How does a Rational Agent Act in an Epidemic?

S. Yagiz Olmez, Shubham Aggarwal, Jin Won Kim, Erik Miehling,

Tamer Başar, Matthew West, and Prashant G. Mehta

Coordinated Science Lab, University of Illinois at Urbana-Champaign, Urbana, IL

{solmez2,sa57,jkim684,miehling,basar1,mwest,mehtapg}@illinois.edu

Abstract— **Evolution of disease in a large population is a function of the top-down policy measures from a centralized planner, as well as the self-interested decisions (to be socially active) of individual agents in a large heterogeneous population. This paper is concerned with understanding the latter based on a mean-field type optimal control model. Specifically, the model is used to investigate the role of partial information on an agent's decision-making, and study the impact of such decisions by a large number of agents on the spread of the virus in the population. The motivation comes from the presymptomatic and asymptomatic spread of the COVID-19 virus where an agent unwittingly spreads the virus. We show that even in a setting with fully rational agents, limited information on the viral state can result in an epidemic growth.**

I. INTRODUCTION

Social behavior of individual agents ultimately drives the case count in an epidemic. Decisions of an individual agent are mediated by the top-down policy guidelines, e.g., lockdown mandates or social distancing guidelines. However, these decisions are also a function of: (i) the agent's risk-reward trade-off based on the agent's health condition, age, and information from media on transmissibility and lethality of the virus (e.g., omicron variant is more transmissible but less deadly than the delta variant), (ii) aggregate positivity rate in the local population, and (iii) the agent's assessment of its own epidemiological status.

This paper is a continuation of prior work from our group on modeling an agent's decision-making in an epidemic based on mean-field game (MFG) formalisms [1]. The primary focus of the current paper is to investigate the role of agent rationality and information on agent behavior. The motivation comes from the following thought experiment: On the planet of Vulcan, the agents are both perfectly rational and perfectly informed. Because an agent has perfect information, it immediately knows its epidemiological status. And because an agent is perfectly rational, it knows to self-quarantine if infected. Therefore, the rational choice of a susceptible (non-infected) agent is to carry on with normal

life. This is because it can count on rational choices of perfectly informed others.

Unfortunately in reality, on the planet of Earth, we continue to experience waves of infection which are doubtlessly caused by emergence of variants but surely exacerbated by our own actions which are often less than rational and certainly almost always based on incomplete information. A case in point is the role played by the presymptomatic and asymptomatic spread of the COVID-19 virus [2]–[5].

The contributions of this paper are as follows. Apart from the development of the model, we derive several qualitative insights into an agent's actions in the midst of an epidemic. Specifically, these insights are related to: (i) the risk-reward tradeoff of a susceptible agent, and (ii) the evolution of belief of an active agent who does not show symptoms. The resulting decision-making and its effect on the disease outbreak in the population are discussed at length. These qualitative insights are drawn from quantitative results presented as part of Props. 3, 4, and 5, and as illustrated in Figs. 3-7. The discussion appears in Sec. IV-D.

Closely related to our work are [6], where an agent's decision variable is its rate of contact with others, and [7], [8], where an agent strives to follow a prescribed rate of contact based on government guidelines. Other MFG-style modeling of epidemics appears in [9]–[13]. The novelty of our work comes from the inherent partial observability of viral status and differences in cost structures. These factors are crucial for modeling the asymptomatic spread of an epidemic.

The remainder of the paper is organized as follows. The problem formulation appears in Sec. II where the two main questions of the paper are also introduced. The answers to these questions are based on the analysis of the HJB equations derived in Sec. III. The analysis is presented in Sec. IV together with qualitative answers to the two main questions. These answers conclude this paper and spur the development of the MFG formalism. This formalism, which will be studied in detail as part of future work, is presented in Sec. V. The proofs appear in the Appendix.

II. PROBLEM FORMULATION

A. Model for a Single Agent

Dynamics: The epidemiological state of a single agent is modeled as a Markov process $X := \{X_t \in \mathcal{X} : t \geq 0\}$
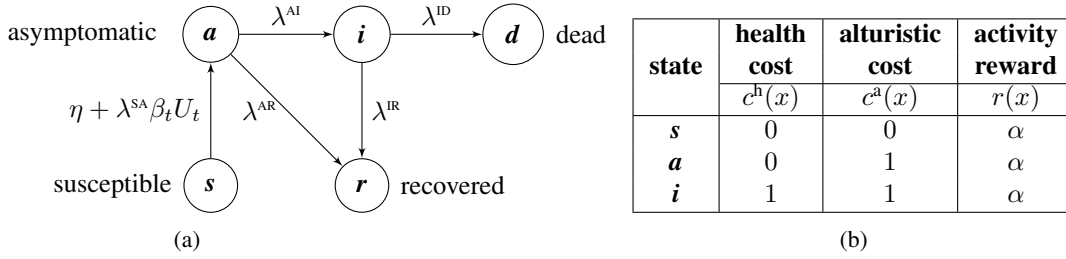
Fig. 1: (a) Epidemiological states and the associated transition graph. (b) Model for running cost: It is assumed that the economic reward of α is smaller than the altruistic cost of 1 for an asymptomatic ($x = \boldsymbol{a}$) or symptomatic ($x = \boldsymbol{i}$) agent.

with state-space $\mathcal{X} := \{\boldsymbol{s}, \boldsymbol{a}, \boldsymbol{i}, \boldsymbol{r}, \boldsymbol{d}\}$. Fig. 1 depicts the transition graph along with a description of the epidemiological meaning of each of the five states. Notably, there are two types of infected states: (i) asymptomatic state, denoted as $\boldsymbol{a}$; and (ii) symptomatic state, denoted as $\boldsymbol{i}$. In either of these states, an agent is infectious, i.e., able to infect other agents. The modeling distinction is that, in partially observed settings, an asymptomatic agent may not *know* its true state but a symptomatic agent does. Note that unlike our previous paper [1], the transition graph now includes an edge from $\boldsymbol{a}$ to $\boldsymbol{r}$. This is more realistic, but significantly complicates modeling and analysis. Broadly, there are two types of transitions:

1) On the subset $\{\boldsymbol{a}, \boldsymbol{i}, \boldsymbol{r}, \boldsymbol{d}\}$, the transition rate depends *only* upon the agent attribute θ, which here represents the age of the agent. For example, an older symptomatic agent has a longer expected recovery time (smaller λ^{IR}) than a younger one.

2) The transition from $\boldsymbol{s} \to \boldsymbol{a}$ depends upon three factors: (i) the intrinsic infectivity of the virus, (ii) the agent behavior (level of social activity), and (iii) the behavior of the infected agents in the population. The following equation is used to model the effect of these three factors:

$$\text{rate}[\boldsymbol{s} \to \boldsymbol{a}] = \eta + \lambda^{\text{SA}} \, \beta_t \, U_t$$

where $\eta > 0$ is a small baseline rate (useful for regularizing the problem) and λ^{SA} is the virus transmissibility parameter. The process U is referred to as the agent's control input, and models the agent's activity level: With $U_t = 0$ (resp., $U_t = 1$) the agent is isolated (resp., normally active) at time t (Fig. 2). The process β models the average activity level of the infected agents (Eq. (1)).

The two main questions driving this work are as follows: (i) How does a single agent choose its control input U? and (ii) How does that choice (made by agents) affect the evolution of disease in a large heterogeneous population? To answer these questions, we adopt an optimal control framework.

