# Epidemic Population Games And Perturbed Best Response Dynamics

Shinkyu Park, *Member, IEEE*, Jair Certorio, *Student Member, IEEE*, Nuno C. Martins, *Senior Member, IEEE*, Richard J. La

*Abstract*— **This paper proposes an approach to mitigate epidemic spread in a population of strategic agents by encouraging safer behaviors through carefully designed rewards. These rewards, which adapt to the evolving state of the epidemic, are ascribed by a dynamic payoff mechanism we seek to design. We use a modified SIRS model to track how the epidemic progresses in response to the agents' strategic choices. By employing perturbed best response evolutionary dynamics to model the population's strategic behavior, we extend previous related work so as to allow for noise in the agents' perceptions of the rewards and intrinsic costs of the available strategies. Central to our approach is the use of system-theoretic methods and passivity concepts to obtain a Lyapunov function, ensuring the global asymptotic stability of an endemic equilibrium with minimized infection prevalence under budget constraints. We leverage the Lyapunov function to analyze how the epidemic's spread rate is influenced by the time scale of the payoff mechanism's dynamics. Additionally, we derive anytime upper bounds on both the infectious fraction of the population and the instantaneous cost a social planner must incur to control the spread, allowing us to quantify the trade-off between peak infection prevalence and the corresponding cost. For a class of one-parameter perturbed best response models, we propose a method to learn the model's parameter from data.**

*Index Terms*— **Epidemic; Evolutionary Dynamics; Lyapunov Stability; Population Games.**

## I. Introduction

Recent studies [1], [2] have used system-theoretic methods to obtain dynamic rewards promoting safer behaviors during an epidemic, aiming to lessen infection prevalence in a large population. The results in [1], [2] apply to a wide range of strategic decision-making behaviors but assume that the agents have precise knowledge of the rewards and costs of the available strategies. Unlike these studies, here we accept that the agents' choices are based on rough estimates of the strategies' net rewards (rewards minus intrinsic costs for adopting the strategies), leading us to explore the so-called *Perturbed Best Response (PBR) rules* to capture the unavoidable noise in the agents' perceptions.

Our work stands on two pillars: First, we use a modified susceptible-infectious-recovered-susceptible (SIRS) epidemic model [3] to track disease spread. The agents' strategic choices affect the infection rate denoted as $\mathcal{B}(t)$. The agents choose from a range of strategies $\{1, \cdots, n\}$, $n \geq 2$, each affecting $\mathcal{B}(t)$. The strategies' net rewards, calculated by rewards minus intrinsic costs, guide the decisions the agents make. This decision-making is described by an evolutionary dynamics model, which assumes that the agents can review and change their strategies over time according to a PBR rule [4], [5] (see §I-B). Second, we seek to design a payoff mechanism for assigning rewards that nudge the population's choices towards safer strategies reducing infection rates, with the constraint of keeping rewards affordable in the long run. The payoff mechanism's state is coupled to the epidemic model's state, creating what we call an *Epidemic Population Game (EPG)* (see §I-B).

The design methods proposed in [1], [2] are not suitable for our context, as they assume the absence of noise. While our approach also builds on system-theoretic passivity as in [1], [2], it requires substantial modifications to the stability proofs and the characterization of the set of endemic equilibria with minimized infection prevalence, which here depends on the noise distribution. For an important subset of PBR rules, which we will describe later, this distribution is governed by a single parameter that can be learned from data. Moreover, we establish local exponential stability for a representative class of PBR dynamics. Additionally, we explore the impact of the time scale governing the dynamics of the payoff mechanism on the upper bound of the infected population. By incorporating this time scale parameter into the design, it becomes an additional lever that a social planner can adjust to mitigate the spread of the endemic.

### A. Evolutionary Dynamics Model

Each agent follows one strategy at a time, which it can revise repeatedly. A payoff vector $p(t)$ in $\mathbb{R}^n$ whose entries quantify the net rewards of the available strategies influences the revision process. Typically, the agents are more likely to choose strategies with higher payoffs. Namely, we define

$$p(t) := r(t) - c(I(t)), \tag{1}$$

where $I(t)$ represents the infectious portion of the population, $c(I(t)) = (c_1(I(t)), \cdots, c_n(I(t)))$ is a vector whose $\ell$-th entry $c_\ell(I(t))$ is the inherent cost of the $\ell$-th strategy, which depends

continuously on $I(t)$ and decrease as $I(t)$ increases, and $r(t) = (r_1(t), \cdots, r_n(t))$ is a reward vector meant to incentivize the adoption of safer (costlier) strategies, where $r_\ell(t)$ is the $\ell$-th strategy's reward.

Rather than focusing on what each strategy may represent (see [1, Remark 1]), in our framework, we assume that a vector $\vec{\beta}$ in $\mathbb{R}_{>0}^n$ is given whose $\ell$-th entry $\vec{\beta}_\ell$ quantifies the effect of the $\ell$-th strategy towards $\mathcal{B}(t)$ according to

$$\mathcal{B}(t) = \vec{\beta}' x(t), \quad t \geq 0, \tag{2}$$

where $x(t)$ is the so-called *population state* taking values in the standard simplex $\mathbb{X}$ defined below and whose $\ell$-th entry $x_\ell(t)$ represents the proportion of the population adopting the $\ell$-th strategy at time $t$.

$$\mathbb{X} := \left\{ x \in [0,1]^n \mid \sum_{i=1}^n x_i = 1 \right\}.$$

Following the standard approach in [6, §4.1.2], the following *Evolutionary Dynamics Model (EDM)* governs the dynamics of $x(t)$ in the large-population limit:

$$\dot{x}(t) = \mathcal{V}(x(t), p(t)), \quad t \geq 0, \tag{EDMa}$$

where the $i$-th component of $\mathcal{V}$ is specified as

$$\mathcal{V}_i(x(t), p(t)) := \underbrace{- \sum_{j=1}^n x_i(t) \mathcal{T}_{ij}(x(t), p(t))}_{\text{outflow switching from strategy } i}$$
$$+ \underbrace{\sum_{j=1}^n x_j(t) \mathcal{T}_{ji}(x(t), p(t))}_{\text{inflow switching to strategy } i}. \tag{EDMb}$$

A Lipschitz continuous map $\mathcal{T} : \mathbb{X} \times \mathbb{R}^n \to [0,1]^{n \times n}$ is referred to as the learning rule (or revision protocol). Specifically, given the current state $x(t)$ and the payoff vector $p(t)$, $\mathcal{T}_{ji}(x(t), p(t))$ represents the probability that an agent will switch from its current strategy $j$ to strategy $i$ when given the opportunity. As discussed in [6, Part II] and [7, §13.3-13.5], which provide an in-depth exploration of various learning rules, the opportunity for strategy revision for each agent is modeled by an i.i.d. Poisson process.

