with disease transmission; see Weiss [55].

### 7 Appendix B: Omitted Derivations

This Appendix collects proofs associated with the acceptance threshold, its limiting values, and the comparative statics discussed in the text.

**Thresholds.** – In equilibrium, the detection thresholds satisfy

\[
P^*(d|x_b) = \frac{[U(h, h) - U(d, h)]}{[U(d, d) - U(h, d)] + [U(h, h) - U(d, h)]}, \tag{A1}
\]

and

\[
P^*(d|y_s) = \frac{[C(d, h) - C(h, h)]}{[C(h, d) - C(d, d)] + [C(d, h) - C(h, h)]}. \tag{A2}
\]

These thresholds vary with \( U(i, j) \) and \( C(i, j) \) in the way stated in Propositions 1 and 2.

The acceptance threshold can be written implicitly as

- **If** \( x < x_b \) and \( y(x) \geq y_s \),

\[
P^*(d|y(x)) = \frac{s[U(h, h) - (r + s)C(d, h)]}{(r + s)[C(d, d) - C(d, h)]} - P^*(d|x) \frac{s[U(h, h) - U(h, d)]}{(r + s)[C(d, d) - C(d, h)]}. \tag{A3}
\]
• If $x < x_b$ and $\bar{y}(x) < y_s$,
\[
P^*(d|\bar{y}(x)) = \frac{sU(h, h) - (r + s)C(h, h)}{(r + s)[C(h, d) - C(h, h)]} - P^*(d|x) \frac{s[U(h, h) - U(h, d)]}{(r + s)[C(h, d) - C(h, h)]}.
\]

(A4)

• If $x \geq x_b$ and $\bar{y}(x) \geq y_s$
\[
P^*(d|\bar{y}(x)) = \frac{sU(d, h) - (r + s)C(d, h)}{(r + s)[C(d, d) - C(d, h)]} - P^*(d|x) \frac{s[U(d, h) - U(d, d)]}{(r + s)[C(d, d) - C(d, h)]}.
\]

(A5)

• If $x \geq x_b$ and $\bar{y}(x) < y_s$,
\[
P^*(d|\bar{y}(x)) = \frac{sU(d, h) - (r + s)C(h, h)}{(r + s)[C(h, d) - C(h, h)]} - P^*(d|x) \frac{s[U(d, h) - U(d, d)]}{(r + s)[C(d, d) - C(h, d)]}.
\]

(A6)

It is possible to see that \{x_b, y_s, \bar{y}(x)\} are homogeneous of degree zero in $U(i, j)$ and $C(i, j)$, since a proportional change in utilities and costs leaves (A1), (A2), and (A3)-(A6) unchanged. To further characterize $\bar{y}(x)$, I consider a more explicit version of Assumption 5, which now lists all possible trading outcomes:

**Assumption 5.** Regardless of the assigned label, it is mutually beneficial for healthy types to trade, i.e.,
\[
sU(h, h) > sU(d, h) > (r + s)C(d, h) > (r + s)C(h, h),
\]

(A7)

and for diseased types not to trade, i.e.,
\[
(r + s)C(h, d) > (r + s)C(d, d) > sU(d, d) > sU(h, d).
\]

(A8)

Moreover,
\[
(r + s)C(d, d) > sU(h, h), \text{ and } (r + s)C(h, h) > sU(d, d).
\]

(A9)

The overall order of utilities and costs is just a combination of the previous expressions, and is given by $(r + s)C(h, d) > (r + s)C(d, d) > sU(h, h) > sU(d, h) > (r + s)C(d, h) > (r + s)C(h, h) > sU(d, d) > sU(h, d)$. Some expressions in this order are relevant in the subsequent derivations. With these expressions in mind, I next consider some of the omitted proofs and derivations.