Optimal control objective: In the following, β is a given deterministic process. The control objective for a single agent is to choose its activity U to minimize

$$\mathsf{J}(U; \beta) = \mathsf{E} \left(\int_0^T e^{-\gamma t} c(X_t, U_t) \, \mathrm{d}t + e^{-\gamma T} \phi(X_T) \right)$$
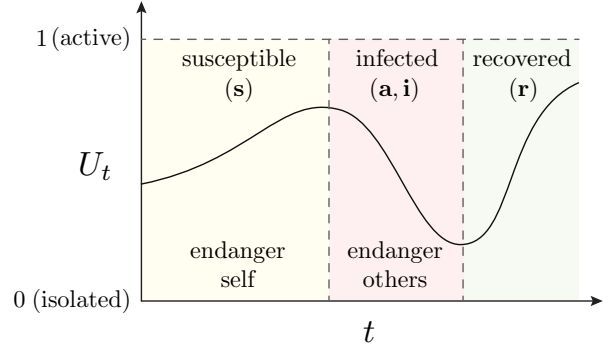

Fig. 2: A qualitative representation of the control input. Agents choose their activity levels based on their own viral status and the disease level within the population.

where $\gamma > 0$ is a discounting rate (assumed to be much smaller than the transition rates), $T = T(\omega) = \inf\{t > 0 : X_t(\omega) \in \{\boldsymbol{d}, \boldsymbol{r}\}\}$ is the random stopping time when the agent either recovers ($X_T = \boldsymbol{r}$) or the agent dies ($X_T = \boldsymbol{d}$); by convention, $\inf \emptyset = \infty$. The cost function is of the following form:

$$c(x, u) = c^{\text{h}}(x) + (c^{\text{a}}(x) - r(x)) \, u$$

where $c^{\text{h}}(\cdot)$, $c^{\text{a}}(\cdot)$, and $r(\cdot)$ model the health cost, altruistic cost, and economic reward, respectively (Fig. 1 (b)). The altruistic cost models an agent's desire to help the greater good and has the effect of inducing an infected agent to isolate. Of course, an agent's choice is mediated by its information, the structure of which is described next.

Information structure: There are two settings of the problem: (i) the fully observed case, and (ii) the partially observed case. In the partially observed setting, the observation process $Y := \{Y_t \in \{0, 1, 2\} : t \geq 0\}$ is defined as

$$Y_t = 1_{[X_t = \boldsymbol{i}]} + 2 \cdot 1_{[X_t = \boldsymbol{d}]}$$

Therefore, an infected agent who is also asymptomatic (in state $\boldsymbol{a}$) does not observe its epidemiological status until it begins to show symptoms (in state $\boldsymbol{i}$).

B. Model for the Mean-Field

To fully specify the problem, we need to define the model for the agent's attribute θ and the process β, henceforth

referred to as the mean-field process. The probability mass function of the attribute θ is denoted by $\mathsf{p}(\cdot)$. Denote by $\rho_t(x, u; \theta)$ the joint distribution of the state-action pair (X_t, U_t) at time t, conditioned on the attribute θ. At time t, the average activity level of infected agents is

$$\beta_t := \sum_\theta \mathsf{p}(\theta) \sum_{x \in \{\boldsymbol{a}, \boldsymbol{i}\}} \int_0^1 u \rho_t(x, u; \theta)\, \mathrm{d}u \qquad (1)$$

which we define to be the mean-field process.

C. Function Spaces

The filtration of the Markov process X is denoted by $\mathcal{F} := \{\mathcal{F}_t : t \geq 0\}$ where $\mathcal{F}_t := \sigma(X_t)$. The filtration of the observation process Y is denoted by $\mathcal{Y} := \{\mathcal{Y}_t : t \geq 0\}$ where $\mathcal{Y}_t := \sigma(\{Y_s : 0 \leq s \leq t\})$. In the two settings of the problem, the spaces of admissible control inputs, denoted by $\mathcal{U}$ in each case, are as follows:

$$\begin{aligned} \text{(fully obsvd.)} \quad & \mathcal{U} = L_\mathcal{F}^2([0, \infty); [0, 1]) \\ \text{(part. obsvd.)} \quad & \mathcal{U} = L_\mathcal{Y}^2([0, \infty); [0, 1]) \end{aligned}$$

i.e., an admissible control input U is a $[0, 1]$-valued stochastic process adapted to $\mathcal{F}$ in the fully observed case, and adapted to $\mathcal{Y}$ in the partially observed case. The use of the common notation $\mathcal{U}$ should not cause any confusion because the two cases are treated in separate subsections. The process β is assumed to be deterministic. The function space for β is $\mathcal{M} = L^2([0, \infty); [0, 1])$.

III. OPTIMALITY EQUATIONS FOR A SINGLE AGENT

A. Fully Observed Case

For each $x \in \mathcal{X}$ and $t \geq 0$, the value function is

$$v_t(x) := \min_{U \in L_\mathcal{F}^2}$$

$$\mathsf{E}\left(\int_t^T e^{-\gamma(s-t)} c(X_s, U_s)\, \mathrm{d}s + e^{-\gamma T} \phi(X_T) \,\middle|\, X_t = x \right) \tag{2}$$

where at the two terminal states $x = \boldsymbol{d}$ and $x = \boldsymbol{r}$, the value function is given by known constants $v_t(\boldsymbol{d}) = \bar{\phi}(\boldsymbol{d})$ and $v_t(\boldsymbol{r}) = \bar{\phi}(\boldsymbol{r})$.

For the fully observed problem, complete characterizations of the value function and the optimal control are described in the following proposition. This result is a minor extension of a similar result appearing in our prior paper [1] (for λ^{AR} and $\eta = 0$) and therefore its proof is omitted.

Proposition 1: Suppose $\alpha < 1$. For $x \in \{\boldsymbol{a}, \boldsymbol{i}, \boldsymbol{r}, \boldsymbol{d}\}$, the value function $v_t(x)$ and the optimal control law $\psi_t(x)$ are stationary as tabulated in Table I. For state $\boldsymbol{s}$, the value function $v_t(\boldsymbol{s})$ solves the HJB equation

$$-\frac{\mathrm{d}v_t}{\mathrm{d}t}(\boldsymbol{s}) + (\gamma + \eta) v_t(\boldsymbol{s})$$
$$= \eta \bar{\phi}(\boldsymbol{a}) + \min_{u \in [0,1]} \left(\lambda^{\text{SA}} \beta_t \left(\bar{\phi}(\boldsymbol{a}) - v_t(\boldsymbol{s}) \right) - \alpha \right) u$$

Remark 1: The optimal control for both asymptomatic and symptomatic agents is zero. This is entirely because

state x	value function $v_t(x)$	optimal control
$\boldsymbol{a}$	$\frac{c^h(\boldsymbol{a}) + \lambda^{\text{AI}} \bar{\phi}(\boldsymbol{i}) + \lambda^{\text{AR}} \bar{\phi}(\boldsymbol{r})}{\gamma + \lambda^{\text{AI}} + \lambda^{\text{AR}}} =: \bar{\phi}(\boldsymbol{a})$	$U_t^{\text{opt}} = \psi_t(\boldsymbol{a}) = 0$
$\boldsymbol{i}$	$\frac{c^h(\boldsymbol{i}) + \lambda^{\text{IR}} \bar{\phi}(\boldsymbol{r}) + \lambda^{\text{ID}} \bar{\phi}(\boldsymbol{d})}{\gamma + \lambda^{\text{IR}} + \lambda^{\text{ID}}} =: \bar{\phi}(\boldsymbol{i})$	$U_t^{\text{opt}} = \psi_t(\boldsymbol{i}) = 0$
$\boldsymbol{r}$	$\bar{\phi}(\boldsymbol{r})$ (given)	-
$\boldsymbol{d}$	$\bar{\phi}(\boldsymbol{d})$ (given)	-

TABLE I: Fully observed case.

$\alpha < 1$. Recall that α is a model for economic reward per unit time. The assumption $\alpha < 1$ means that the economic reward (α) is outweighed by the altruistic cost (1).