### B. PBR Learning Rule and Epidemic Population Game

In contrast to earlier work in [1], [2], [8], here we adopt the class of PBR learning rules [4], [5] as a way to incorporate noise in the agents' perceived payoffs. Specifically, we consider

$$\mathcal{T}_{ji}(x, p) = C_i(p) := \mathbb{P}\left( p_i + v_i \geq \max_{1 \leq \ell \leq n} (p_\ell + v_\ell) \right),$$
$$\forall j \in \{1, \cdots, n\}, \tag{4}$$

where $v_1, \cdots, v_n$ are random variables admitting a positive joint probability density function over $\mathbb{R}^n$. In the following example, we discuss a learning rule that has been widely studied in the literature.

*Example 1:* The logit learning rule [5] is specified as

$$C_i(p) \underset{\text{logit}}{=} \frac{e^{p_i/\mu}}{\sum_{\ell=1}^n e^{p_\ell/\mu}}, \tag{5}$$

where $\mu > 0$ quantifies the noise intensity. Namely, $v_1, \cdots, v_n$ are i.i.d. random variables characterized by $\mathbb{P}(v_i \leq \zeta) =$

$e^{-e^{-\zeta/\mu - \bar{\zeta}}}$, where $\bar{\zeta}$ is Euler's constant. In the noise-free limit as $\mu$ tends to zero, $C$ will tend to the best response learning rule.

According to [4], we can represent the choice function as[1]

$$C(p) = (C_1(p), \cdots, C_n(p)) = \underset{z \in \text{int}(\mathbb{X})}{\arg\max}(z'p - Q(z)), \tag{6}$$

where $Q : \text{int}(\mathbb{X}) \to \mathbb{R}$ is called the *admissible payoff perturbation*, which is twice continuously differentiable and satisfies the following conditions:

$$\tilde{z}' \nabla^2 Q(z) \tilde{z} > 0, \ z \in \mathbb{X}, \tilde{z} \in T\mathbb{X} \setminus \{0\} \tag{7a}$$
$$\lim_{z_{\min} \to 0} \|\nabla Q(z)\|_2 = \infty, \quad \text{where } z_{\min} = \min_{1 \leq i \leq n} z_i, \tag{7b}$$

where $T\mathbb{X}$ is the tangent space of $\mathbb{X}$. The logit learning rule is obtained via the perturbation function $Q$ defined as $Q(x) = \mu \sum_{i=1}^n x_i \ln x_i$, with $\mu > 0$.

When (EDM) is defined by the choice function (6), we refer to it as the PBR EDM. Note that when the PBR EDM reaches its equilibrium state, i.e., $\mathcal{V}(x^*, p^*) = 0$, the population state $x^*$ satisfies $x^* = C(p^*)$.

We adopt the EPG formalism from [1] as follows.

$$\dot{I}(t) = \big(\mathcal{B}(t)(1 - I(t) - R(t)) - \sigma\big)I(t) \tag{EPGa}$$
$$\dot{R}(t) = \gamma I(t) - \omega R(t) \tag{EPGb}$$
$$\dot{q}(t) = G\big(I(t), R(t), x(t), q(t)\big) \tag{EPGc}$$
$$r(t) = H\big(I(t), R(t), x(t), q(t)\big), \tag{EPGd}$$

where $I(t)$, $R(t)$, and $S(t) := 1 - I(t) - R(t)$ take values in $[0,1]$ and represent the proportions of the population which are infectious, have recovered, and are susceptible to infection at time $t$, respectively. Here, (EPGa,b) is a normalized SIRS model with $\sigma := \gamma + \theta$ and $\omega := \psi + \theta$, where $\gamma$ and $\psi$ denote the daily recovery rate and the daily rate at which recovered individuals become susceptible (due to waning immunity), respectively, and $\theta$ is the daily birth rate. Additionally, (EPGc,d) is a payoff mechanism we seek to design, where $r(t)$ appears in (1) and $q(t)$ belongs to $\mathbb{R}^m$ with $m \geq 1$.

### C. Problem Formulation and Paper Structure

We order the entries of $\vec{\beta}$ and $c(I)$ as follows:

$$\vec{\beta}_i < \vec{\beta}_{i+1} \text{ and } c_i(I) > c_{i+1}(I), \ i \in \{1, \cdots, n-1\}, \ I \in [0,1].$$

Henceforth, $\vec{\beta}$ and $c : [0,1] \to \mathbb{R}^n$ satisfying the conditions above are assumed given. We will use $\tilde{c}(I)$ defined below to specify cost constraints because for a planner seeking to promote the $i$-th strategy it suffices to offer incentives to offset the differential $\tilde{c}_i(I)$.

$$\tilde{c}_i(I) := c_i(I) - c_n(I), \ i \in \{1, \cdots, n\}, \ I \in [0,1]. \tag{8}$$

*Definition 1:* Given a cost budget $c^* > 0$, the optimal endemic transmission rate $\beta^*$ is determined as the minimum

---

[1]Since $Q$ satisfies (7), according to the analysis presented in [4], the maximization in (6) admits a unique solution.

value of the following constrained optimization problem:

$$\min_{\beta \in (\sigma, \vec{\beta}_n), r \in \mathbb{R}^n_{\geq 0}, I \in [0,1]} \beta \qquad (9)$$

$$\text{subject to } r'C(r - \tilde{c}(I)) \leq c^*$$
$$\beta = \vec{\beta}'C(r - \tilde{c}(I))$$
$$I = \eta(1 - \sigma/\beta),$$

where $C$ is the choice function defined in (6).

Note that at the optimal solution $(\beta^*, r^*, I^*)$ for (9), $C(r^* - \tilde{c}(I^*))$ represents the population state at the equilibrium of the PBR EDM, where the infectious population $I^*$ is minimized while satisfying the budget constraint $r^{*'}C(r^* - \tilde{c}(I^*)) \leq c^*$. Additionally, we can infer that the optimal endemic transmission rate $\beta^*$ decreases as the cost budget $c^*$ increases, implying that with a higher $c^*$, the planner can further reduce $I^*$. Throughout the paper, we assume that an average transmission rate less than or equal to $\sigma$ is considered too costly for a social planner to sustain. Therefore, we focus on scenarios where the minimum $\beta^*$ is greater than $\sigma$.

