IEEE Control Systems Letters paper presented at
2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

SIS Epidemic Propagation under Strategic Non-myopic Protection: A Dynamic Population Game Approach

Urmee Maitra[1], Ashish R. Hota[1] and Vaibhav Srivastava[2]

***Abstract*—We consider a dynamic game setting in which a large population of strategic individuals decides whether to adopt protective measures to protect themselves against an infectious disease, specifically the susceptible-infected-susceptible (SIS) epidemic. Protection is costly and partially effective, and adopting protection reduces the probability of becoming infected for susceptible individuals and the probability of transmitting the infection for infected individuals. In a departure from most prior works that assume the decision-makers to be myopic, we model individuals who choose their actions to maximize the infinite horizon discounted expected reward. We define the notion of best response and stationary Nash equilibrium in this class of games, and completely characterize the equilibrium policy and stationary state distribution for different parameter regimes. Numerical results illustrate the evolution and convergence of the infected proportion and the policy of protection adoption to the equilibrium.**

***Index Terms*—Game theory, Stochastic systems.**

I. Introduction

As observed during the COVID-19 pandemic, the evolution or spread of infectious diseases depends critically on the behavior of susceptible and healthy individuals in society. In particular, the adoption of (partially effective) protective measures such as wearing masks and adhering to social distancing plays a critical role in shaping the future growth of the disease. Due to the selfish and strategic nature of individual behavior, past work has examined the interplay of epidemic evolution and individual behavior in the framework of game theory; see [1] for a comprehensive review. Since the number of decision-makers is typically large in this setting, the frameworks of population games [2] and mean-field games [3] have been adopted.

In a seminal work [4], the authors consider the class of susceptible-infected-susceptible (SIS) epidemic where a large population of individuals decide whether to adopt protection or not and analyze the equilibrium of the epidemic dynamics under Nash equilibrium strategies. Follow-up works [5] and [6] extended the above setting to include network interactions. However, these works assumed that adopting protection completely stops disease transmission, and did not consider (evolutionary) learning dynamics for agents to update their strategies and converge to the Nash equilibrium. Coupled evolution of epidemic and behavioral dynamics have been studied very recently in [7]–[12]. Specifically, [7], [8] consider partially effective protection against the SIS and SIRI epidemics with the cost function defined via the instantaneous risk of becoming infected and model behavioral evolution via the replicator dynamics. In [12], the authors consider a social influence factor in the cost function and model behavioral evolution via imitation dynamics. Such coupled evolution often leads to oscillatory convergence behavior in the SIS epidemic as shown in [7], [10], [12]. The authors in [9] examine how to dynamically update the reward matrix to ensure convergence to a desired equilibrium. In the above works, each individual strategically maximizes a myopic reward function which is possibly a function of the current infection prevalence, but does not include the impact of the chosen action on the future infection state and reward.

However, in the context of epidemics, individuals often reason about their future infection state while choosing their present action and aim to maximize the aggregate reward over a longer time horizon, i.e., they are forward-looking as opposed to myopic. Despite the motivation, there have been limited investigations on epidemic games with forward-looking agents. A notable recent exception is [13] which leverages the framework of dynamic population games [14] and studies the strategies of individuals who maximize infinite horizon discounted expected reward in the context of a variant of the susceptible-infected-recovered (SIR) epidemic that includes asymptomatic infections and recovery. While the dynamic population game framework is related to the class of mean-field games which has been examined in the context of epidemics (mostly the SIR model and its variants) [3], unlike mean-field games, it enables applying evolutionary learning algorithms to study the evolution of individual behavior and convergence to equilibrium strategies.

In this paper, we consider the SIS epidemic setting with a large population of agents who decide whether to adopt protection or not at each time step in order to maximize the infinite horizon discounted expected reward. Protection is partially effective in the sense that infected individuals adopting protection are less likely to infect others and susceptible individuals adopting it are less likely to become infected compared to unprotected individuals. The instantaneous reward consists of the cost of adopting protection, the cost of infection, and the cost of being unprotected while infected (e.g., the cost of violating quarantine norms). Similar to [14], we define the best response as a Markovian policy that maximizes the single-stage deviation value function [15][1] and accordingly,

This work was supported in part by a joint Indo-US DST-NSF research project awarded by IDEAS, Technology Innovation Hub, ISI Kolkata, India, and by ARO grant W911NF-18-1-0325.

[1]U. Maitra and A. R. Hota are with the Department of Electrical Engineering, IIT Kharagpur, Kharagpur, West Bengal, India, 721302. Email: urmeemaitra93@kgpian.iitkgp.ac.in, ahota@ee.iitkgp.ac.in.

[2]V. Srivastava is with the Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI 48824. Email: vaibhav@msu.edu

[1]analogous to the notion of the state-action value function in the Markov decision process literature.

Copyright ©2023 IEEE

introduce the notion of stationary Nash equilibrium at which both the infection state distribution and the policy are invariant. We then prove a complete characterization of the stationary Nash equilibrium (including the infected proportion at the endemic equilibrium and the policy adopted by susceptible and infected individuals) for different parameter regimes. Finally, we examine the impact of the discount factor and cost of adopting protection on equilibrium infection level and policy via simulations.

II. SIS Epidemic Model and State Transitions under Protection Adoption

We consider a homogeneous large population of agents where each agent can be in either of the two states: susceptible ($\mathtt{S}$) or infected ($\mathtt{I}$) with $\mathtt{Z} := \{\mathtt{S}, \mathtt{I}\}$. Susceptible agents are at risk of becoming infected when they come in contact with infected agents, whereas infected agents recover at an exogenous rate and once again become susceptible to the disease. The state distribution $d \in \mathcal{D} := \Delta(\mathtt{Z})$, where $\Delta(\mathcal{X})$ is the space of probability distributions supported on $\mathcal{X}$. Specifically, $d[\mathtt{s}] \in [0,1]$ denotes the proportion of agents in state $\mathtt{s}$, and $\sum_{\mathtt{s}\in\mathtt{Z}} d[\mathtt{s}] = 1$.