B. Partially Observed Case

The partially observed problem is first converted to a fully observed one by introducing the belief state, as in [14], [15], which at time t is denoted by

$$\pi_t := \begin{bmatrix} \pi_t(\boldsymbol{s}) & \pi_t(\boldsymbol{a}) & \pi_t(\boldsymbol{i}) & \pi_t(\boldsymbol{r}) & \pi_t(\boldsymbol{d}) \end{bmatrix}$$

where $\pi_t(x) := \mathsf{P}([X_t = x] \mid \mathcal{Y}_t)$ for $x \in \mathcal{X}$. Since the events $[X_t = \boldsymbol{i}]$ and $[X_t = \boldsymbol{d}]$ are both contained in $\mathcal{Y}_t$, π_t is not an arbitrary element of the probability simplex in $\mathbb{R}^5$. Let $\mathcal{P}^1$ denote the set of pmfs on $\{\boldsymbol{s}, \boldsymbol{a}, \boldsymbol{r}\}$ and let $\mathcal{P}^2 = \{\delta_{\boldsymbol{i}}, \delta_{\boldsymbol{r}}, \delta_{\boldsymbol{d}}\}$. Then, the state-space for the belief is $\mathcal{P}^1 \cup \mathcal{P}^2$. For $t \geq 0$ and $\mu \in \mathcal{P}^1 \cup \mathcal{P}^2$, the value function is

$$v_t(\mu) := \min_{U \in L_\mathcal{Y}^2}$$

$$\mathsf{E}\left(\int_t^T e^{-\gamma(s-t)} c(X_s, U_s; \alpha)\, \mathrm{d}s + e^{-\gamma T} \phi(X_T) \mid \pi_t = \mu \right)$$

There are two cases to consider: (i) when $\mu \in \mathcal{P}^2$, and (ii) when $\mu \in \mathcal{P}^1 \setminus \{\delta_{\boldsymbol{r}}\}$. In the first case, when $\mu \in \mathcal{P}^2$, the problem reduces to the fully-observed setting, and the value function is given by

$$v_t(\delta_{\boldsymbol{d}}) = \bar{\phi}(\boldsymbol{d}), \quad v_t(\delta_{\boldsymbol{r}}) = \bar{\phi}(\boldsymbol{r}), \quad v_t(\delta_{\boldsymbol{i}}) = \bar{\phi}(\boldsymbol{i})$$

The optimal control for the agent in the symptomatic state ($\pi_t = \delta_{\boldsymbol{i}}$) is $U_t^{\text{opt}} = \psi_t(\delta_{\boldsymbol{i}}) = 0$.

For the second case, when $\mu \in \mathcal{P}^1 \setminus \{\delta_{\boldsymbol{r}}\}$, a nonlinear filter is used to obtain the evolution of the belief. For this purpose, consider first the random variable $\tau = \tau(\omega) = \inf\{t > 0 : X_t(\omega) = \boldsymbol{i}\}$. Now, τ is a $\mathcal{Y}_t$-stopping time and

$$\pi_t = \begin{bmatrix} \pi_t(\boldsymbol{s}) & \pi_t(\boldsymbol{a}) & 0 & \pi_t(\boldsymbol{r}) & 0 \end{bmatrix} \quad \text{for } t < \tau$$

Let $S_t := \pi_t(\boldsymbol{s})$, $A_t := \pi_t(\boldsymbol{a})$ and $R_t := \pi_t(\boldsymbol{r})$ for $t < \tau$. Then the stochastic process $\{(S_t, A_t, R_t) \in [0, 1]^3 : S_t + A_t + R_t = 1, 0 \leq t < \tau\}$ evolves according to the nonlinear filter

$$\frac{\mathrm{d}S_t}{\mathrm{d}t} = (-\lambda^{\text{SA}} \beta_t U_t - \eta + A_t \lambda^{\text{AI}}) S_t \tag{3a}$$

$$\frac{\mathrm{d}A_t}{\mathrm{d}t} = (\lambda^{\text{SA}} \beta_t U_t + \eta) S_t + A_t (-\lambda^{\text{AI}} - \lambda^{\text{AR}} + \lambda^{\text{AI}} A_t) \tag{3b}$$

$$\frac{\mathrm{d}R_t}{\mathrm{d}t} = A_t (\lambda^{\text{AR}} + \lambda^{\text{AI}} R_t) \tag{3c}$$

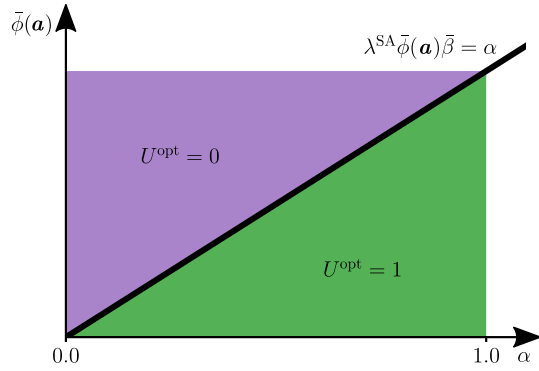which is derived from the general form detailed by [16].

Fig. 3: A susceptible agent's risk-reward diagram. The line $\lambda^{\text{SA}}\bar{\phi}(\boldsymbol{a})\bar{\beta} = \alpha$ is where the risk equals the reward.

We identify $\mathcal{P}^1$ with the domain $D := \{(s,a) \in [0,1]^2 : s + a \leq 1\}$ where s is the value of S_t and a is the value of A_t. For an arbitrary element $\mu = [s, a, 0, 1 - s - a, 0]$ in $\mathcal{P}^1$, we denote the value function with respect to the (s,a)-coordinates as

$$\phi_t(s,a) := v_t(\mu), \quad (s,a) \in D, \quad t \geq 0$$

The following proposition provides the HJB equation whose derivation appears in Appendix C.

Proposition 2: The value function $\phi_t(s,a)$ solves the HJB equation

$$\left(-\frac{\partial \phi_t}{\partial t} + \gamma \phi_t\right)(s,a) = (\mathcal{L}\phi_t)(s,a) + \min_{u \in [0,1]} \mathcal{M}_t(s,a)u \tag{4}$$

together with boundary conditions

$$\phi_t(s,a) - \phi_t(s, 1-s-a) = (2a+s-1)\left(\bar{\phi}(\boldsymbol{a}) - \bar{\phi}(\boldsymbol{r})\right) \tag{5a}$$

$$\phi_t(0,a) = a\bar{\phi}(\boldsymbol{a}) + (1-a)\bar{\phi}(\boldsymbol{r}) \tag{5b}$$

where the formulae for the linear operators $\mathcal{L}$ and $\mathcal{M}_t$ appear in Appendix C.

In the remainder of this paper, it is assumed that a unique solution for the HJB equation exists and yields a well-posed optimal control law, denoted as $\psi_t(s,a)$ for $(s,a) \in D$.

IV. ANALYSIS OF THE DECISION-MAKING BY AN AGENT

In this section, an analysis of the HJB and the nonlinear filter equations is given. The analysis is used to develop insights into the decisions of a single agent in an epidemic (given β). To aid the analysis, it is assumed that $\beta_t = \bar{\beta}$ for all $t \geq 0$ (i.e. stationary). The main results are described in the form of three propositions (Props. 3-5) and illustrated with accompanying figures (Fig. 3- 7). The section concludes with a qualitative discussion of an agent's decision-making in Sec. IV-D.

A. Risk-Reward Tradeoff for a Fully Observed Susceptible Agent

Proposition 3 (Stationary Solution): Suppose $\alpha < 1$, $\gamma > 0$, $\bar{\phi}(\boldsymbol{i}) > 0$, $\beta_t \equiv \bar{\beta}$. Set $\bar{\beta}^{\text{crit}} := \frac{\alpha}{\lambda^{\text{SA}}\bar{\phi}(\boldsymbol{a})}\left(1 + \frac{\eta}{\gamma}\right)$. Then the optimal control for a susceptible agent is stationary and described by the following cases:
1) If $\bar{\beta} < \bar{\beta}^{\text{crit}}$, then the optimal control $U_t^{\text{opt}} = \psi_t(\boldsymbol{s}) = 1$ and the optimal value $v_t(\boldsymbol{s}) = \frac{(\lambda^{\text{SA}}\bar{\beta}+\eta)\bar{\phi}(\boldsymbol{a})-\alpha}{\lambda^{\text{SA}}\bar{\beta}+\gamma+\eta}$.
2) If $\bar{\beta} \geq \bar{\beta}^{\text{crit}}$, then the optimal control $U_t^{\text{opt}} = \psi_t(\boldsymbol{s}) = 0$ and the optimal value $v_t(\boldsymbol{s}) = \frac{\eta}{\gamma+\eta}\bar{\phi}(\boldsymbol{a})$.