**Main Problem:** We seek to obtain Lipschitz continuous $G$ and $H$ for which the following hold for any $I(0)$ in $(0, 1]$, $R(0)$ in $[0, 1 - I(0)]$, $x(0)$ in $\mathbb{X}$, and $q(0)$ in $\mathbb{R}^m$:[2]

$$\lim_{t \to \infty} (I, R, \mathcal{B})(t) = (I^*, R^*, \beta^*), \qquad \text{(P1)}$$

$$\limsup_{t \to \infty} r'(t)x(t) \leq c^*, \qquad \text{(P2)}$$

where, from Picard's Theorem, $\{(I, R, x, q)(t) \mid t \geq 0\}$ is the unique solution of the initial value problem for the closed-loop model formed by (EDM) and (EPG). Here, the nontrivial endemic equilibrium for (EPGa,b) is:

$$I^* := \eta(1 - \tfrac{\sigma}{\beta^*}), \ \ R^* := (1 - \eta)(1 - \tfrac{\sigma}{\beta^*}), \ \ \eta := \tfrac{\omega}{\omega + \gamma}.$$

We will seek $G$ and $H$ for which a Lyapunov function for the closed-loop model exists. We will do so not only to establish (P1) but, crucially, also to leverage the Lyapunov function to obtain anytime upper bounds for $I(t)$. This is relevant because, as has been pointed out in studies [9], [10] employing $\mathcal{B}(t)$ as a control variable, $I(t)$ tends to significantly overshoot its endemic equilibrium $I^*$ when $I(0) < I^*$, unless the control policy prevents it.

**Paper Structure:** After the introduction, in §II we outline motivation for our work and provide a brief overview of related work. The paper's main section is §III where we describe $G$ and $H$ that constitute a solution to our Main Problem and present Theorem 1 and Corollary 1 stating the precise conditions under which (P1)-(P2) are guaranteed globally for the chosen $G$ and $H$ along with a method to select key parameters. In §III, we also discuss convergence rate, present an example, and explain how to construct anytime upper bounds for the prevalence of infections and how to learn from data the noise parameter specifying a class of choice functions when it is unknown a priori. In §IV, we illustrate our methods via simulation for two scenarios.

---

[2]In our main results, specifically Corollary 1, the design of (EPGc) and (EPGd) guarantees that $r(t)$ converges to $r^*$, as defined by the social planner in (9). It is important to note that $r^*$ is a vector with non-negative elements.

## II. Motivation and Comparative Literature Review

In this work, we investigate agent decision-making in epidemic processes where the agents' strategy selection is determined by PBR models. Our study is motivated by prevalent empirical observations, as exemplified in [11], which indicate that the Nash equilibrium is not an accurate predictor of human decision-making processes. We provide a review of prior studies on epidemic models and PBR models.

There are a number of recent studies that investigated the problem of managing an epidemic using control theory: di Lauro et al. [12] and Sontag [10] studied the problem of identifying the optimal timing for non-pharmaceutical interventions (NPIs), e.g., quarantine and lockdowns, to minimize the peak infections. Al-Radhawi et al. [13] examined the problem of tuning NPIs to regulate infection rates as an adaptive control problem and investigated the stability of disease-free and endemic steady states. Godara et al. [9] studied the problem of controlling the infection rate to minimize the total cost till herd immunity is attained as an optimal control problem subject to a constraint on the fraction of infectious population. But, these studies did not consider the strategic decision-making.

Several recent studies employed game theory, including evolutionary or population games, to study epidemic processes with strategic agents [14]–[20]. For instance, [14] studied the effect of risk perception on whether individuals choose to self-quarantine or not, and how increased perceived risks could lead to multiple infection peaks. Khazaei et al. [20] adopted the SEIR epidemic model with the replicator dynamics to study the interplay between the underlying epidemic state and the behavioral response of a single population. They showed that as the disease prevalence changes over time, the level of public cooperation varies as well in response, which results in oscillations of infection level.

The dynamics of epidemic processes on networks have been studied extensively, e.g., [3], [21]–[24]. Other studies also considered epidemics with multiple populations [25], [26]; each population represents a group of similar agents, a community or a geographic area (e.g., a city). The interactions among populations are often modeled using a graph, in which edge weights indicate the contact or interactions rates across different populations. We refer an interested reader to [3], [21] for a comprehensive literature survey. Considering multiple populations complicates the analysis of epidemic models, even without modeling agents' strategic interactions and leads to richer dynamics [25]. The impact of asymptomatic infections over complex networks has also been studied, along with seasonal transmission rate changes, e.g., a high tourist season or the start of a new school year [27].

Kuniya and Muroya [26] studied a multi-group SIS model with population migration and established global convergence to the endemic equilibrium when the basic reproduction number exceeds one, and to the disease free equilibrium otherwise. These studies, however, assume fixed transmission rates that do not depend on the strategies chosen by the agents in different populations.

The authors of [1], [2] introduced a new framework that combines the strategic decision-making process of agents

(evolutionary dynamics) and a compartmental epidemic model (SIRS model) for a single population. This framework allowed them to design a dynamic payoff mechanism that ensures the convergence to an endemic equilibrium where the disease transmission rate is minimized subject to a budget constraint. A key contribution of these studies is that they provide anytime bounds on the peak infection, which are universal and hold for any protocol that meets certain assumptions. The framework was extended to two-population scenarios in [8]. However, these studies only considered learning rules that satisfy Nash stationarity. Although many learning rules are Nash stationary (e.g., impartial pairwise protocol), there are some well-known learning rules that are not Nash stationary (e.g., imitation and PBR dynamics).

Here, we extend the framework proposed in [1], [2] to study PBR dynamics in epidemic population games. The PBR model (often referred to as the *better response model*), which is rooted in bounded rationality and accounts for random noise in the payoff mechanism, has been extensively investigated in the economics literature [28]–[34]. An earlier work relevant to the perturbed decision-making is [30] which presents a game formulation with randomly perturbed payoffs in finite-agent game settings. The mathematical analysis presented in the paper provides a rigorous approach to identifying the equilibrium, referred to as the *Bayesian equilibrium* [35], of such a game with perturbed payoffs.

Rosenthal [31] discussed the limitations of the Nash equilibrium in predicting outcomes of empirical studies in non-cooperative games. Motivated by such limitations, the author proposed a one-parameter better response model to capture the key characteristics in human decision-making: humans are equally likely to choose strategies that are equally effective. The study also provided concrete examples to illustrate the distinction between the equilibrium determined by the model and that of Nash's.