A. Actions and Policies

At any time instant, each individual chooses to either adopt protection (i.e., using masks or personal protective equipment (PPE) kits or adhering to social distancing guidelines) denoted by $\mathtt{P}$, or decides to remain unprotected, denoted by $\mathtt{U}$. Formally, we denote this choice as an action $a \in \mathcal{A} = \{\mathtt{P}, \mathtt{U}\}$. A Markovian policy denoted by $\pi : \mathtt{Z} \to \Delta(\mathcal{A})$, maps an agents state $\mathtt{s} \in \mathtt{Z}$ to a randomization over the actions $a \in \mathcal{A}$, and $\pi[a|\mathtt{s}]$ denotes the probability that an agent in infection state $\mathtt{s}$ chooses action a. All agents are assumed to be homogeneous and follow the same policy π. The (time-varying) social state is defined to be the concatenation of the policy and state distribution (π, d).

B. Individual State Transitions

We assume that a susceptible agent upon adopting protection is $\gamma \in (0,1)$ times (less) prone to infection compared to an unprotected susceptible agent. Similarly, an infected agent adopting protection is less likely to transmit or cause a new infection. Formally, let $\beta_{\mathtt{U}} \in (0,1)$ and $\beta_{\mathtt{P}} \in (0,1)$ denote the probabilities of an infected individual causing a new infection if it is unprotected and protected, respectively. We assume that $\beta_{\mathtt{U}} > \beta_{\mathtt{P}}$. In other words, protection is partially effective in preventing as well as transmitting new infections. Finally, an infected agent recovers from the infection and transits back to the susceptible state with probability $\delta \in (0,1)$ which is independent of any external process.

We now define the transition probabilities for every state-action pair. Since $\mathbb{P}[\mathtt{s}^+ = \mathtt{S}|\mathtt{s}, a] = 1 - \mathbb{P}[\mathtt{s}^+ = \mathtt{I}|\mathtt{s}, a]$, we will only define $\mathbb{P}[\mathtt{s}^+ = \mathtt{I}|\mathtt{s}, a]$. For an infected agent, the probability that it will become susceptible in the next time-step is δ independent of any action it chooses and the current social state. Accordingly, we have

$$\mathbb{P}[\mathtt{s}^+ = \mathtt{I}|\mathtt{s} = \mathtt{I}, a](\pi, d) = 1 - \delta,$$

for $a \in \{\mathtt{P}, \mathtt{U}\}$. For a susceptible agent, the probability of becoming infected depends on its own action as well as the proportion of protected and unprotected infected agents; the latter depends on the policy $\pi[a|\mathtt{I}]$ and the proportion of infected agents $d[\mathtt{I}]$. Define $\beta_\pi := \beta_{\mathtt{U}}\pi[\mathtt{U}|\mathtt{I}] + \beta_{\mathtt{P}}\pi[\mathtt{P}|\mathtt{I}]$, then the probability of a susceptible agent becoming infected, depending on its chosen action, is given by

$$\mathbb{P}[\mathtt{s}^+ = \mathtt{I}|\mathtt{s} = \mathtt{S}, a = \mathtt{U}](\pi, d) = \beta_\pi d[\mathtt{I}],$$
$$\mathbb{P}[\mathtt{s}^+ = \mathtt{I}|\mathtt{s} = \mathtt{S}, a = \mathtt{P}](\pi, d) = \gamma\beta_\pi d[\mathtt{I}].$$

The above expressions define the state transition probabilities for each state-action pair. When an agent chooses its actions following a policy π, the transition probability from current state $\mathtt{s}$ to future state $\mathtt{s}^+$ is given by

$$P_{\mathtt{eff}}[\mathtt{s}^+|\mathtt{s}](\pi, d) = \sum_{a\in\mathcal{A}} \pi[a|\mathtt{s}]\mathbb{P}[\mathtt{s}^+|\mathtt{s}, a](\pi, d). \qquad (1)$$

As a result, under the assumption that agents are homogeneous and follow the same policy π, the state distribution evolves in discrete-time as

$$d_{k+1} = P_{\mathtt{eff}}(\pi_k, d_k)^\top d_k, \qquad (2)$$

where $P_{\mathtt{eff}}(\pi_k, d_k)$ is the stochastic matrix at time-step k with policy π_k and state distribution d_k.

Before we define the rewards associated with different actions and the decision-making framework of the agents, we characterize the infected proportion at the equilibria of the above dynamics for a given Markovian policy π. Under (2), the proportion of infected agents evolves as

$$\begin{aligned} d_{k+1}[\mathtt{I}] &= (1 - d_k[\mathtt{I}])\beta_\pi d_k[\mathtt{I}](\gamma\pi[\mathtt{P}|\mathtt{S}] + \pi[\mathtt{U}|\mathtt{S}]) \\ &\quad + d_k[\mathtt{I}](1-\delta) \\ &=: d_k[\mathtt{I}]\big(1 - \delta + (1 - d_k[\mathtt{I}])\beta_{\pi,\mathtt{eff}}\big), \end{aligned} \qquad (3)$$

where $\beta_{\pi,\mathtt{eff}} := \beta_\pi(\pi[\mathtt{U}|\mathtt{S}] + \gamma\pi[\mathtt{P}|\mathtt{S}])$ captures the effect of protection adoption on the disease transmission rate under policy π. We now state the following result.

Lemma 2.1: At a given policy π adopted by all the agents, there are at most two equilibrium points of the dynamics (3):

- the disease-free equilibrium given by $d^*_{\mathtt{dfe}}[\mathtt{I}] = 0$; and
- the endemic equilibrium given by $d^*_{\mathtt{ee},\pi}[\mathtt{I}] = 1 - \frac{\delta}{\beta_{\pi,\mathtt{eff}}}$.

The endemic equilibrium $d^*_{\mathtt{ee},\pi}[\mathtt{I}]$ exists when $\beta_{\pi,\mathtt{eff}} > \delta$ and is stable in this regime. The disease-free equilibrium always exists and is stable when $\beta_{\pi,\mathtt{eff}} \leq \delta$.

Proof: Note that (3) is analogous to the scalar discrete-time SIS epidemic dynamics with infection rate given by $\beta_{\pi,\mathtt{eff}}$. Thus, the result follows from analogous arguments as the proof of Case 2 of Proposition 4.6 in [6]. ■

In other words, the disease-free equilibrium is stable when the natural recovery rate is larger than the effective infection rate. Consequently, the endemic equilibrium is more interesting and relevant to study the impact of protection adoption and is the focus of this work. We impose the following assumption which guarantees that $d^*_{\mathtt{ee},\pi}[\mathtt{I}]$ always exists.