Remark 2: A susceptible agent's decision is best understood using a risk-reward diagram depicted in Fig. 3. The horizontal axis of this diagram is the economic reward per unit time (α). The vertical axis is the potential health cost (value function $\bar{\phi}(\boldsymbol{a})$). The diagram helps show that the optimal decision for an agent is to be active if the reward is greater than the risk. For any given α and $\bar{\phi}(\boldsymbol{a})$, the risk-reward analysis reveals a critical threshold $\bar{\beta}^{\text{crit}}$ above which the agent ceases to be active. It is noted that the critical threshold scales inversely with the product $\lambda^{\text{SA}}\bar{\phi}(\boldsymbol{a})$. This is useful in several ways for analysis:
1) An older agent will have a greater potential health cost $\bar{\phi}(\boldsymbol{a})$ and therefore a smaller critical threshold $\bar{\beta}^{\text{crit}}$.
2) A more contagious (high λ^{SA}) variant may still have a larger critical threshold, if it is significantly less severe (low $\bar{\phi}(\boldsymbol{a})$). For instance, the omicron variant is two times more transmissible ($2\times$ greater λ^{SA}) than the delta variant [17]. However, the potential health cost $\bar{\phi}(\boldsymbol{a})$ of the omicron is also smaller (because it is less lethal).

B. Threshold Optimal Policies for Partially Observed Case

In our prior paper [1], we derived the following result for the solutions of the HJB equation in a special case:

Proposition 4: Suppose $\beta_t \equiv \bar{\beta}$, $S_0 + A_0 = 1$, and $\lambda^{\text{AR}} = 0$. Then, for $t < \tau$, $S_t + A_t = 1$, and
1) If $\bar{\beta} \geq \bar{\beta}^{\text{crit}}$, then the optimal control law $\bar{\psi}(\cdot, a) = 0$.
2) For each fixed $\bar{\beta} < \frac{\alpha}{\lambda^{\text{SA}}\bar{\phi}(\boldsymbol{i})} < \bar{\beta}^{\text{crit}}$ there exists a $\underline{\lambda}^{\text{AI}} = \underline{\lambda}^{\text{AI}}(\bar{\beta})$ such that for all $\lambda^{\text{AI}} > \underline{\lambda}^{\text{AI}}$, the optimal control law is of threshold type:

$$\bar{\psi}(\cdot, a) = \begin{cases} 1 & \text{if } 0 \leq a < a^{\text{thresh}} \\ 0 & \text{if } a^{\text{thresh}} < a \leq 1 \end{cases}$$
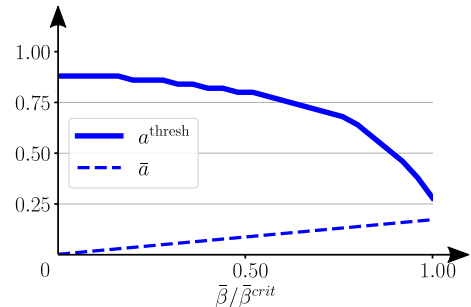

Fig. 4: Plot of a^{thresh} and $\bar{a}$ as a function of $\bar{\beta}/\bar{\beta}^{\text{crit}}$.
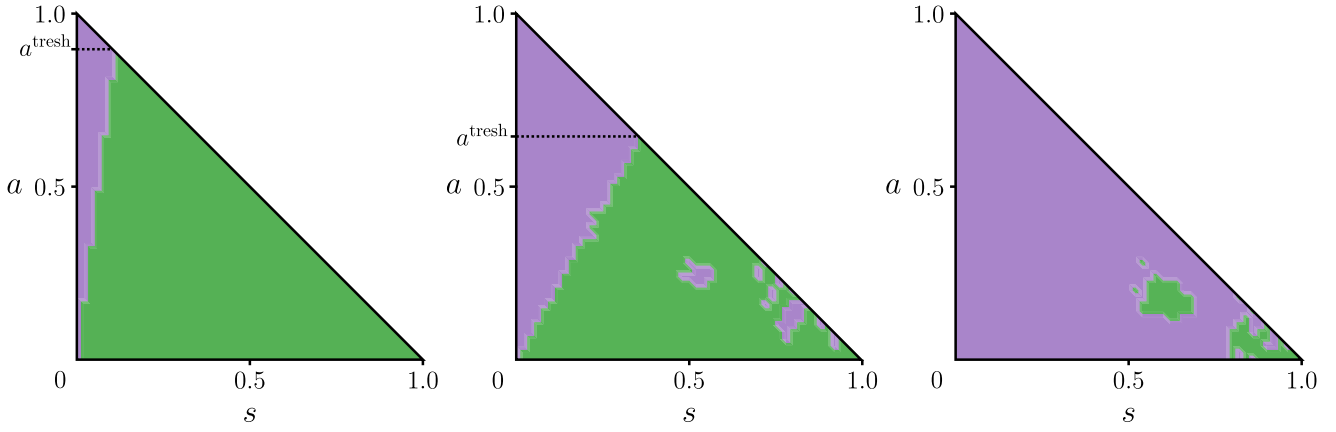
Fig. 5: Stationary optimal control policies for the partially observed case; $\alpha = 0.9$ and $\bar{\beta}/\beta^{\mathrm{crit}} = 0.1, 0.8, 1.1$ for the three plots. Purple region is $U^{\mathrm{opt}} = 0$ and the green region is $U^{\mathrm{opt}} = 1$. The islands are numerical artifacts.

where the threshold $a^{\mathrm{thresh}} \in (0,1)$. Furthermore $\lambda^{\mathrm{SA}}\bar{\beta}/\lambda^{\mathrm{AI}} < a^{\mathrm{thresh}}$. The function $\underline{\lambda}^{\mathrm{AI}}(\bar{\beta})$ is monotonic in its argument and $\lim_{\bar{\beta}\downarrow 0}\underline{\lambda}^{\mathrm{AI}}(\bar{\beta}) = 0$.

With $\lambda^{\mathrm{AR}} > 0$, an analytical treatment has so far not been possible. However, numerical evidence suggests that the policies are also of threshold type even with $\lambda^{\mathrm{AR}} > 0$. Three such numerically obtained stationary policies are depicted in Fig. 5. The policies were computed using the method of lines numerical algorithm described in Appendix D. From these numerically computed policies, a threshold of a^{thresh} is identified such that an agent becomes inactive for $A_t > a^{\mathrm{thresh}}$ (see Fig. 5). By identifying the thresholds for different choices of $\bar{\beta} < \bar{\beta}^{\mathrm{crit}}$, a plot of a^{thresh} as a function of $\frac{\bar{\beta}}{\beta^{\mathrm{crit}}}$ is obtained as depicted in Fig. 4. As $\bar{\beta} \uparrow \bar{\beta}^{\mathrm{crit}}$, the threshold $a^{\mathrm{thresh}} \downarrow 0$ and the agent ceases to be active.

C. Belief of an Active Agent who Shows No Symptoms

The analysis thus far has revealed two important insights captured by the critical value for $\bar{\beta}$ and the threshold for a:

1) A susceptible agent in the fully observed case is active if $\bar{\beta}$ is small enough ($\bar{\beta} < \bar{\beta}^{\mathrm{crit}}$).
2) In the absence of symptoms, a partially observed agent is active if its belief A_t is small enough ($A_t < a^{\mathrm{thresh}}$).