McKelvey and Palfrey [32] introduced a quantal choice model that includes the logit protocol as its special case and analyzed its equilibrium, referred to as the *quantal response equilibrium (QRE)*. By examining data from various human decision-making experiments, they illustrated that QRE for the logit learning rule consistently provides a better prediction across these experiments compared to the Nash equilibrium. Additionally, the study reports variability in the selection of the optimal model parameter during the analysis of experimental data, indicating the need for developing a parameter estimation method. More in-depth treatment of the logit learning rule is presented in [29]. The authors explained why the Nash's approach to defining an equilibrium, which relies only on the signs of payoff differences, falls short in predicting the outcomes of the experiments adopted from [36].

Other related studies include [33] which presents a model that captures the noise in agent decision-making to study the robustness of the *stochastic stability*. [34] introduces the idea of *stochastic potential* for a coordination game, which adds an effort cost to an ordinary potential function of the game, and draws the connection between the equilibrium maximizing the stochastic potential and the logit equilibrium. Using experimental data, the authors claim that such equilibrium is a better predictor of the experiment outcomes than that of Nash's.

## III. A SOLUTION TO MAIN PROBLEM

Recall the state equation of the PBR EDM given by

$$
\begin{aligned}
\dot{x}(t) &= \mathcal{V}(x(t), p(t)) \\
&= \arg\max_{z \in \text{int}(\mathbb{X})} (z'p(t) - Q(z)) - x(t), \quad \text{(PBR EDM)}
\end{aligned}
$$

where the payoff vector $p(t)$ is defined by the reward $r(t)$ and the intrinsic cost function $\tilde{c} : [0,1] \to \mathbb{R}^n_{\geq 0}$ as in (1) and (8). We adopt the notion of $\delta$-passivity from [37] to analyze the stability of (PBR EDM) when it is interconnected with (EPG). (PBR EDM) is $\delta$-*passive* in that there is a continuously differentiable function $\mathcal{S} : \mathbb{X} \times \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ given by

$$
\mathcal{S}(x, p) = \max_{z \in \text{int}(\mathbb{X})} (z'p - Q(z)) - (x'p - Q(x)), \quad (10)
$$

which satisfies

$$
\begin{aligned}
\nabla'_x \mathcal{S}(x,p)\mathcal{V}(x,p) + \nabla'_p \mathcal{S}(x,p)u &\leq u'\mathcal{V}(x,p), \\
\forall x \in \mathbb{X}, p \in \mathbb{R}^n, u \in \mathbb{R}^n. \quad (11)
\end{aligned}
$$

Note that $\nabla'_x \mathcal{S}(x,p)\mathcal{V}(x,p) \leq 0$ holds for all $x \in \mathbb{X}$, $p \in \mathbb{R}^n$, and $\mathcal{S}(x,p) = 0$ if and only if $\mathcal{V}(x,p) = 0$. We refer to $\mathcal{S}$ as the $\delta$-*storage function* of (PBR EDM).

Following the similar idea as in [1], we define $G$ and $H$ as follows.

$$
\begin{aligned}
G(I, R, x, q) = \kappa\Big( &(\hat{I} - I) + \eta(\ln I - \ln \hat{I}) + \upsilon^2(\bar{\beta} - \mathcal{B}) \\
&+ \frac{\mathcal{B}}{\gamma}(R - \hat{R})(1 - \eta - R)\Big) \quad (12a)
\end{aligned}
$$

$$
H(I, R, x, q) = q\vec{\beta} + \bar{r} + \tilde{c}(I) - \tilde{c}(\bar{I}), \quad (12b)
$$

where $\bar{I} = \eta(1 - \sigma/\bar{\beta})$, $\hat{I} = \eta(1 - \sigma/\mathcal{B})$, $\hat{R} = (1 - \eta)(1 - \sigma/\mathcal{B})$, $\mathcal{B} = \vec{\beta}'x$, and $\eta = \omega/(\omega + \gamma)$. The constant $\kappa$ represents the time scale of the dynamics for $q(t)$, and its effect on the convergence of $I(t)$ is discussed in §III-A. The constants $\bar{\beta} \in (\sigma, \bar{\beta}_n)$ and $\bar{r} \in \mathbb{R}^n_{\geq 0}$ are the endemic transmission rate and stationary reward vector when $q(t)$ vanishes as $t$ tends to infinity, respectively. These are design parameters of (EPG). The variables $\hat{I}$ and $\hat{R}$ constitute the equilibrium state of (EPGa) and (EPGb), and are considered as the *reference* endemic variables. From this perspective, the parameter $\upsilon$ can be interpreted as the weight that determines the importance of the current transmission rate $\mathcal{B}$ attaining the desired value $\bar{\beta}$ compared to $(I, R)$ attaining $(\hat{I}, \hat{R})$.

The following are remarks on the choice of $G$ and $H$.

*Remark 1:* According to the dynamics for $q$ defined in (12), $q$ is a signal adjusted by the social planner to control the portion $q\vec{\beta}$ in the reward vector (12b). Specifically, when $q < 0$, the reward vector incentivizes more effective strategies – those associated with lower values of $\vec{\beta}_i$. Additionally, $G$ in (12a) is defined such that $\frac{\partial G}{\partial \mathcal{B}} < 0$ holds. As a result, when agents switch to strategies that increase the average transmission rate (i.e., $\dot{\mathcal{B}}(t) > 0$), the value of $G$ would decrease. If this trend persists, it will further lower the value of $q(t)$, encouraging agents to adopt more effective strategies.

Furthermore, using Lemma 2 (in Appendix B), we can infer that $\mathcal{B}(t) > \sigma$, $\forall t \geq 0$, provided that $\mathcal{B}(0) > \sigma$.

If we want to ensure that the planner is offering rewards (non-negative incentives) to the population, the design of $H$ can be modified in the following way.

$$H(I, R, x, q) = q\vec{\beta} + \bar{r} + \tilde{c}(I) - \tilde{c}(\bar{I}) \\ - \min_{1 \leq i \leq n}(q\vec{\beta}_i + \bar{r}_i + \tilde{c}_i(I) - \tilde{c}_i(\bar{I}))\mathbf{1}. \quad (13)$$

Since adding an equal value to each entry of the payoff vector $p$ does not affect the results of this work, we use (12b) for concise presentation, unless otherwise specified.

### A. Optimal Design of $\bar{r}$, $\bar{\beta}$

We begin by presenting the main convergence results.