Assumption 2.2: The parameters satisfy $\gamma\beta_{\mathtt{P}} > \delta$ which guarantees that $\beta_{\pi,\mathtt{eff}} > \delta$ for any policy π.

In the following, for simplicity of notation, we will denote $d^*_{\mathtt{ee},\pi}[\mathtt{I}]$ by $\mathtt{I}^*_{\mathtt{ee},\pi}$.

C. Rewards

We now define the rewards or costs associated with different actions. We assume that an agent adopting the protection at a given time incurs an instantaneous cost $C_{\mathtt{P}} > 0$ irrespective of its infection state. An infected agent incurs an instantaneous cost $C_{\mathtt{I}} > 0$ which represents the discomfort associated with infection. In addition, an infected agent who does not adopt protection incurs an additional cost $C_{\mathtt{U}} > 0$ which could represent the penalty imposed by authorities for violation of quarantine norms. Accordingly, the static or instantaneous reward matrix is given by

$$R = \begin{bmatrix} -C_{\mathtt{P}} & 0 \\ -C_{\mathtt{P}} - C_{\mathtt{I}} & -C_{\mathtt{U}} - C_{\mathtt{I}} \end{bmatrix},$$

where the first (second) row represents the reward for an agent who is susceptible (infected) while the first (second) column represents the reward for an agent who adopts protection (remains unprotected). The stage reward of an agent in state $\mathtt{s}$ choosing an action a is denoted by $R[\mathtt{s}, a]$. The expected instantaneous reward of an agent under policy π is defined as $R_{\mathtt{eff}}[\mathtt{s}](\pi) = \sum_{a \in \mathcal{A}} \pi[a|\mathtt{s}] R[\mathtt{s}, a]$.

In a departure from previous works such as [7] which assume that agents are myopic and choose their actions to maximize their instantaneous rewards, we here assume that agents are forward-looking and consider the impact of their actions on the future state. In particular, agents seek to maximize discounted infinite horizon expected reward. To this end, we introduce the discount factor $\alpha \in [0, 1)$. At a given stationary social state (π, d), the value function for an agent belonging to the state $\mathtt{s}$ satisfies

$$V[\mathtt{s}](\pi, d) = R_{\mathtt{eff}}[\mathtt{s}](\pi) + \alpha \sum_{\mathtt{s}^+ \in \mathtt{Z}} P_{\mathtt{eff}}[\mathtt{s}^+|\mathtt{s}](\pi, d) V[\mathtt{s}^+](\pi, d), \quad (4)$$

following Bellman's principle of optimality. In order to define the best response of an agent to a social state (π, d), we define the *single-stage deviation value function* for an agent in state $\mathtt{s}$ choosing action a for the present time step and subsequently following the homogeneous policy π as

$$Q[\mathtt{s}, a](\pi, d) = R[\mathtt{s}, a] + \alpha \sum_{\mathtt{s}^+ \in \mathtt{Z}} \mathbb{P}[\mathtt{s}^+|\mathtt{s}, a](\pi, d) V[\mathtt{s}^+](\pi, d). \quad (5)$$

Thus, while computing Q, the agent is fully aware of the immediate reward and the effect of its present action on the future state, and chooses its action to maximize Q.

D. Best Response and Stationary Nash Equilibrium

We now define the best response map and stationary Nash equilibrium in this setting.

Definition 2.3 (Best Response): The best response map is a set-valued correspondence from the space of social states to the set of policies $B : \Pi \times \mathcal{D} \rightrightarrows \Pi = \Delta(\mathcal{A})^{|\mathtt{Z}|}$ given by

$$B(\pi, d) := \{\{\sigma_{\mathtt{s}} \in \Delta(\mathcal{A})\}_{\mathtt{s} \in \mathtt{Z}} | \sum_{a \in \mathcal{A}} (\sigma_{\mathtt{s}}[a] - \sigma'_{\mathtt{s}}[a]) Q[\mathtt{s}, a](\pi, d) \geq 0, \forall \sigma'_{\mathtt{s}} \in \Delta(\mathcal{A}), \mathtt{s} \in \mathtt{Z}\}.$$

In other words, at a given social state (π, d), a randomized strategy that maximizes the expected single-stage deviation value (5) is a best response. As the set of actions is finite (two) in our setting, a best response always exists, and the correspondence B is guaranteed to be non-empty. The above definition is analogous to the single-stage deviation principle studied in stochastic games literature [15]. We now define the notion of stationary Nash equilibrium.

Definition 2.4 (Stationary Nash Equilibrium [14]): A social state $(\pi^*, d^*) \in \Pi \times \mathcal{D}$ is a stationary Nash equilibrium if it satisfies

$$\pi^* \in B(\pi^*, d^*),$$
$$d^* = P_{\mathtt{eff}}(\pi^*, d^*)^\top d^*,$$

where $B(\pi^*, d^*)$ is the best response map defined in Definition 2.3 and $P_{\mathtt{eff}}(\pi^*, d^*)$ is defined in (1).

Thus, at this stationary Nash equilibrium, the state distribution d^* is invariant under the time-homogeneous stochastic matrix $P_{\mathtt{eff}}(\pi^*, d^*)$, and π^* is a best response to the equilibrium social state. Since the game-theoretic setting that we have considered has a finite number of states (i.e. two) and actions (i.e. two), and the state transition and reward functions are continuous in the social state, it follows from [14] that a stationary Nash equilibrium is guaranteed to exist.

III. Analysis of Value Function

In this section, we derive several intermediate results pertaining to the value function V and the single-stage deviation value function Q for each of the individual states. Since the best response is defined in terms of the Q function, it is necessary to analyze the difference in the Q function values to determine optimal or equilibrium policy and state distribution.