**Proof of Proposition 4.** To obtain $\Pi^*(x, y) = 1$, one only needs to check $su(x) - (r + s)c(y) \geq 0$. Suppose, to the contrary, that acceptance is not determined by the seller, i.e., $su(x) - (r + s)c(y) \geq 0$ but $(r + b)u(x) - bc(y) < 0$. Then, $c(y) > (1 + r/b)u(x)$ and $u(x) \geq (1 + rs)c(y)$; a contradiction as long as $r > 0$; see (11). The threshold draws indifference from the seller’s perspective. It is possible to see that all the ‘intercept’ and ‘slope’ values in (A3)-(A6) are positive by Assumptions 2, 3, and 5. As $P^*(d|y)$ and $P^*(d|x)$ are strictly
increasing functions, the previous expressions imply that the acceptance threshold must be strictly decreasing in $x$.

Since the indirect utility and cost functions are continuous, the threshold is $\bar{y}(x)$ also continuous. If the likelihood function $L(x)$ is assumed differentiable, the inversion formula associated with the acceptance threshold fails to be differentiable only for the finite points associated with the detection threshold kinks. For all other points, the acceptance threshold satisfies $\bar{y}_x(x) = su_x(x)/(r + s)c_y(\bar{y}(x)) < 0$.

Expressions (A3)-(A6) also show that all the ‘intercept’ and ‘slope’ values in (A3)-(A6) are (weakly) increasing in $U(i, j)$ and decreasing in $C(i, j)$, so the threshold shifts outwards as $U(i, j)$ increases or $C(i, j)$ decreases.

**Healthiest-looking sellers.** To check that $\bar{y}(x_-) \geq y_s$, one requires

$$-\frac{sU(h, h) - (r + s)C(d, h)}{sU(h, h) - (r + s)C(d, d)} \geq \frac{C(d, h) - C(h, h)}{C(h, d) - C(d, d)}. \quad (A10)$$

The right-hand-side is positive (Assumption 3). The left-hand side is positive by (A7) and (A9). If differences are ‘large enough,’ the previous inequality is satisfied. Sufficient conditions for (A10) are $sU(h, h) - (r + s)C(d, h) \geq (r + s)C(d, d) - sU(h, h)$ and $C(h, d) - C(d, d) \geq C(d, h) - C(h, h)$, or simply

$$sU(h, h) \geq (r + s)\frac{C(h, h) + C(h, d)}{2} \geq (r + s)\frac{C(d, d) + C(d, h)}{2}. \quad (A11)$$

**Unhealthiest-looking sellers.** Similarly, to check that $\bar{y}(x_+) < y_s$, one requires

$$-\frac{sU(d, d) - (r + s)C(h, h)}{sU(d, d) - (r + s)C(h, d)} < \frac{C(d, h) - C(h, h)}{C(h, d) - C(d, d)}.$$  

The left-hand-side of the previous expression is negative by (A8) and (A9), so the inequality is verified.

**Proof of Corollary 1.** The comparative statics follow because the stationary detection thresholds behave qualitatively as in Propositions 1 and 2. For example, the threshold $x_b$ is increasing in $U(h, j)$, and decreasing in $U(d, j)$, for $j = h, d$; see (A1). (The comparative statics for $y_s$ are reversed; see (A2).) The acceptance threshold is implicitly defined through (A3)-(A6), as in

$$P^*(d|\bar{y}(x)) = n(x_b, y_s) - m(x_b, y_s)P^*(d|x),$$

where the ‘intercept’ $n(x_b, y_s)$ and ‘slope’ $m(x_b, y_s)$ depend on the relationship between $(x, y)$ and the detection thresholds $(x_b, y_s)$. By Proposition 4, the acceptance threshold $\bar{y}(x)$ is increasing in $U(i, j)$ for $i, j = h, d$. The limiting case with $\{x_b, y_s\} \rightarrow \{x_-, x_+\}$ satisfies (A5) which requires $\bar{y}(x) \rightarrow x_-$ for all $x$ in $X$. Essentially, since $P^*(d|x_-) = 0$, expression (A5) would imply that $P^*(d|\bar{y}(x)) = [U(d, d) - C(d, h)]/[C(d, d) - C(d, h)]$. The right-hand-side is negative by Assumption 5. Probabilities are non-negative so for $P^*(d|\bar{y}(x))$ to tend toward a negative value, one needs $\bar{y}(x) \rightarrow x_-$ and $P^*(d|\bar{y}(x)) \rightarrow 0$ as $x \rightarrow x_-$ for all $x$. ■
The overall comparative statics are collected in the following table:

<table>
<thead>
<tr>
<th></th>
<th>$U(h, h)$</th>
<th>$U(h, d)$</th>
<th>$U(d, h)$</th>
<th>$U(d, d)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_b$</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>$y_b(x)$</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>$y_s$</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>$y_s(x)$</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
</tbody>
</table>

Although not discussed in the text, additional comparative statics are also intuitive. For example, suppose that the utility (cost) of correct disease rejection declines (increases). The detection and acceptance thresholds decline. Society also becomes less tolerant to disease signs because the gains from trade are reduced (Assumption 5). Correct rejections of disease are relevant as they signal valuable trading opportunities. If the value of these opportunities decreases, the acceptance set shrinks and a “clean” label loses some of its value. Finally, suppose that the utility (cost) of false alarms or correct disease detection declines (increases). The detection thresholds increase and the acceptance threshold declines. In both cases, the cost of assigning a “disease” label increases hence society becomes more tolerant to disease signs. Buyers and sellers have a lower incentive to over-diagnose (i.e., false alarms) or correctly detect a diseased partner.

**Proof of Corollary 2.** The (approximated) disease prevalence beliefs (8) yield stationary detection thresholds

\[
x_b = \mathcal{L}^{-1}\left(\frac{\delta}{\eta} \left[\frac{U(h, h) - U(d, h)}{U(d, d) - U(h, d)}\right]^2\right)
\]

and

\[
y_s = \mathcal{L}^{-1}\left(\frac{\delta}{\eta} \left[\frac{C(d, h) - C(h, h)}{C(h, d) - C(d, d)}\right]^2\right),
\]

so $x_b$ moves with $\delta$ since $\mathcal{L}(x)$ is increasing in $x$. As $\delta \to 0$, then $x_b \to x_-$, as stated in the text.

At low values of $\delta$, the acceptance threshold satisfies (A5) since all values of $x$ and $y$ exceed the thresholds. (The arguments apply to all other segments relevant for $P^*(d|\bar{y}(x))$.) Rearranging terms using (8) gives

\[
\sqrt{\mathcal{L}(\bar{y}(x))} = \frac{\delta [sU(d, h) - (r + s)C(d, h)] + \sqrt{\delta \eta} \sqrt{\mathcal{L}(\bar{y}(x))} [sU(d, d) - (r + s)C(d, h)]}{\sqrt{\delta \eta} (r + s) [C(d, d) - sU(d, h)] + \eta \sqrt{\mathcal{L}(\bar{y}(x))} (r + s) [C(d, d) - sU(d, d)]}
\]

This expression shows that $\bar{y}(x)$ is increasing in $\delta$. As $\delta \to 0$, the right-hand-side goes to zero so $\bar{y}(x) \to x_-$ for all $x$ in $X$, following the same arguments at the end of the previous proof. Although not part of the corollary, as $\eta \to 0$, the right-hand-side goes to infinity so $\bar{y}(x) \to x_+$ for all $x$, making all potential exchanges mutually beneficial.

**Log-normal signals.**— Under log-normal and opposite signals, the (logistic) approximation
for the stationary beliefs (3) is
\[ P^*(d|x) \approx \frac{x^\alpha}{\sqrt{\delta/\eta + x^\alpha}} = 1 - \frac{x^{-\alpha}}{\sqrt{\eta/\delta + x^{-\alpha}}}. \quad (A13) \]

The detection thresholds are
\[ x_b = \left( \sqrt{\frac{\delta}{\eta} U(h, h) - U(d, h)} \right)^{1/\alpha} \text{ and } y_s = \left( \sqrt{\frac{\delta}{\eta} C(d, h) - C(h, h)} \right)^{1/\alpha}. \quad (A14) \]