*Theorem 1:* Consider the closed-loop model consisting of (EPG) and (PBR EDM), with the parameters $\bar{\beta} \in (\sigma, \vec{\beta}_n)$ and $\bar{r} \in \mathbb{R}_{\geq 0}^n$ given. Suppose $\mathcal{B}(t) \geq \mathcal{B}_{\min}$, $\forall t \geq 0$ holds for some $\bar{\beta} > \mathcal{B}_{\min} > \sigma$; then the following statements hold:

1) $\lim_{t \to \infty} q(t) = \bar{q}$ and $\lim_{t \to \infty} \mathcal{B}(t) = \bar{\beta}$, where $\bar{q}$ is a unique solution to $\bar{\beta} = \vec{\beta}'C(\bar{q}\vec{\beta} + \bar{r} - \tilde{c}(\bar{I}))$ with $\bar{I} = \eta(1 - \sigma/\bar{\beta})$, and
2) $\lim_{t \to \infty} I(t) = \eta(1 - \sigma/\bar{\beta})$, $\lim_{t \to \infty} R(t) = (1-\eta)(1 - \sigma/\bar{\beta})$, and $\lim_{t \to \infty} x(t) = C(\bar{q}\vec{\beta} + \bar{r} - \tilde{c}(\bar{I}))$.

*Proposition 1:* When $C$ in (PBR EDM) is the logit learning rule (5) and $\upsilon$ in (EPGc) is sufficiently small, the closed-loop model, consisting of (EPG) and (PBR EDM), achieves local exponential stability.

The proofs are provided in Appendices B and C.

*Remark 2:* The assumption $\mathcal{B}(t) \geq \mathcal{B}_{\min}$, $\forall t \geq 0$ is imposed to ensure that the term $\ln I(t) - \ln \hat{I}(t)$ in (12a) converges to zero as does $I(t) - \hat{I}(t)$. Note that if $\vec{\beta}_1 > \sigma$, this assumption is satisfied with $\mathcal{B}_{\min} = \vec{\beta}_1$. In cases where $\vec{\beta}_1 < \sigma$ and $\mathcal{B}(0) \geq \mathcal{B}_{\min}$, the requirement can be strictly enforced by redefining the reward vector $r(t)$ as

$$r(t) = \begin{cases} q_{\min}\vec{\beta} + \bar{r} + \tilde{c}(I(t)) - \tilde{c}(\bar{I}) & \text{if } \mathcal{B}(t) = \mathcal{B}_{\min}, q(t) \leq q_{\min} \\ q(t)\vec{\beta} + \bar{r} + \tilde{c}(I(t)) - \tilde{c}(\bar{I}) & \text{otherwise,} \end{cases}$$

where $q_{\min}$ is selected to satisfy $\mathcal{B}_{\min} = \vec{\beta}'C(q_{\min}\vec{\beta} + \bar{r} - \tilde{c}(\bar{I})) < \bar{\beta}$, and $\mathcal{B}_{\min}$ can be chosen arbitrarily close to $\sigma$.

In light of Theorem 1, the budget constraint (P2) can be expressed as $(\bar{q}\vec{\beta} + \bar{r})'C(\bar{q}\vec{\beta} + \bar{r} - \tilde{c}(\bar{I})) \leq c^*$ where $c^*$ is the cost budget, and $\bar{q}$ is the limit of $q(t)$ for given $\bar{r}$ and $\bar{\beta}$. Consider the optimal choice $\bar{r} = r^*$ and $\bar{\beta} = \beta^*$, where $r^*$ and $\beta^*$ are part of the optimal solution to (9). In this case, we can conclude that $\bar{q} = 0$ and thus, $r^{*\prime}C(r^* - \tilde{c}(I^*))$ represents the average cost that the social planner needs to spend to maintain the optimal transmission rate $\beta^*$ when the closed-loop model reaches equilibrium. As a corollary of Theorem 1, we can show that with $\bar{r} = r^*$ and $\bar{\beta} = \beta^*$, the reward vector $r(t)$ converges to $r^*$, and the transmission rate $\mathcal{B}(t)$ converges to the minimum $\beta^*$ that satisfies the budget constraint.

*Corollary 1:* Suppose $\bar{r}$ and $\bar{\beta}$ are determined as the optimal solution to (9): $\bar{r} = r^*$ and $\bar{\beta} = \beta^* = \vec{\beta}'C(r^* - \tilde{c}(I^*))$. For the feedback interconnection of (EPG) and (PBR EDM), assume

that $\mathcal{B}(t) \geq \mathcal{B}_{\min}$, $\forall t \geq 0$, where $\beta^* > \mathcal{B}_{\min} > \sigma$. Then, the following hold:

1) $\lim_{t \to \infty} \mathcal{B}(t) = \beta^*$, and
2) $\lim_{t \to \infty} r(t) = r^*$ while satisfying $r^{*\prime}C(r^* - \tilde{c}(I^*)) \leq c^*$.

*Remark 3:* The following are remarks regarding Theorem 1 and Corollary 1.

1) For the planner to compute $r^*$ and $\beta^*$ using (9), the knowledge of the choice function $C$ is essential. In §III-C, we consider a scenario where $C$ is initially unknown to the planner. In such cases, $C$ must be estimated based on data gathered over time from the agents' strategy selections.
2) The proof of Theorem 1 adopts the Lyapunov stability technique which, in conjunction with Corollary 1, implies that the infectious population $I(t)$ converges to $I^*$. In other words, the reward mechanism defined by (EPGc,d) asymptotically attains the minimum achievable infectious population while asymptotically satisfying the given budget constraint.
3) The higher the value of $\kappa$, in (12a), the faster the dynamics of $q(t)$. Consequently, with large $\kappa$, the planner can more quickly adjust $q(t)$ to control the spread of the epidemic. To see this, consider the Lyapunov function $\mathscr{S}(\mathcal{I}, \mathcal{R}, \mathcal{B})$, defined in (27), and used to establish the convergence results in Theorem 1. Here, $\mathcal{I}$ and $\mathcal{R}$ are given as $\mathcal{I} = \mathcal{B}I$ and $\mathcal{R} = \mathcal{B}R$, respectively. The following conditions hold:

$$\mathscr{S}(\mathcal{I}(t), \mathcal{R}(t), \mathcal{B}(t)) = 0 \implies \dot{\mathcal{I}}(t) = 0$$
$$\frac{\mathrm{d}\mathscr{S}}{\mathrm{d}t}(\mathcal{I}(t), \mathcal{R}(t), \mathcal{B}(t)) = -\kappa^{-1}G(I(t), R(t), x(t), q(t))\dot{\mathcal{B}}(t) \\ - (\mathcal{I}(t) - \hat{\mathcal{I}}(t))^2 - \frac{\omega}{\gamma}(\mathcal{R}(t) - \hat{\mathcal{R}}(t))^2.$$