Therefore, to determine the actions of an agent in the partially observed case, it becomes important to understand the evolution of the process $\{A_t : 0 \le t < \tau\}$. This is the subject of the following proposition.

Proposition 5: Set $\bar{a} := \frac{\lambda^{\mathrm{SA}}\bar{\beta}+\eta}{\lambda^{\mathrm{AI}}}$. Suppose $t < \tau$, and

$$S_0 = 1 \qquad \text{(agent starts out as susceptible)}$$
$$U_s = 1 \quad \forall\, s < t \qquad \text{(agent has been active upto time } t)$$

Then,

$$A_t < \bar{a}, \qquad 0 \le t < \tau$$

Moreover if $\lambda^{\mathrm{AR}} = 0$, then $A_t \uparrow \bar{a}$ as $t \to \infty$.

A numerical illustration of the result of this proposition appears as part of Fig. 6. The most important point about this plot is that the agent's belief that it is asymptomatic scales as $\eta + \mathcal{O}(\bar{\beta})$.
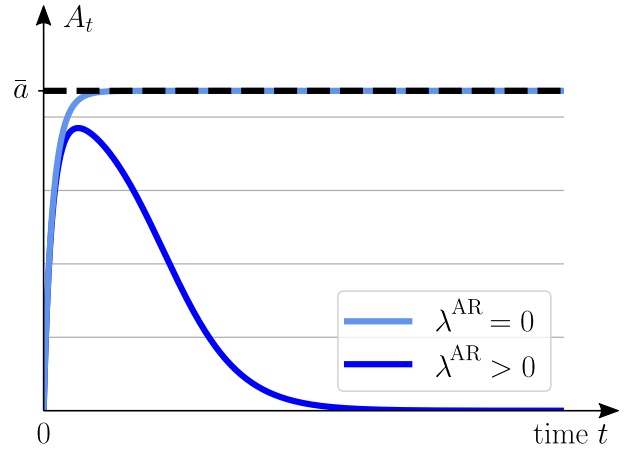

Fig. 6: Evolution of A_t for an active agent who starts out as susceptible.

D. Qualitative Discussion

We are now ready to provide a qualitative answer to the two primary questions we had raised: (i) How does a single agent choose its control input U?; and (ii) How does that choice (made individually and independently by many agents) affect the evolution of disease in a large heterogeneous population?

The answer to the first question is as follows:

1) In the fully observed setting, a susceptible agent is active if $\bar{\beta} < \bar{\beta}^{\mathrm{crit}}$. In this setting, an asymptomatic agent isolates (Table I).
2) In the partially observed setting, an agent's decision is based on its belief (S_t, A_t, R_t). An agent who starts out as susceptible and active has a belief $A_t < \bar{a} = \eta + \mathcal{O}(\bar{\beta})$ (Fig. 6).
3) For small values of $\bar{\beta} < \bar{\beta}^{\mathrm{crit}}$, the quantity $\bar{a} < a^{\mathrm{thresh}}$ (Fig. 4). Therefore, an agent's optimal decision is to continue to be active. This is illustrated in Fig. 7 for $\bar{\beta}/\bar{\beta}^{\mathrm{crit}} = 0.5$.

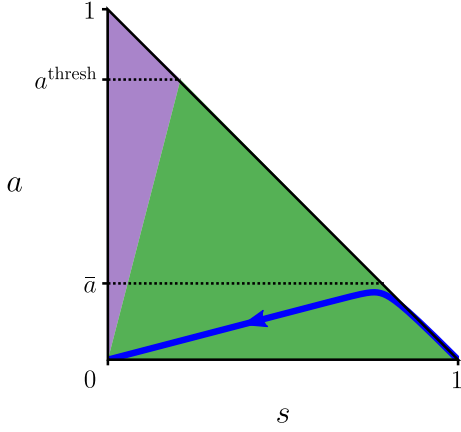The answer to the second question is as follows:

Fig. 7: The trajectory of Fig. 6 (for $\lambda^{\text{AR}} > 0$) depicted on the simplex. An agent's optimal decision is to remain active, unless it starts to show symptoms.

1) Even though an individual agent's conditional probability A_t is small at any time $t < \tau$, the law of large numbers (LLN) dictates that in a large population with N agents, $\mathsf{E}(A_t)$ fraction of agents are asymptomatic with high probability.

2) The basic reproduction number R_o (pronounced R nought) is defined as $\frac{T^{\text{r}}}{T^{\text{c}}}$ where T^{r} is the average time until removal (in our settings, either showing symptoms $x = \boldsymbol{i}$ or recovery $x = \boldsymbol{r}$) and T^{c} is the average time between infectious contacts. For our model, this ratio is $\frac{\lambda^{\text{SA}}}{\lambda^{\text{AI}} + \lambda^{\text{AR}}} = R_o$.

3) If $R_o > 1$ then for small values of $\bar{\beta}$, optimal actions of individual agents (to remain active) still causes an epidemic.

The above is an example of rational irrationality where rational choices of individual agents lead to irrational outcomes for the population [18], [19]. It is notable that the irrationality here is entirely a consequence of the lack of information (with full information, asymptomatic agents isolate).

A limitation of these arguments is that the analysis is rather qualitative and moreover relies on the assumption of a given stationarity $\beta_t \equiv \bar{\beta}$. Indeed, this is not necessarily true in the case of an epidemic: The process β is non-stationary and also a result of the individual actions of many agents. This motivates the need for a more refined analysis based on a mean-field game model as described next.

V. MEAN-FIELD GAME MODEL

In this section, we consider a population of heterogeneous agents. As already noted, $\rho_t(x, u; \theta)$ is the joint distribution of the state-action pair (X_t, U_t) at time t, conditioned on the attribute θ. The marginal pmf $\rho_t(x; \theta)$ evolves according to the mean-field equations:

$$\frac{\mathrm{d}\rho_t}{\mathrm{d}t}(\boldsymbol{s}; \theta) = -\lambda^{\text{SA}} \beta_t \int_0^1 u \rho_t(\boldsymbol{s}, u; \theta) \, \mathrm{d}u - \eta \rho_t(\boldsymbol{s}; \theta) \quad (6\text{a})$$

$$\frac{\mathrm{d}\rho_t}{\mathrm{d}t}(x; \theta) = (\mathcal{A}^\dagger \rho_t)(x; \theta), \quad x \in \{\boldsymbol{i}, \boldsymbol{a}, \boldsymbol{r}, \boldsymbol{d}\} \quad (6\text{b})$$

from a given initial condition $\rho_0(x; \theta)$; $\mathcal{A}^\dagger$ is the adjoint of the generator $\mathcal{A}$ of the Markov process X.

Each agent in the population chooses its control according to the optimal control law $\psi_t(\cdot; \theta)$ obtained from solving the HJB equation (the notation denotes the dependence on the attribute θ). In the two cases,

$$\begin{aligned} \text{(fully obsvd.)} \qquad & U_t = \psi_t(X_t; \theta) \\ \text{(part. obsvd.)} \qquad & U_t = \psi_t(\pi_t; \theta) \end{aligned}$$

where $\pi_t = \mathsf{P}(X_t \mid \mathcal{Y}_t)$ is the belief state; the use of the common notation $\psi_t(\cdot; \theta)$ should not cause any confusion because the two cases are treated in separate subsections.

The following operators are of interest (see Sec. II-C for definition of function space $\mathcal{M}$ and $\mathcal{U}$):

1) The operator $\Psi : \mathcal{M} \to \mathcal{U}$, defined as

$$\Psi(\beta) = \arg\min_{U \in \mathcal{U}} \mathsf{J}(U; \beta)$$

2) The operator $\Xi : \mathcal{U} \to \mathcal{M}$, defined according to (6) where β is obtained using (1).

Assuming that the two operators are well-defined, we have:

Definition 1: A *mean-field equilibrium* (MFE) is a fixed point β such that $\Xi(\Psi(\beta)) = (\beta)$.