First, we consider the infected state $\mathtt{I}$. If agents follow policy π, then we have

$$R_{\mathtt{eff}}[\mathtt{I}](\pi) = -\pi[\mathtt{P}|\mathtt{I}] C_{\mathtt{P}} - \pi[\mathtt{U}|\mathtt{I}] C_{\mathtt{U}} - C_{\mathtt{I}}, \quad (6)$$
$$V[\mathtt{I}] = R_{\mathtt{eff}}[\mathtt{I}] + \alpha(\delta V[\mathtt{S}] + (1 - \delta) V[\mathtt{I}]), \quad (7)$$

where the dependence of the value function on the social state (π, d) is suppressed for better readability. The single-stage deviation values upon adopting protection and remaining unprotected are respectively given by:

$$Q[\mathtt{I}, \mathtt{P}] = -C_{\mathtt{P}} - C_{\mathtt{I}} + \alpha(\delta V[\mathtt{S}] + (1 - \delta) V[\mathtt{I}]), \quad (8)$$
$$Q[\mathtt{I}, \mathtt{U}] = -C_{\mathtt{U}} - C_{\mathtt{I}} + \alpha(\delta V[\mathtt{S}] + (1 - \delta) V[\mathtt{I}]), \quad (9)$$
$$\Rightarrow Q[\mathtt{I}, \mathtt{P}] - Q[\mathtt{I}, \mathtt{U}] = C_{\mathtt{U}} - C_{\mathtt{P}}. \quad (10)$$

In other words, the best response of an infected agent depends only on the difference between the penalty of avoiding quarantine and the cost of adopting protection, irrespective of the social state and discount factor. Consequently, we have the following result on the best response and equilibrium policy of an infected agent.

Lemma 3.1: Let π^* denote the policy at a stationary Nash equilibrium. If $C_{\mathtt{P}} < C_{\mathtt{U}}$, we have $\pi^*[\mathtt{P}|\mathtt{I}] = 1$ and $\pi^*[\mathtt{U}|\mathtt{I}] = 0$. Similarly, if $C_{\mathtt{P}} > C_{\mathtt{U}}$, we have $\pi^*[\mathtt{P}|\mathtt{I}] = 0$ and $\pi^*[\mathtt{U}|\mathtt{I}] = 1$.

We now consider the behavior of agents in the susceptible state $\mathtt{S}$. When agents follow policy π at state distribution d, the expected stage reward and the value function are

$$R_{\mathtt{eff}}[\mathtt{S}] = -\pi[\mathtt{P}|\mathtt{S}]C_{\mathtt{P}}, \tag{11}$$
$$V[\mathtt{S}] = R_{\mathtt{eff}}[\mathtt{S}] + \alpha(V[\mathtt{S}] + \beta_{\pi,\mathtt{eff}} d[\mathtt{I}](V[\mathtt{I}] - V[\mathtt{S}])), \tag{12}$$

where the dependence on the social state is omitted for better readability. Accordingly, the single-stage deviation value functions for a susceptible agent are given by

$$Q[\mathtt{S},\mathtt{P}] = -C_{\mathtt{P}} + \alpha((1-\gamma\beta_\pi d[\mathtt{I}])V[\mathtt{S}] + \gamma\beta_\pi d[\mathtt{I}]V[\mathtt{I}]), \tag{13}$$
$$Q[\mathtt{S},\mathtt{U}] = \alpha((1-\beta_\pi d[\mathtt{I}])V[\mathtt{S}] + \beta_\pi d[\mathtt{I}]V[\mathtt{I}]). \tag{14}$$

The difference is computed as

$$Q[\mathtt{S},\mathtt{P}] - Q[\mathtt{S},\mathtt{U}] = -C_{\mathtt{P}} + \alpha\beta_\pi d[\mathtt{I}](1-\gamma)(V[\mathtt{S}] - V[\mathtt{I}]), \tag{15}$$

where the right-hand side depends on the difference in the value functions associated with infected and susceptible states. The following lemma derives a useful expression on this quantity when the state distribution corresponds to the endemic equilibrium under a given policy.

Lemma 3.2: Suppose agents follow a given policy π, and the state distribution is at the corresponding endemic equilibrium d^*_π with infected proportion given by $\mathtt{I}^*_{\mathtt{ee},\pi}$. Then,

$$V[\mathtt{S}](\pi, d^*_\pi) - V[\mathtt{I}](\pi, d^*_\pi) = \frac{R_{\mathtt{eff}}[\mathtt{S}](\pi) - R_{\mathtt{eff}}[\mathtt{I}](\pi)}{1-\alpha(1-\beta_{\pi,\mathtt{eff}})}.$$

Proof: Subtracting (7) from (12), and rearranging the terms, we obtain

$$\begin{aligned} V[\mathtt{S}](\pi, d) - V[\mathtt{I}](\pi, d) &= \frac{R_{\mathtt{eff}}[\mathtt{S}](\pi) - R_{\mathtt{eff}}[\mathtt{I}](\pi)}{1-\alpha(1-\delta-\beta_{\pi,\mathtt{eff}} d[\mathtt{I}])} \\ &= \frac{R_{\mathtt{eff}}[\mathtt{S}](\pi) - R_{\mathtt{eff}}[\mathtt{I}](\pi)}{1-\alpha(1-\beta_{\pi,\mathtt{eff}})}, \end{aligned}$$

where at the second equality, we have used the fact that at $d = d^*_\pi$, we have $d[\mathtt{I}] = \mathtt{I}^*_{\mathtt{ee},\pi} = 1 - \frac{\delta}{\beta_{\pi,\mathtt{eff}}}$ following Lemma 2.1, which implies $\delta + \beta_{\pi,\mathtt{eff}}\mathtt{I}^*_{\mathtt{ee},\pi} = \beta_{\pi,\mathtt{eff}}$. ■

IV. Characterization of Stationary Nash Equilibrium

In this section, we present our main findings regarding the characterization of stationary Nash equilibrium in different parameter regimes. We first tackle the case where $C_{\mathtt{P}} < C_{\mathtt{U}}$. We first introduce several quantities of interest as

$$\begin{aligned} C^1_{\mathtt{P},\max} &= \frac{\alpha C_{\mathtt{I}}(\beta_{\mathtt{P}}-\delta)(1-\gamma)}{1-\alpha(1-\beta_{\mathtt{P}}+(\beta_{\mathtt{P}}-\delta)(1-\gamma))}, \\ C^1_{\mathtt{P},\min} &= \frac{\alpha C_{\mathtt{I}}(\gamma\beta_{\mathtt{P}}-\delta)(1-\gamma)}{\gamma(1-\alpha(1-\gamma\beta_{\mathtt{P}}))}, \\ \mathtt{I}^\dagger &= \frac{C_{\mathtt{P}}[1-\alpha(1-\delta)]}{\alpha\beta_{\mathtt{P}}[C_{\mathtt{I}}(1-\gamma) - C_{\mathtt{P}}\gamma]}, \quad (16) \\ x^\star &= \frac{1}{1-\gamma}\Big[1 - \frac{\delta}{\beta_{\mathtt{P}}(1-\mathtt{I}^\dagger)}\Big]. \quad (17) \end{aligned}$$

We are now ready to establish our main result.