We explain how large $\kappa$ facilitates the reduction of $\mathscr{S}(\mathcal{I}(t), \mathcal{R}(t), \mathcal{B}(t))$, thereby slowing the growth of the infectious population $I(t)$, particularly when $\dot{I}(t) > 0$. To achieve this, the planner can incentivize agents to adopt strategies that minimize the upper bound $-\kappa^{-1}G(I(t), R(t), x(t), q(t))\dot{\mathcal{B}}(t)$ of $\frac{\mathrm{d}\mathscr{S}}{\mathrm{d}t}(\mathcal{I}(t), \mathcal{R}(t), \mathcal{B}(t))$, which directly depends on $q(t)$. When $\kappa$ is sufficiently large, $\kappa^{-1}G(I(t), R(t), x(t), q(t)) < 0$ leads to a rapid decrease in $q(t)$, which in turn reduces $\dot{\mathcal{B}}(t) = \vec{\beta}' \arg\max_{z \in \text{int}(\mathbb{X})}(z'p(t) - Q(z)) - \mathcal{B}(t)$ by driving $\vec{\beta}' \arg\max_{z \in \text{int}(\mathbb{X})}(z'p(t) - Q(z))$ towards its lower bound $\vec{\beta}_1$, as indicated by Lemma 2 in Appendix B. Conversely, when $\kappa^{-1}G(I(t), R(t), x(t), q(t)) > 0$, the variable $q(t)$ increases rapidly, leading to an increase in $\dot{\mathcal{B}}(t)$ as $\vec{\beta}' \arg\max_{z \in \text{int}(\mathbb{X})}(z'p(t) - Q(z))$ moves towards its upper bound $\vec{\beta}_n$.

The following example illustrates the convergence result of Corollary 1.

*Example 2:* Consider that (PBR EDM) is defined by the logit learning rule with $\mu = 1$ and is interconnected with (EPG), where $\gamma = 0.1$ (infectiousness period $\sim 10$ days), $\sigma = $
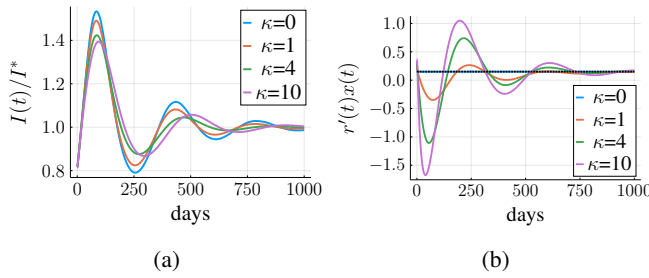
Fig. 1: Simulation results for Example 2 illustrating (a) the infectious fraction $I(t)$ of the population with respect to $I^*$ and (b) the average cost $r'(t)x(t)$, evaluated with different values of $\kappa$ for (12a).

$\gamma$, $\omega = 0.005$ (immunity period $\sim 200$ days), and $\upsilon = 3$. With the transmission rates $\vec{\beta} = (0.15, 0.19)$, intrinsic cost $c(I) = (0.2038 - 0.2I, 0)$, and budget $c^* = 0.15$, we obtain $r^* \approx (0.287, 0)$ and $\beta^* = 0.1691$ from (9) under which $x(t)$ and $(I(t), R(t))$ converge to $x^* = C(r^* - \tilde{c}(I^*)) \approx (0.522, 0.478)$ and $(I^*, R^*) \approx (0.019, 0.389)$, respectively.

Using the initial condition $x(0) = (1, 0)$, $I(0) = 0.0158$, $R(0) = 0.3170$, and $q(0) = 0$, Fig. 1 illustrates the ratio $I(t)/I^*$ and the instantaneous average cost $r'(t)x(t)$ with varying values of $\kappa$ in (EPGc). We can observe that $I(t)/I^*$ converges to 1 in Fig. 1a, indicating that the infectious population $I(t)$ approaches the endemic equilibrium $I^*$, and $r'(t)x(t)$ converges to $c^*$, marked by the dotted horizontal line in Fig. 1b. In addition, as we can observe from the figure, increasing $\kappa$ will reduce the overshoot in $I(t)/I^*$ (see Fig. 1a) at the expense of larger instantaneous cost (see Fig. 1b).

### B. Anytime Bound

We establish an anytime bound on the infectious fraction $I(t)$ of the population. By following the same arguments as in [1, §4], we proceed with defining

$$\pi_\upsilon(\alpha_\kappa) = \bar{I}^{-1} \sup \{I \mid \mathscr{S}(\mathcal{I}, \mathcal{R}, \mathcal{B}) \leq \alpha_\kappa\}, \quad (14)$$

where $\mathcal{I} = \mathcal{B}I$, $\mathcal{R} = \mathcal{B}R$, $\bar{I} = \eta(1 - \sigma/\bar{\beta})$ for fixed $\bar{\beta} \in (\sigma, \vec{\beta}_n)$, and $\mathscr{S}(\mathcal{I}, \mathcal{R}, \mathcal{B})$ is the Lyapunov function defined in (27). The value $\pi_\upsilon(\alpha_\kappa)$ represents the upper bound of $I(t)/\bar{I}$, where $\pi_\upsilon(0) = 1$ and $\alpha_\kappa$ is determined by the initial condition $(I(0), R(0), q(0), x(0))$ of the closed-loop model. Specifically, $\alpha_\kappa$ is given by $\alpha_\kappa = \kappa^{-1}\mathscr{S}(x(0), p(0)) + \mathscr{S}(\mathcal{I}(0), \mathcal{R}(0), \mathcal{B}(0))$, where $\mathscr{S}(x, p)$ is the $\delta$-storage function of (PBR EDM) defined in (10).