A. Fully Observed Case

For the fully observed problem, a complete characterization of the MFG equilibrium is described in the following proposition. The result is a minor extension of a similar result appearing in our prior paper [1] (for λ^{AR} and $\eta = 0$) and therefore its proof is omitted.

Proposition 6: The unique mean-field equilibrium is

$$\beta_t = 0 \quad \forall \ t \geq 0$$

This result represents a mathematical expression of the thought experiment described in Sec. I. Its main utility is to set up the problem whereby the effects of some of the underlying assumptions – perfect rationality and perfect information – can be investigated. As with the preceding sections, the main focus of the MFG modeling is in the partially observed settings, which is the subject of the following subsection.

B. Partially Observed Case

Consider the space of probability distributions on the belief space $\mathcal{P}^1 \cup \mathcal{P}^2$. The random vector (S_t, A_t) is well-defined on the set $[t < \tau]$, and we denote by $p_t(s, a)$ its density for $0 \leq s \leq 1$, $s + a \leq 1$:

$$\begin{aligned} & \mathsf{P}([s < S_t < s + \mathrm{d}s] \cap [a < A_t < a + \mathrm{d}a] \cap [t < \tau]) \\ & = p_t(s, a) \, \mathrm{d}a \, \mathrm{d}s, \quad t \geq 0 \end{aligned}$$

Assumption 1: The density $p_t(s, 1-s) = p_t(s, 0) = 0$ for all $t \geq 0$, $s \in [0, 1]$.

Under Assumption 1 that an agent uses the optimal control $U_t = U_t^{\text{opt}} = \psi_t(S_t, A_t)$ for $0 \leq t < \tau$, the density process $\{p_t(s, a) \in [0, \infty) : 0 \leq s \leq 1, s + a \leq 1, t \geq 0\}$ solves the

FPK equation whose derivation, along with the expressions for $f_s(s,a,t)$ and $f_{\boldsymbol{a}}(s,a,t)$, appears in Appendix C:

$$\frac{\partial p_t}{\partial t}(s,a) = -\frac{\partial}{\partial s}\big(f_s(s,a,t)p_t(s,a)\big)$$
$$-\frac{\partial}{\partial a}\big(f_{\boldsymbol{a}}(s,a,t)p_t(s,a)\big) - a\lambda^{\text{AI}}p_t(s,a) \quad (7)$$

where $p_0(s,a)$ is the initial density (assumed given). By using the tower property,

$$\rho_t(\boldsymbol{a}) = \mathsf{P}([X_t = \boldsymbol{a}])$$
$$= \mathsf{E}(\pi_t(\boldsymbol{a})) = \mathsf{E}(A_t 1_{[t<\tau]}) = \int_0^1\int_0^{1-s} a p_t(s,a)\,\mathrm{d}a\,\mathrm{d}s$$

and therefore we have

$$\frac{\mathrm{d}\rho_t}{\mathrm{d}t}(\boldsymbol{i}) = \lambda^{\text{AI}}\int_0^1\int_0^{1-s} a p_t(s,a)\,\mathrm{d}a\,\mathrm{d}s - (\lambda^{\text{ID}}+\lambda^{\text{IR}})\rho_t(\boldsymbol{i}) \tag{8a}$$

$$\frac{\mathrm{d}\rho_t}{\mathrm{d}t}(x) = (\mathcal{A}^\dagger \rho_t)(x), \quad x \in \{\boldsymbol{r},\boldsymbol{d}\} \tag{8b}$$

where expression for $\mathcal{A}^\dagger$ is obtained from the transition graph.

With a heterogeneous population, the notation $p_t(s,a;\theta)$ is used to denote the density conditioned on the attribute θ. The mean-field process is then consistently obtained as

$$\beta_t = \sum_\theta \mathsf{p}(\theta)\int_0^1\int_0^{1-s} a\psi_t(s,a;\theta)p_t(s,a;\theta)\,\mathrm{d}a\,\mathrm{d}s \quad (9)$$

This completes the derivation of the system of equations for the partially observed MFG: Eq. (7)-(8) is the forward FPK equation. Eq. (4) is the backward HJB equation. Eq. (9) defines the consistency relationship that links the two equations. Its solution is an MFE (satisfies Defn. 1).

The thrust of the ongoing work is to obtain numerical solutions of the MFG equations and use it to explain differences between the data from the omicron and delta waves.

References

[1] S. Y. Olmez, S. Aggarwal, J. W. Kim, E. Miehling, T. Başar, M. West, and P. G. Mehta, "Modeling presymptomatic spread in epidemics via mean-field games," *arXiv preprint arXiv:2111.10422*, 2021.

[2] L. Rivett, S. Sridhar, D. Sparkes, M. Routledge, N. K. Jones, S. Forrest, J. Young, J. Pereira-Dias, W. L. Hamilton, M. Ferris *et al.*, "Screening of healthcare workers for SARS-CoV-2 highlights the role of asymptomatic carriage in COVID-19 transmission," *Elife*, vol. 9, p. e58728, 2020.

[3] D. C. Buitrago-Garcia, D. Egli-Gany, M. J. Counotte, S. Hossmann, H. Imeri, G. Salanti, and N. Low, "The role of asymptomatic SARS-CoV-2 infections: rapid living systematic review and meta-analysis," *MedRxiv*, 2020.

[4] J. K. Bender, M. Brandl, M. Höhle, U. Buchholz, and N. Zeitlmann, "Analysis of asymptomatic and presymptomatic transmission in SARS-CoV-2 outbreak, Germany, 2020," *Emerging infectious diseases*, vol. 27, no. 4, p. 1159, 2021.

[5] W. E. Wei, Z. Li, C. J. Chiew, S. E. Yong, M. P. Toh, and V. J. Lee, "Presymptomatic transmission of SARS-CoV-2—Singapore, january 23–march 16, 2020," *Morbidity and Mortality Weekly Report*, vol. 69, no. 14, p. 411, 2020.

[6] R. Elie, E. Hubert, and G. Turinici, "Contact rate epidemic control of COVID-19: an equilibrium view," *Mathematical Modelling of Natural Phenomena*, vol. 15, p. 35, 2020.

[7] A. Aurell, R. Carmona, G. Dayanikli, and M. Lauriere, "Optimal incentives to mitigate epidemics: a Stackelberg mean field game approach," *arXiv preprint arXiv:2011.03105*, 2020.

[8] ——, "Finite state graphon games with applications to epidemics," *arXiv preprint arXiv:2106.07859*, 2021.

[9] E. Hubert, T. Mastrolia, D. Possamaï, and X. Warin, "Incentives, lockdown, and testing: from thucydides's analysis to the covid-19 pandemic," *arXiv preprint arXiv:2009.00484*, 2020.

[10] W. Lee, S. Liu, H. Tembine, W. Li, and S. Osher, "Controlling propagation of epidemics via mean-field control," *arXiv preprint arXiv:2006.01249*, 2020.

[11] S. Cho, "Mean-field game analysis of SIR model with social distancing," *arXiv preprint arXiv:2005.06758*, 2020.

[12] J. Doncel, N. Gast, and B. Gaujal, "A mean field game analysis of SIR dynamics with vaccination," *Probability in the Engineering and Informational Sciences*, pp. 1–18, 2020.

[13] H. Tembine, "COVID-19: Data-driven mean-field-type game perspective," *Games*, vol. 11, no. 4, p. 51, 2020.

[14] N. Saldi, T. Başar, and M. Raginsky, "Partially-observed discrete-time risk-sensitive mean-field games," in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 317–322.

[15] N. Saldi, T. Başar, and M. Raginsky, "Approximate nash equilibria in partially observed stochastic games with mean-field interactions," *Mathematics of Operations Research*, vol. 44, no. 3, pp. 1006–1033, 2019. [Online]. Available: https://doi.org/10.1287/moor.2018.0957

[16] F. Confortola and M. Fuhrman, "Filtering of continuous-time markov chains with noise-free observation and applications," *Stochastics An International Journal of Probability and Stochastic Processes*, vol. 85, no. 2, pp. 216–251, 2013.