Theorem 4.1: Suppose Assumption 2.2 holds. When $C_{\mathtt{P}} < C_{\mathtt{U}}$, we have the following characterization of the stationary Nash equilibrium.

- policy $\pi^*[\mathtt{P}|\mathtt{I}] = 1, \pi^*[\mathtt{P}|\mathtt{S}] = 0$, and infected proportion $\mathtt{I}^*_{\mathtt{ee},\pi} = 1 - \dfrac{\delta}{\beta_{\mathtt{P}}}$, if and only if $C_{\mathtt{P}} \geq C^1_{\mathtt{P},\max}$;
- $\pi^*[\mathtt{P}|\mathtt{I}] = 1, \pi^*[\mathtt{P}|\mathtt{S}] = 1$, and $\mathtt{I}^*_{\mathtt{ee},\pi} = 1 - \dfrac{\delta}{\gamma\beta_{\mathtt{P}}}$, if and only if $C_{\mathtt{P}} \leq C^1_{\mathtt{P},\min}$; and
- $\pi^*[\mathtt{P}|\mathtt{I}] = 1, \pi^*[\mathtt{P}|\mathtt{S}] = x^\star$, and $\mathtt{I}^*_{\mathtt{ee},\pi} = \mathtt{I}^\dagger$, if and only if $C^1_{\mathtt{P},\min} < C_{\mathtt{P}} < C^1_{\mathtt{P},\max}$.

Proof: When $C_{\mathtt{P}} < C_{\mathtt{U}}$, it follows from Lemma 3.1 that the optimal policy for the infected agents is given by $\pi^*[\mathtt{P}|\mathtt{I}] = 1$ and $\pi^*[\mathtt{U}|\mathtt{I}] = 0$ at the stationary Nash equilibrium. Consequently, $\beta_\pi = \beta_{\mathtt{P}}$. Let $\pi^*[\mathtt{P}|\mathtt{S}] = x$ for some $x \in [0,1]$ at the stationary Nash equilibrium. Then, $\pi^*[\mathtt{U}|\mathtt{S}] = 1-x$ and $\beta_{\pi,\mathtt{eff}}(x) = \beta_{\mathtt{P}}(1-x+\gamma x) \in [0,1]$. Furthermore, the proportion of infected agents at the endemic equilibrium induced by this policy is given by

$$\mathtt{I}^\star(x) = 1 - \frac{\delta}{\beta_{\pi,\mathtt{eff}}(x)} = 1 - \frac{\delta}{\beta_{\mathtt{P}}(1-x+\gamma x)}, \tag{18}$$

which is strictly decreasing in x for $x \in [0,1]$.

Following (15) and Lemma 3.2, we compute the difference in the Q function values at the Nash equilibrium as

$$\begin{aligned} & Q[\mathtt{S},\mathtt{P}] - Q[\mathtt{S},\mathtt{U}] \\ &\quad = -C_{\mathtt{P}} + \alpha\beta_{\mathtt{P}}\mathtt{I}^\star(x)(1-\gamma)(V[\mathtt{S}] - V[\mathtt{I}]) \\ &\quad = -C_{\mathtt{P}} + \alpha\beta_{\mathtt{P}}\mathtt{I}^\star(x)(1-\gamma)\frac{C_{\mathtt{P}}(1-x) + C_{\mathtt{I}}}{1-\alpha(1-\beta_{\pi,\mathtt{eff}}(x))} \\ \Rightarrow & (1-\alpha(1-\beta_{\pi,\mathtt{eff}}(x)))(Q[\mathtt{S},\mathtt{P}] - Q[\mathtt{S},\mathtt{U}]) \\ &\quad = C_{\mathtt{I}}\alpha\beta_{\mathtt{P}}(1-\gamma)\mathtt{I}^\star(x) + C_{\mathtt{P}}[-1+\alpha-\alpha\beta_{\pi,\mathtt{eff}}(x) \\ &\qquad + \alpha\beta_{\mathtt{P}}(1-x)(1-\gamma)\mathtt{I}^\star(x)] \\ &\quad = C_{\mathtt{I}}\alpha\beta_{\mathtt{P}}(1-\gamma)\mathtt{I}^\star(x) \\ &\qquad + C_{\mathtt{P}}[-1+\alpha-\alpha\delta-\alpha\beta_{\mathtt{P}}\gamma\mathtt{I}^\star(x)] \\ &\quad = \mathtt{I}^\star(x)\alpha\beta_{\mathtt{P}}[C_{\mathtt{I}}(1-\gamma) - C_{\mathtt{P}}\gamma] - C_{\mathtt{P}}[1-\alpha(1-\delta)], \end{aligned} \tag{19}$$

where in the second last step, we have used the fact that $\beta_{\pi,\mathtt{eff}}(x)\mathtt{I}^\star(x) = \beta_{\pi,\mathtt{eff}}(x) - \delta$.

We now examine the following cases.

Case 1: $C_{\mathtt{P}} \geq C^1_{\mathtt{P},\max}$. With some algebraic calculations, it can be shown that $C^1_{\mathtt{P},\max} < \frac{C_{\mathtt{I}}(1-\gamma)}{\gamma}$. Therefore, we consider two sub-cases. First, we consider $C^1_{\mathtt{P},\max}\gamma < C_{\mathtt{I}}(1-\gamma) \leq C_{\mathtt{P}}\gamma$. In this regime, the R.H.S. of (19) is strictly less than 0. Consequently, we have $Q[\mathtt{S},\mathtt{P}] < Q[\mathtt{S},\mathtt{U}]$ and $x = \pi^*[\mathtt{P}|\mathtt{S}] = 0$ is the only possible best response. Consequently, we have $\mathtt{I}^\star(0) = 1 - \frac{\delta}{\beta_{\mathtt{P}}}$ at the equilibrium.