Beliefs can be (further) approximated by Pareto-type distributions. For small values of \( x \) and \( \delta \simeq \eta \), beliefs can be approximated by \( x^\alpha \), while for large values of \( x \), I consider \( 1 - x^{-\alpha} \). Using these approximations at (A3)-(A6) and some calculations yields:

If \( x < x_b \) and \( \bar{y}(x) \geq y_s \),
\[ \bar{y}(x) = \left( \frac{(r + s)C(d, d) - sU(h, h)}{(r + s)[C(d, d) - C(h, h)]} - \frac{1}{x^{-\alpha}} \frac{s[U(h, h) - U(h, d)]}{(r + s)[C(d, d) - C(h, h)]} \right)^{-1/\alpha}, \quad (A15) \]

if \( x < x_b \) and \( \bar{y}(x) < y_s \),
\[ \bar{y}(x) = \left( \frac{sU(h, h) - (r + s)C(h, h)}{(r + s)[C(h, h) - C(h, h)]} - x^{-\alpha} \frac{s[U(h, h) - U(h, d)]}{(r + s)[C(h, d) - C(h, h)]} \right)^{1/\alpha}, \quad (A16) \]

if \( x \geq x_b \) and \( \bar{y}(x) \geq y_s \),
\[ \bar{y}(x) = \left( \frac{(r + s)C(d, d) - sU(h, h)}{(r + s)[C(d, d) - C(h, h)]} - x^{-\alpha} \frac{s[U(h, d) - U(h, d)]}{(r + s)[C(d, d) - C(h, h)]} \right)^{-1/\alpha}, \quad (A17) \]

if \( x \geq x_b \) and \( \bar{y}(x) < y_s \),
\[ \bar{y}(x) = \left( \frac{sU(d, d) - (r + s)C(h, h)}{(r + s)[C(h, h) - C(h, h)]} + \frac{1}{x^{\alpha}} \frac{s[U(d, h) - U(d, d)]}{(r + s)[C(h, d) - C(h, h)]} \right)^{1/\alpha}. \quad (A18) \]

**Proof of Corollary 3.** The buyer’s detection thresholds uses (A12) or (A14), with
\[ \frac{dx_b}{d\alpha} = -\ln \left( \frac{\delta}{\eta} \frac{U(h, h) - U(h, d)}{U(d, d) - U(h, d)} \right) \frac{x_b}{\alpha^2}. \]

The seller’s case is symmetric and hence I omit it. The threshold \( x_b \) is increasing in \( \alpha \) as long as right-hand-side is smaller than one, i.e., \( \sqrt{\delta} [U(h, h) - U(h, d)] < \sqrt{\eta} [U(d, d) - U(h, d)] \). This condition states that the disease state is more persistent than the healthy state, and that the marginal gains from avoiding detection errors are larger in the disease state. This condition implies that \( x_b \) itself is smaller than one. I assume that this and a symmetric condition for the costs holds, so that both thresholds decline toward \( x_\alpha = 0 \) as \( \alpha \) decreases.

Given that \( \{x_s, y_b\} \to \{x_\alpha, x_\alpha\} \) as \( \alpha \to 0 \), one only needs to consider (A17) to locally
characterize the acceptance threshold. (This is the relevant expression when the thresholds become small.) A log-linear approximation yields

\[
\ln \bar{\psi}(\psi) \approx -\frac{1}{\alpha} \left( \ln \frac{(r + s)C(d, d) - sU(d, d)}{(r + s)[C(d, d) - C(d, h)]} + \alpha \ln x - \ln \frac{s[U(d, h) - U(d, d)]}{(r + s)[C(d, d) - C(d, h)]} \right),
\]

which can be written as

\[
\bar{\psi}(x) \approx \frac{1}{x} \left( \frac{s[U(d, h) - U(d, d)]}{(r + s)[C(d, d) - sU(d, d)]} \right)^{1/\alpha}.
\]

Thus, an increase \( \alpha \) shifts outwards \( \bar{\psi}(x) \) if the term inside the parenthesis is smaller than one, i.e., \( sU(d, h) < (r + s)C(d, d) \), which holds by Assumption 5. \( \blacksquare \)