[17] S. Cobey, J. Bloom, T. Starr, and N. Lash, "We study virus evolution. here's where we think the coronavirus is going." *New York Times*, 2022. [Online]. Available: https://www.nytimes.com/interactive/2022/03/28/opinion/coronavirus-mutation-future.html

[18] J. Cassidy, "Rational irrationality: The real reason that capitalism is so crash-prone," Oct. 5th 2009.

[19] H. Yin, P. G. Mehta, S. P. Meyn, and U. V. Shanbhag, "Synchronization of oscillators is a game," in *IEEE Transactions on Automatic Control*, vol. 57, no. 4, April 2012, pp. 920–935.

[20] J. Xiong, *An Introduction to Stochastic Filtering Theory*. Oxford University Press, 2008.

[21] W. H. Fleming and H. M. Soner, *Controlled Markov Processes and Viscosity Solutions*, 2nd ed., ser. Stochastic Modelling and Applied Probability ; 25. New York: Springer-Verlag, 2006.

Appendix

A. Proof of Proposition 3

With $\beta_t = \bar{\beta}$, the HJB equation for $v_t(\boldsymbol{s})$ is

$$-\frac{\mathrm{d}v_t}{\mathrm{d}t}(\boldsymbol{s}) + (\gamma+\eta)v_t(\boldsymbol{s})$$
$$= \eta\bar{\phi}(\boldsymbol{a}) + \min_{u\in[0,1]}\big(\lambda^{\text{SA}}\bar{\beta}\big(\bar{\phi}(\boldsymbol{a})-v_t(\boldsymbol{s})\big)-\alpha\big)u$$

We investigate its stationary solutions $v_t(\boldsymbol{s}) = \bar{\phi}(\boldsymbol{s})$ in which case the stationary HJB equation is

$$(\gamma+\eta)\bar{\phi}(\boldsymbol{s}) = \eta\bar{\phi}(\boldsymbol{a})$$
$$+ \min_{u\in[0,1]}\underbrace{\big(\lambda^{\text{SA}}\bar{\beta}\big(\bar{\phi}(\boldsymbol{a})-\bar{\phi}(\boldsymbol{s})\big)-\alpha\big)}_{=:M}u \quad (10)$$

We have the following two cases:

- If $\bar{\beta} < \frac{\alpha}{\lambda^{\text{SA}}\bar{\phi}(\boldsymbol{a})}\left(1+\frac{\eta}{\gamma}\right)$, then $\bar{\phi}(\boldsymbol{s}) = \frac{(\lambda^{\text{SA}}\bar{\beta}+\eta)\bar{\phi}(\boldsymbol{a})-\alpha}{\lambda^{\text{SA}}\bar{\beta}+\gamma+\eta}$ solves the HJB equation with the minimizing choice of $u=1$, because

$$M = \lambda^{\text{SA}}\bar{\beta}\left(\frac{\gamma\bar{\phi}(\boldsymbol{a})+\alpha}{\lambda^{\text{SA}}\bar{\beta}+\gamma+\eta}\right) - \alpha < 0$$
$$\iff \lambda^{\text{SA}}\bar{\beta}\alpha + \gamma\alpha + \eta\alpha > \lambda^{\text{SA}}\bar{\beta}\gamma\bar{\phi}(\boldsymbol{a}) + \lambda^{\text{SA}}\bar{\beta}\alpha$$
$$\iff \bar{\beta} < \frac{\alpha}{\lambda^{\text{SA}}\bar{\phi}(\boldsymbol{a})}\left(1+\frac{\eta}{\gamma}\right)$$

- If $\bar{\beta} > \frac{\alpha}{\lambda^{\text{SA}}\bar{\phi}(\boldsymbol{a})}\left(1 + \frac{\eta}{\gamma}\right)$, then $\bar{\phi}(\boldsymbol{s}) = \frac{\eta}{\gamma+\eta}\bar{\phi}(\boldsymbol{a})$ solves the HJB equation with the minimizing choice of $u = 0$, because

$$M = \frac{\lambda^{\text{SA}}\bar{\beta}\gamma\bar{\phi}(\boldsymbol{a})}{\gamma + \eta} - \alpha > 0$$

B. Proof of Proposition 5

Because $A_0 = 0$, and the solution is continuous as a function of time t, it suffices to show that

$$A_t = \frac{\lambda^{\text{SA}}\bar{\beta} + \eta}{\lambda^{\text{AI}}} \implies \frac{dA_t}{dt} < 0$$

Setting $A_t = \frac{\lambda^{\text{SA}}\bar{\beta}+\eta}{\lambda^{\text{AI}}}$ on the righthand side of (3),

$$\frac{dA_t}{dt} = (\lambda^{\text{SA}}\bar{\beta} + \eta)\left(S_t - 1 + \frac{\lambda^{\text{SA}}\bar{\beta} + \eta - \lambda^{\text{AR}}}{\lambda^{\text{AI}}}\right)$$

The desired conclusion follows because $S_t + A_t \le 1$ and therefore $S_t - 1 + \frac{\lambda^{\text{SA}}\bar{\beta}+\eta}{\lambda^{\text{AI}}} \le 0$ if $A_t = \frac{\lambda^{\text{SA}}\bar{\beta}+\eta}{\lambda^{\text{AI}}}$.

C. Derivations

Let $\mathcal{P} := \mathcal{P}^1 \cup \mathcal{P}^2$. The belief process $\pi := \{\pi_t \in \mathcal{P} : t \ge 0\}$ is a Markov process [20, Theorem 1.7]. The HJB equations and FPK equations are easily derived once we obtain the infinitesimal generator of the process.

Infinitesimal generator: Consider a smooth test function $v : \mathcal{P} \to \mathbb{R}$. Let $\mu \in \mathcal{P}$. The infinitesimal generator (for the general time-inhomogeneous case) is

$$(\mathcal{A}_t^u v)(\mu) = \lim_{\delta t \downarrow 0} \frac{\mathsf{E}(v(\pi_{t+\delta t})|\pi_t = \mu) - v(\mu)}{\delta t}$$

There are two cases to consider:

- If $\mu \in \mathcal{P}^2$, upon identifying the measures $\{\delta_{\boldsymbol{i}}, \delta_{\boldsymbol{r}}, \delta_{\boldsymbol{d}}\}$ with the states $\{\boldsymbol{i}, \boldsymbol{r}, \boldsymbol{d}\}$, the generator is the same as the generator for the Markov process.
- If $\mu \in \mathcal{P}^1 \setminus \{\delta_{\boldsymbol{r}}\}$, then using the coordinates (s, a) for $\mathcal{P}^1$, $\mu = [s, a, 0, 1 - s - a, 0]$ for $a \in [0, 1]$. With $\pi_t = \mu$, in the asymptotic limit as $\delta t \to 0$,

$$\pi_{t+\delta t} = \begin{cases} \mu + e_{\boldsymbol{s}}f_{\boldsymbol{s}}(s, a, t)\delta t \\ +e_{\boldsymbol{a}}f_{\boldsymbol{a}}(s, a, t)\delta t + o(\delta t) & \text{w.p. } (1 - a\lambda^{\text{AI}}\delta t) + o(\delta t) \\ \delta_{\boldsymbol{i}} & \text{w.p. } a\lambda^{\text{AI}}\delta t + o(\delta t) \end{cases}$$

where

$$f_{\boldsymbol{s}}(s, a, t) = (-\lambda^{\text{SA}}\beta_t u - \eta + a\lambda^{\text{AI}})s$$
$$f_{\boldsymbol{a}}(s, a, t) = (\lambda^{\text{SA}}\beta_t u + \eta)s + a(-\lambda^{\text{AI}} - \lambda^{\text{AR}} + \lambda^{\text{AI}}a)$$

and $e_{\boldsymbol{s}} = [1, 0, 0, -1, 0]$, $e_{\boldsymbol{a}} = [0, 1, 0, -1, 0]$. Denoting $v(\mu) = \phi(s, a)$, the generator is then easily calculated to be

$$(\mathcal{A}_t^u v)(\mu) = f_{\boldsymbol{s}}(s, a, t)\frac{\partial\phi}{\partial s}(s, a) + f_{\boldsymbol{a}}(s, a, t)\frac{\partial\phi}{\partial a}(s, a)$$
$$+ a\lambda^{\text{AI}}(v(\delta_{\boldsymbol{i}}) - \phi(s, a))$$

where the superscript u denotes the fact that $f_{\boldsymbol{s}}(s, a, t)$, $f_{\boldsymbol{a}}(s, a, t)$, and therefore also the generator, depends also upon u. The subscript t denotes the fact that the generator is

for a time-inhomogeneous Markov process (because β may depend upon time).