We now consider $C^1_{\mathtt{P},\max}\gamma \leq C_{\mathtt{P}}\gamma < C_{\mathtt{I}}(1-\gamma)$. In this regime, the R.H.S. of (19) is strictly decreasing since $\mathtt{I}^\star(x)$ is a strictly decreasing function of x. As a result, if the R.H.S. is non-positive (non-negative) at $x = 0$ ($x = 1$), it stays negative (positive) over the entire range $x \in (0,1]$ ($x \in [0,1)$). We obtain conditions for the R.H.S. of (19) to be non-positive as

$$\begin{aligned} & \mathtt{I}^\star(0)\alpha\beta_{\mathtt{P}}[C_{\mathtt{I}}(1-\gamma) - C_{\mathtt{P}}\gamma] - C_{\mathtt{P}}[1-\alpha(1-\delta)] \leq 0 \\ \Leftrightarrow & (\beta_{\mathtt{P}}-\delta)\alpha C_{\mathtt{I}}(1-\gamma) \leq C_{\mathtt{P}}[(\beta_{\mathtt{P}}-\delta)\alpha\gamma + 1 - \alpha(1-\delta)] \\ \Leftrightarrow & C_{\mathtt{P}} \geq \frac{C_{\mathtt{I}}\alpha(\beta_{\mathtt{P}}-\delta)(1-\gamma)}{1-\alpha(1-\delta-\gamma(\beta_{\mathtt{P}}-\delta))} =: C^1_{\mathtt{P},\max}. \end{aligned}$$

In other words, $Q[\texttt{S},\texttt{P}] < Q[\texttt{S},\texttt{U}]$ for all $x \in (0,1]$ (with possible equality at $x=0$) if and only if $C_{\texttt{P}} \geq C^1_{\texttt{P},\max}$. As a result, $\pi^*[\texttt{P}|\texttt{S}] = 0$ is the only possible best response at which $\texttt{I}^\star(0) = 1 - \frac{\delta}{\beta_{\texttt{P}}}$.

Case 2: $C_{\texttt{P}} \leq C^1_{\texttt{P},\min}$. In this regime, the term multiplying $\texttt{I}^\star(x)$ in (19) is positive and following similar arguments as Case 1, we obtain conditions for the R.H.S. of (19) to be non-negative at $x=1$ as

$$\begin{aligned}&\texttt{I}^\star(1)\alpha\beta_{\texttt{P}}[C_{\texttt{I}}(1-\gamma) - C_{\texttt{P}}\gamma] - C_{\texttt{P}}[1-\alpha(1-\delta)] \geq 0\\ \Leftrightarrow\ &(\gamma\beta_{\texttt{P}} - \delta)\alpha C_{\texttt{I}}(1-\gamma) \geq C_{\texttt{P}}\gamma[(\gamma\beta_{\texttt{P}} - \delta)\alpha + 1 - \alpha(1-\delta)]\\ \Leftrightarrow\ &C_{\texttt{P}} \leq \frac{C_{\texttt{I}}\alpha(\gamma\beta_{\texttt{P}} - \delta)(1-\gamma)}{\gamma(1-\alpha(1-\gamma\beta_{\texttt{P}}))} =: C^1_{\texttt{P},\min}.\end{aligned}$$

Thus, $C_{\texttt{P}} \leq C^1_{\texttt{P},\min}$ is necessary and sufficient for $x = \pi^*[\texttt{P}|\texttt{S}] = 1$ to be the policy at a stationary Nash equilibrium, and at $x=1$, $\texttt{I}^\star(1) = 1 - \frac{\delta}{\gamma\beta_{\texttt{P}}}$.

Case 3: $C^1_{\texttt{P},\min} < C_{\texttt{P}} < C^1_{\texttt{P},\max}$. In this regime, the term multiplying $\texttt{I}^\star(x)$ in (19) is positive as before. Additionally, the R.H.S. of (19) is positive at $x=0$ and negative at $x=1$. Consequently, there is a unique $x^\star \in [0,1]$ at which the R.H.S. is equal to zero, or equivalently, $Q[\texttt{S},\texttt{P}] = Q[\texttt{S},\texttt{U}]$. For any $x < x^\star$, we have $Q[\texttt{S},\texttt{P}] > Q[\texttt{S},\texttt{U}]$ leading to the best response of an agent being $\pi[\texttt{P}|\texttt{S}] = 1$, and similarly, for $x > x^\star$, the best response is $\pi[\texttt{P}|\texttt{S}] = 0$. Thus, we must have $x = x^\star$ and $Q[\texttt{S},\texttt{P}] = Q[\texttt{S},\texttt{U}]$ at the stationary Nash equilibrium. By setting the R.H.S. of (19) to zero, we obtain $\texttt{I}^\star(x^\star) = \texttt{I}^\dagger$ and $x^\star$ is obtained using (16) and (17). ■

Remark 4.2: The above theorem shows that adopting protection is preferred when the associated cost $C_{\texttt{P}}$ is sufficiently small. Further, both $C^1_{\texttt{P},\min}$ and $C^1_{\texttt{P},\max}$ are monotonically increasing in the discount factor α. Thus, for a given $C_{\texttt{P}}$, as individuals become more forward-looking, they are more likely to adopt protection. Thus, non-myopic behavior plays a key role in individuals adopting protection in a strategic setting. As protection becomes more effective (i.e., as γ decreases), $C^1_{\texttt{P},\max}$ increases, which leads to adopting protection becoming a more attractive choice at the equilibrium. Finally, as $C_{\texttt{P}}$ increases from $C^1_{\texttt{P},\min}$ to $C^1_{\texttt{P},\max}$, the policy of adopting protection $\pi^*[\texttt{P}|\texttt{S}]$ decreases monotonically from 1 to 0 and the infected proportion $I^\dagger$ increases monotonically from $1 - \frac{\delta}{\gamma\beta_{\texttt{P}}}$ to $1 - \frac{\delta}{\beta_{\texttt{P}}}$. The above observations provide valuable insights to the policymakers on how to choose the values of $C_{\texttt{P}}$ or $C_{\texttt{U}}$ as a function of epidemic parameters, the effectiveness of protection and degree of non-myopic behavior to induce a desired level of infection at the equilibrium.