To be more specific, according to (28), the function $\kappa^{-1}\mathscr{S}(x(t), p(t)) + \mathscr{S}(\mathcal{I}(t), \mathcal{R}(t), \mathcal{B}(t))$ is decreasing in $t$ and, hence, it holds that

$$\mathscr{S}(\mathcal{I}(t), \mathcal{R}(t), \mathcal{B}(t)) \leq \alpha_\kappa, \ \forall t \geq 0. \quad (15)$$

From (14), we can establish that $I(t) \leq \bar{I}\pi_\upsilon(\alpha_\kappa)$ holds for all $t \geq 0$. Additionally, note that the constant $\alpha_\kappa$ decreases as $\kappa$ increases, as does the upper bound $\bar{I}\pi_\upsilon(\alpha_\kappa)$, as observed in Fig. 1a.

Now, to evaluate how the choice of the parameters $\bar{r}$ and $\bar{\beta}$ affects the anytime bound, consider that the state of the closed-loop model, consisting of (EPG) and (PBR EDM),

starts from the endemic equilibrium resulting from a prior use of (EPG) with $r^o$ and $\beta^o$, respectively, as the stationary reward vector and endemic transmission rate. Hence, $I(0) = \hat{I}(0) = \eta(1 - \sigma/\mathcal{B}(0))$ and $R(0) = \hat{R}(0) = (1 - \eta)(1 - \sigma/\mathcal{B}(0))$ hold. Suppose the social planner adopts $\bar{r}$ and $\bar{\beta}$ as the revised stationary reward vector and target endemic transmission rate, respectively. Note that in this case, $\mathcal{B}(0) = \beta^o$, $p(0) = q(0)\vec{\beta} + \bar{r} - \tilde{c}(\bar{I})$, and $\mathscr{S}(\mathcal{I}(0), \mathcal{R}(0), \mathcal{B}(0)) = \upsilon^2\tilde{\beta}^2/2$ hold, where $\bar{I} = \eta(1 - \sigma/\bar{\beta})$ and $\tilde{\beta} = \beta^o - \bar{\beta}$. Also, we can evaluate the $\delta$-storage function $\mathcal{S}$ of (PBR EDM) at $t = 0$ as

$$\mathcal{S}(x(0), p(0)) = \max_{z \in \text{int}(\mathbb{X})} (z'(q(0)\vec{\beta} + \bar{r} - \tilde{c}(\bar{I})) - Q(z))$$
$$- (x'(0)(q(0)\vec{\beta} + \bar{r} - \tilde{c}(\bar{I})) - Q(x(0)))$$
$$= \max_{z \in \text{int}(\mathbb{X})} (z'(q(0)\vec{\beta} + \bar{r} - \tilde{c}(\bar{I})) - Q(z))$$
$$- \max_{z \in \text{int}(\mathbb{X})} (z'(q(0)\vec{\beta} + r^o - \tilde{c}(I^o)) - Q(z))$$
$$- x'(0) \left(\bar{r} - \tilde{c}(\bar{I}) - (r^o - \tilde{c}(I^o))\right), \quad (16)$$

where $I^o = \eta(1 - \sigma/\beta^o)$. To derive the latter equality, we use the fact that before the revision to $\bar{r}$ and $\bar{\beta}$, the state was at the previous endemic equilibrium. Consequently, $\mathcal{S}(x(0), q(0)\vec{\beta} + r^o - \tilde{c}(I^o)) = 0$ holds. With fixed $r^o$ and $\beta^o$, we can express $\mathcal{S}(x(0), p(0))$ as a continuous function $B_\mathcal{S}(\tilde{r}, \tilde{\beta})$ of $\tilde{r}$ and $\tilde{\beta}$ satisfying $B_\mathcal{S}(0, 0) = 0$, where $\tilde{r} = r^o - \bar{r}$.[3]

In conjunction with (14) and (15), by selecting $\alpha_\kappa = \kappa^{-1}\mathcal{S}(x(0), p(0)) + \mathscr{S}(\mathcal{I}(0), \mathcal{R}(0), \mathcal{B}(0)) = \kappa^{-1}B_\mathcal{S}(\tilde{r}, \tilde{\beta}) + \upsilon^2\tilde{\beta}^2/2$, we can establish the anytime bound on $I(t)$ as

$$I(t) \leq \bar{I}\pi_\upsilon(\kappa^{-1}B_\mathcal{S}(\tilde{r}, \tilde{\beta}) + \upsilon^2\tilde{\beta}^2/2), \forall t \geq 0. \quad (17)$$

Therefore, the bound (17) indicates how much the infectious fraction $I(t)$ exceeds its target value $\bar{I} = \eta(1 - \sigma/\bar{\beta})$ when the planner revises the stationary reward vector and endemic transmission rate to $\bar{r}$ and $\bar{\beta}$, respectively. Additionally, (17) highlights how the choice of $\kappa$ and $\upsilon$ influences the overshoot of $I(t)$.

Using a similar analysis, we can establish an anytime bound on the instantaneous average reward $r'(t)x(t)$ as follows:

$$\sup \{(p + \tilde{c}(I))'x \mid \kappa^{-1}\mathcal{S}(x, p) + \mathscr{S}(\mathcal{I}, \mathcal{R}, \mathcal{B}) \leq \alpha_\kappa\},$$

where $p = q\vec{\beta} + \bar{r} - \tilde{c}(\bar{I})$, $\mathcal{B} = \vec{\beta}'x$, $\mathcal{I} = \mathcal{B}I$, $\mathcal{R} = \mathcal{B}R$, $\bar{I} = \eta(1 - \sigma/\bar{\beta})$, and $\alpha_\kappa = \kappa^{-1}\mathcal{S}(x(0), p(0)) + \mathscr{S}(\mathcal{I}(0), \mathcal{R}(0), \mathcal{B}(0))$.

### C. Learning One-Parameter Choice Function $\mathsf{C}^\mu$

In §III-A and §III-B, we discussed computing the optimal stationary reward vector $r^*$ and endemic transmission rate $\beta^*$ using the minimization (9), and establishing the anytime bound (17). To solve the minimization and compute the bound, the planner needs to know the choice function $C$, which dictates how agents in the population revise their strategies. In this section, we discuss a scenario wherein the planner needs to estimate $C$.