Derivation of the HJB equation: The HJB equation (See Sec.III.7 of [21]) is

$$-\frac{\partial v_t}{\partial t}(\mu) + \gamma v_t(\mu) = \min_{u \in [0,1]} (\mathcal{A}_t^u v_t(\mu) + \mu(c(\cdot, u; \alpha)))$$

In order to calculate $\mu(c(\cdot, u; \alpha))$, we need to assume a running cost for state $\boldsymbol{r}$. If we assume $c(\boldsymbol{r}, \cdot; \cdot) = \gamma\bar{\phi}(\boldsymbol{r})$, then we get

$$\int_0^\infty c(\boldsymbol{r}, \cdot; \cdot)e^{-\gamma t}dt = \bar{\phi}(\boldsymbol{r})$$

as desired. Let

$$\mathcal{L}\phi := (1 - s - a)\gamma\bar{\phi}(\boldsymbol{r}) + a\lambda^{\text{AI}}(\bar{\phi}(\boldsymbol{i}) - \phi)$$
$$+ (a\lambda^{\text{AI}} - \eta)s\frac{\partial\phi}{\partial s} + (\eta s + a(a\lambda^{\text{AI}} - \lambda^{\text{AI}} - \lambda^{\text{AR}}))\frac{\partial\phi}{\partial a}$$
$$\mathcal{M}_t(s, a) := a - (s + a)\alpha + \lambda^{\text{SA}}\beta_t s\left(\frac{\partial\phi_t}{\partial a} - \frac{\partial\phi_t}{\partial s}\right)$$

Thus for $\mu = [s, a, 0, 1 - s - a, 0]$ and $v_t(\mu) = \phi_t(s, a)$, the HJB equation (4) is obtained because

$$\mu(c(\cdot, u; \alpha)) = (1 - s - a)\gamma\bar{\phi}(\boldsymbol{r}) + s(-\alpha u) + a(1 - \alpha)u$$
$$= (1 - s - a)\gamma\bar{\phi}(\boldsymbol{r}) + (a - (s + a)\alpha)u$$

The boundary conditions (5) follow from the observation that

$$\phi_t(s, a)$$
$$= s\min_{U \in L_{\mathcal{F}}^2}\mathsf{E}\left(\int_t^T e^{-\gamma s}c(X_s, U_s; \alpha)ds + e^{-\gamma T}\phi(X_T)\right.$$
$$\left.\mid X_t = \boldsymbol{s}\right) + a\bar{\phi}(\boldsymbol{a}) + (1 - s - a)\bar{\phi}(\boldsymbol{r})$$

Derivation of the FPK equation: We derive the adjoint of the generator $\mathcal{A}^u$ where dependence on t is suppressed for notational ease. Let ρ be a measure on $\mathcal{P}$. On $\mathcal{P}^1$, ρ has density $p(s, a)$. Consider $\rho(\mathcal{A}^u v) = \int \mathcal{A}^u v(\mu)\rho(d\mu) =$

$$\int_0^1 \int_0^{1-s} p(s, a)\left[f_{\boldsymbol{s}}(s, a)\frac{\partial\phi}{\partial s} + f_{\boldsymbol{a}}(s, a)\frac{\partial\phi}{\partial a}\right.$$
$$\left. + a\lambda^{\text{AI}}(v(\delta_{\boldsymbol{i}}) - \phi(s, a))\right]da\,ds - \rho(\delta_{\boldsymbol{i}})(\lambda^{\text{IR}} + \lambda^{\text{ID}})v(\delta_{\boldsymbol{i}})$$
$$+ \rho(\delta_{\boldsymbol{r}})\lambda^{\text{IR}}v(\delta_{\boldsymbol{r}}) + \rho(\delta_{\boldsymbol{d}})\lambda^{\text{ID}}v(\delta_{\boldsymbol{d}}) \tag{11}$$

First note that

$$\int_0^{1-a} p(s, a)f_{\boldsymbol{s}}(s, a)\frac{\partial\phi}{\partial s}ds = p(s, a)f_{\boldsymbol{s}}(s, a)\phi(s, a)|_{s=0}^{s=1-a}$$
$$- \int_0^{1-a} \phi(s, a)\frac{\partial}{\partial s}(p(s, a)f_{\boldsymbol{s}}(s, a))ds$$

where boundary terms vanish because $f_{\boldsymbol{s}}(0, a) = 0$ and $p(1 - a, a) = 0$ (Assumption 1). One can also write

$$\int_0^{1-s} p(s, a)f_{\boldsymbol{a}}(s, a)\frac{\partial\phi}{\partial a}da = p(s, a)f_{\boldsymbol{a}}(s, a)\phi(s, a)|_{a=0}^{a=1-s}$$
$$- \int_0^{1-s} \phi(s, a)\frac{\partial}{\partial a}(p(s, a)f_{\boldsymbol{a}}(s, a))da$$

where boundary terms vanish because $p(s, 1-s) = p(s, 0) = 0$ (Assumption 1). Using these two results, one can rewrite the righthand side of (11) as

$$
-\int_0^1 \int_0^{1-s} \phi(s, a) \left[\frac{\partial}{\partial s} \big(p(s, a) f_s(s, a) \big) + \frac{\partial}{\partial a} \big(p(s, a) f_a(s, a) \big) \right.
$$
$$
\left. + a \lambda^{\text{AI}} p(s, a) \right] \mathrm{d}a \, \mathrm{d}s + v(\delta_r) \lambda^{\text{IR}} \rho(\delta_r) + v(\delta_d) \lambda^{\text{ID}} \rho(\delta_d)
$$
$$
+ v(\delta_i) \left(\int_0^1 \int_0^{1-s} a \lambda^{\text{AI}} p(s, a) \, \mathrm{d}a \, \mathrm{d}s - (\lambda^{\text{IR}} + \lambda^{\text{ID}}) \rho(\delta_i) \right)
$$

D. Numerical Algorithm

The HJB equation (4) is numerically solved using the method of lines. In this method, the spatial coordinate $(s, a) \in D$ is discretized over a finite grid. The partial derivatives $\partial \phi_t / \partial s$ and $\partial \phi_t / \partial a$ are approximated using a finite difference approximation at each point on the grid. The time variable t is treated as a continuous variable and after spatial discretization, the resulting ordinary differential equations are numerically integrated (backward in time) using a standard numerical procedure. For this purpose, a sufficiently large terminal time T is chosen with the following terminal condition:

$$
\phi_T(s, a) = \begin{cases} a\bar{\phi}(a) + (1-a)\bar{\phi}(r) & a \le \frac{1-s}{2} \\ \phi_T(s, 1-s-a) & \\ + (2a + s - 1)\big(\bar{\phi}(a) - \bar{\phi}(r)\big) & a > \frac{1-s}{2} \end{cases}
$$

The algorithm was used to compute the stationary solutions depicted in Fig. 5.