Remark 4.3: While the above result holds under Assumption 2.2, we briefly discuss the consequence of relaxing it. When $\delta > \beta_{\texttt{P}}$, we have $C^1_{\texttt{P},\max} < 0$, and thus, for any $C_{\texttt{P}} > 0$, we have $C_{\texttt{P}} > C^1_{\texttt{P},\max}$. It can be easily shown that at a stationary Nash equilibrium, $\pi^*[\texttt{P}|\texttt{S}] = 0$ (similar to Case 1 of Theorem 4.1) and infected proportion $\texttt{I}^*_{\texttt{dfe},\pi} = 0$. Similarly, when $\gamma\beta_{\texttt{P}} < \delta < \beta_{\texttt{P}}$, the protection adoption behavior would be governed in a similar manner as Case 1 and 3 of Theorem 4.1. The complete derivation is omitted due to space constraints.

Remark 4.4: The equilibria derived in Theorem 4.1 have a similar flavour as the results obtained in [7] for the myopic setting. However, there does not exist a specific value of the loss parameter L in [7] at which the results obtained in Theorem 4.1 would map exactly to the results obtained in [7]. Further connections between the myopic and non-myopic settings will be explored in a follow-up work.

We now tackle the case where $C_{\texttt{P}} > C_{\texttt{U}}$. We define the following quantities of interest.

$$\begin{aligned}C^2_{\texttt{P},\max} &= \frac{\alpha(\beta_{\texttt{U}} - \delta)(1-\gamma)(C_{\texttt{U}} + C_{\texttt{I}})}{1 - \alpha + \alpha\beta_{\texttt{U}}},\\ C^2_{\texttt{P},\min} &= \frac{\alpha(\gamma\beta_{\texttt{U}} - \delta)(1-\gamma)(C_{\texttt{U}} + C_{\texttt{I}})}{\alpha(\gamma\beta_{\texttt{U}} - \delta) + \gamma(1-\alpha) + \gamma\alpha\delta},\\ \texttt{I}^\ddagger &= \frac{C_{\texttt{P}}[1-\alpha(1-\delta)]}{\alpha\beta_{\texttt{U}}[(1-\gamma)(C_{\texttt{I}} + C_{\texttt{U}}) - C_{\texttt{P}}]}, \qquad (20)\\ z^\star &= \frac{1}{1-\gamma}\Big[1 - \frac{\delta}{\beta_{\texttt{U}}(1-\texttt{I}^\ddagger)}\Big]. \qquad (21)\end{aligned}$$

Our result is stated below.

Theorem 4.5: When $C_{\texttt{P}} > C_{\texttt{U}}$, we have the following characterization of the stationary Nash equilibrium.

- $\pi^*[\texttt{P}|\texttt{I}] = 0, \pi^*[\texttt{P}|\texttt{S}] = 0$, and $\texttt{I}^*_{\texttt{ee},\pi} = 1 - \dfrac{\delta}{\beta_{\texttt{U}}}$, if and only if $C_{\texttt{P}} \geq C^2_{\texttt{P},\max}$;
- $\pi^*[\texttt{P}|\texttt{I}] = 0, \pi^*[\texttt{P}|\texttt{S}] = 1$, and $\texttt{I}^*_{\texttt{ee},\pi} = 1 - \dfrac{\delta}{\gamma\beta_{\texttt{U}}}$, if and only if $C_{\texttt{P}} \leq C^2_{\texttt{P},\min}$; and
- $\pi^*[\texttt{P}|\texttt{I}] = 0, \pi^*[\texttt{P}|\texttt{S}] = z^\star$, and $\texttt{I}^*_{\texttt{ee},\pi} = \texttt{I}^\ddagger$, if and only if $C^2_{\texttt{P},\min} < C_{\texttt{P}} < C^2_{\texttt{P},\max}$.

The proof follows from analogous arguments as the proof of Theorem 4.1, and is omitted in the interest of space.

When agents are myopic, i.e., the discount factor $\alpha = 0$, then $C^1_{\texttt{P},\max} = C^2_{\texttt{P},\max} = 0$, and only the first sub-case of both the theorems are operative.

V. Numerical Results

We now illustrate our findings using numerical simulations. For brevity, we focus on the case $C_{\texttt{P}} < C_{\texttt{U}}$. We use the following parameter values in our simulations.

α	$\beta_{\texttt{P}}$	$\beta_{\texttt{U}}$	γ	δ	$d_0[\texttt{I}]$
0.9	0.6	0.7	0.3	0.05	0.1

Here, $d_0[\texttt{I}]$ denotes the probability of an individual being infected at time $k=0$.

We select the costs $C_{\texttt{U}} = 10$ and $C_{\texttt{I}} = 5$, which yields $C^1_{\texttt{P},\max} = 5.9029$ and $C^1_{\texttt{P},\min} = 5.2099$. Accordingly, we select two different values of $C_{\texttt{P}}$ consistent with the last two cases in Theorem 4.1 as 3 and 5.75, respectively. The infection state distribution is updated according to (2).

We adopt the perturbed best response dynamics given by the *logit choice function* [2], [6] to update the policy (for every state action pair) as

$$\pi_{k+1}[a|\texttt{s}] = \frac{\exp(\lambda Q[\texttt{s},a](\pi_k,d_k))}{\sum_{a'}\exp(\lambda Q[\texttt{s},a'](\pi_k,d_k))}.$$

In other words, the policy is updated such that the action with a larger Q function value is chosen with a higher probability. The degree of preference towards the action associated with the higher Q function value is controlled by a parameter $\lambda > 0$ which also captures the degree of bounded rationality

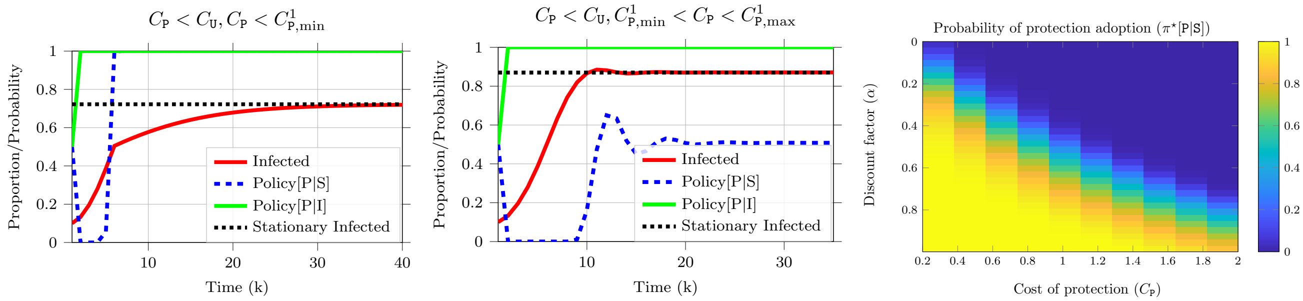

Fig. 1. Evolution and convergence of infected proportion and policy of adopting protection when $C_{\mathtt{P}} < C_{\mathtt{U}}$ and $\lambda = 20$ (left and middle panels), and protection adoption policy at the stationary Nash equilibrium for the range of α from 0 to 0.99 and the range of $C_{\mathtt{P}}$ from 0.2 to 2 (right panel).

associated with human decision-making [2], [6]. We omit further discussion along these lines in the interest of space.

The plots in the left and the middle panels of Fig. 1 show that the evolution of system trajectories is consistent with Theorem 4.1. Interestingly, the evolution of $d_k[\mathtt{I}]$ in the left panel of Fig. 1 is monotonic, while they are oscillatory in the middle panel. During the transient phase, when the fraction of the infected population is low, $Q[\mathtt{S},\mathtt{P}] < Q[\mathtt{S},\mathtt{U}]$ and the behavioral policy adopts protection with a small probability, which increases the fraction of infected population. When the fraction of the infected population is sufficiently high, then $Q[\mathtt{S},\mathtt{P}] > Q[\mathtt{S},\mathtt{U}]$ and the behavioral policy adopts protection with higher probability, which decreases the fraction of the infected population. These interleaving processes lead to the observed oscillatory behavior.

In the right panel of Fig. 1, we explore the influence of parameters α (which determines the extent to which an individual is forward-looking) and $C_{\mathtt{P}}$ (cost of adopting protection) on the policy of susceptible agents adopting protection at equilibrium. We have used the following parameters for this study.

$\beta_{\mathtt{P}}$	$\beta_{\mathtt{U}}$	γ	δ	$C_{\mathtt{I}}$	$C_{\mathtt{U}}$	λ
0.6	0.8	0.6	0.28	12	10	15

For a fixed cost of protection $C_{\mathtt{P}}$, increasing α which leads to more forward-looking behavior, results in increasing adoption of the protection at the stationary Nash equilibrium. Similarly, for a given value of α, increasing the cost of protection $C_{\mathtt{P}}$ results in decreasing adoption of the protection in accordance with the observations stated in Remark 4.2.

VI. Conclusions

We studied the influence of strategic non-myopic protection adoption policies on the spread of the SIS epidemic using a dynamic population game framework. We rigorously characterized the protection adoption policy and the associated fraction of the infected population at the Nash equilibrium under different parameter regimes. We numerically illustrated the transient response properties by using perturbed best response dynamics, and showed that system trajectories are monotonic in certain parameter regimes while they are oscillatory in some parameter regimes. Finally, we illustrated the influence of the degree of non-myopia and the cost of protection on the behavioral policy. There are several promising directions for future research of the proposed framework with non-myopic agents (both in the context of epidemics and other applications), including shaping the reward matrix to induce desired equilibrium behavior, analysis of agents with heterogeneous activity patterns and establishing convergence of evolutionary learning dynamics.

References

[1] Y. Huang and Q. Zhu, “Game-theoretic frameworks for epidemic spreading and human decision-making: A review,” *Dynamic Games and Applications*, vol. 12, pp. 7–48, 2022.

[2] W. H. Sandholm, *Population Games and Evolutionary Dynamics*. MIT press, 2010.

[3] A. Roy, C. Singh, and Y. Narahari, “Recent advances in modeling and control of epidemics using a mean field approach,” *arXiv preprint arXiv:2208.14765*, 2022.

[4] G. Theodorakopoulos, J.-Y. Le Boudec, and J. S. Baras, “Selfish response to epidemic propagation,” *IEEE Transactions on Automatic Control*, vol. 58, no. 2, pp. 363–376, 2013.

[5] C. Eksin, J. S. Shamma, and J. S. Weitz, “Disease dynamics on a network game: A little empathy goes a long way,” *Scientific Reports*, vol. 7, p. 44122, 2017.

[6] A. R. Hota, T. Sneh, and K. Gupta, “Impacts of game-theoretic activation on epidemic spread over dynamical networks,” *SIAM Journal on Control and Optimization*, pp. S92–S118, 2022, Preprint: https://arxiv.org/abs/2011.00445.

[7] A. Satapathi, N. K. Dhar, A. R. Hota, and V. Srivastava, “Epidemic propagation under evolutionary behavioral dynamics: Stability and bifurcation analysis,” in *American Control Conference*, 2022, pp. 3662–3667.

[8] ——, “Coupled evolutionary behavioral and disease dynamics under reinfection risk,” *arXiv preprint arXiv:2209.07348*, 2022.

[9] N. C. Martins, J. Certorio, and R. J. La, “Epidemic population games and evolutionary dynamics,” *arXiv preprint arXiv:2201.10529*, 2022.

[10] H. Khazaei, K. Paarporn, A. Garcia, and C. Eksin, “Disease spread coupled with evolutionary social distancing dynamics can lead to growing oscillations,” *IEEE Conference on Decision and Control*, pp. 4280 – 4286, 2021.

[11] S. Liu, Y. Zhao, and Q. Zhu, “Herd behaviors in epidemics: A dynamics-coupled evolutionary games approach,” *Dynamic Games and Applications*, vol. 12, no. 1, pp. 183–213, 2022.

[12] K. Frieswijk, L. Zino, M. Ye, A. Rizzo, and M. Cao, “A mean-field analysis of a network behavioral–epidemic model,” *IEEE Control Systems Letters*, vol. 6, pp. 2533–2538, 2022.

[13] E. Elokda, S. Bolognani, and A. R. Hota, “A dynamic population model of strategic interaction and migration under epidemic risk,” in *IEEE Conference on Decision and Control*, 2021, pp. 2085–2091.

[14] E. Elokda, A. Censi, and S. Bolognani, “Dynamic population games,” *arXiv preprint arXiv:2104.14662*, 2021.

[15] J. Filar and K. Vrieze, *Competitive Markov Decision Processes*. Springer Science & Business Media, 2012.