We focus on a class of one-parameter choice functions, motivated by empirical studies [31], [34] demonstrating the

---

[3]The function $B_\mathcal{S}$ is continuous since $\max_{z \in \text{int}(\mathbb{X})}(z'(q(0)\vec{\beta} + r) - Q(z))$ is differentiable in $r$ [4] and $\bar{r} - \tilde{c}(\bar{I}) - (r^o - \tilde{c}(I^o))$ is continuous in $\tilde{r}$ and $\tilde{\beta}$.

predictive power of such a class of models.[4] To proceed, we express $C$ as

$$C(p) = C^\mu(p) = \arg\max_{z \in \text{int}(\mathbb{X})} (z'p - \mu\bar{Q}(z)), \qquad (18)$$

where $\bar{Q} : \text{int}(\mathbb{X}) \to \mathbb{R}$ is a known deterministic payoff perturbation whereas $\mu > 0$ is an unknown parameter that needs to be estimated. The parameter $\mu$ quantifies the level of the perturbation. To see this, according to (4), recall that the choice function is given as $C_i^\mu(p) = \mathbb{P}(p_i + v_i \geq \max_{1 \leq j \leq n}(p_j + v_j))$, where $v_1, \cdots, v_n$ are random variables associated with the deterministic perturbation $\mu\bar{Q}$. By defining $C_i^1(p) = \mathbb{P}(p_i + \bar{v}_i \geq \max_{1 \leq j \leq n}(p_j + \bar{v}_j))$, we can establish that $v_i = \mu\bar{v}_i$, $\forall i \in \{1, \cdots, n\}$, where $\bar{v}_1, \cdots, \bar{v}_n$ are random variables associated with $\bar{Q}$. Consequently, $\mu$ can be interpreted as the scaling factor for $\bar{v}_i$. Hence, the larger the parameter $\mu$ is, the greater the noise is in the agent decision making. The assumption that $\bar{Q}$ is known implies that the planner knows the joint probability density function of $\bar{v}_1, \cdots, \bar{v}_n$.

In the following lemma, we examine an important property of the choice function $C^\mu$ that is useful in the estimation of $\mu$.

*Lemma 1:* Let $C^\mu$ be a one-parameter choice function (18). For any given $r \in \mathbb{R}_{\geq 0}^n$ and $\tilde{c}(I) \in \mathbb{R}_{\geq 0}^n$ associated with fixed $I$, where not all entries of $r - \tilde{c}(I)$ are identical, $(r - \tilde{c}(I))'C^\mu(r - \tilde{c}(I))$ is a decreasing function of $\mu$.

The proof is provided in Appendix D. Consequently, the choice function $C^\mu$ satisfies $(r - \tilde{c}(I))'C^{\hat{\mu}}(r - \tilde{c}(I)) = (r - \tilde{c}(I))'C^\mu(r - \tilde{c}(I))$ if and only if $\hat{\mu} = \mu$ holds. Hence, when $(r, C^\mu(r - \tilde{c}(I)))$ is given as data for the estimation of $\mu$, the social planner can learn unique $\mu$. In what follows, we discuss how the planner can devise a parameter estimation scheme based on Lemma 1 and then use it in conjunction with (EPG) to determine $r(t)$ for the population.

At the beginning of a pandemic, a social planner, who is initially unaware of the parameter $\mu$, can employ (EPG) with an initial selection of $\bar{r}$ and $\bar{\beta}$ to curb the spread of the pandemic. Simultaneously, the planner can collect data on the agents' strategy selections to estimate $\mu$. For data collection, we consider a survey method in which randomly selected agents are asked which strategies they would choose given a fixed reward vector $r$, which may differ from $\bar{r}$ used in (EPG). Note that each agent's response represents a specific strategy randomly drawn from the probability distribution $C^\mu(r - \tilde{c}(I))$. The following proposition outlines the method for calculating probabilistic upper and lower bounds of $\mu$ using the survey data.

*Proposition 2:* Let $r = \tilde{r} + \tilde{c}(I)$ represent the reward vector used in the survey, where $I$ denotes the current infectious population and $\tilde{r}$ is a vector that satisfies the condition:

$$\max_{1 \leq i \leq n} \tilde{r}_i - \min_{1 \leq i \leq n} \tilde{r}_i = 2. \qquad (19)$$

[4] Also, as explained in Appendix A, the choice of probability distribution for the noise in (4) does not substantially impact the outcomes of (EPG). This observation highlights that the one-parameter choice functions would be effective models in studying epidemic population games.

Let $\{\xi_k\}_{k=1}^K \subset \{e_1, \cdots, e_n\}$ represent the strategy selections of agents based on the probability distribution $C^\mu(r - \tilde{c}(I))$, where the standard basis vector $e_i$ corresponds to the selection of strategy $i$. Given a positive constant $\epsilon$, compute $\mu_L$ and $\mu_U$ satisfying

$$\tilde{r}'C^{\mu_L}(\tilde{r}) = \tfrac{1}{K}\sum_{k=1}^K \tilde{r}'\xi_k + \epsilon \qquad (20a)$$

$$\tilde{r}'C^{\mu_U}(\tilde{r}) = \tfrac{1}{K}\sum_{k=1}^K \tilde{r}'\xi_k - \epsilon. \qquad (20b)$$

We can establish the following probabilistic bounds on $\mu$:

$$\mathbb{P}(\mu_L \leq \mu \leq \mu_U) \geq 1 - 1/\epsilon^2 K. \qquad (21)$$

The proof is provided in Appendix E. Once enough data has been collected to estimate $\mu$ with the required accuracy, the planner can optimize the choice of $\bar{r}$ and $\bar{\beta}$ using (9). Specifically, with a sufficiently large number of data points $K$, a small enough $\epsilon$ can be chosen to minimize the gap between the lower bound $\mu_L$ and upper bound $\mu_U$, while maintaining a desired confidence level of $1 - 1/\epsilon^2 K$.

During this parameter learning phase, since the planner does not know the actual value of $\mu$, it can be difficult to predict the limit $\bar{q}$ of $q(t)$ and ensure that the budget constraint $(\bar{q}\bar{\beta} + \bar{r})'C^\mu(\bar{q}\bar{\beta} + \bar{r} - \tilde{c}(\bar{I})) \leq c^*$ is satisfied. To address this issue, using the bounds on $\mu$, we explain how the planner can estimate an upper bound on $(\bar{q}\bar{\beta} + \bar{r})'C^\mu(\bar{q}\bar{\beta} + \bar{r} - \tilde{c}(\bar{I}))$ and use the bound to meet the budget requirement.

Suppose $\mu$ belongs to $[\mu_L, \mu_U]$. Then, the upper bound on $(\bar{q}\bar{\beta} + \bar{r})'C^\mu(\bar{q}\bar{\beta} + \bar{r} - \tilde{c}(\bar{I}))$ can be derived as

$$\max_{\mu_L \leq \mu \leq \mu_U} (\bar{q}(\mu)\vec{\beta} + \bar{r})'C^\mu(\bar{q}(\mu)\vec{\beta} + \bar{r} - \tilde{c}(\bar{I})), \qquad (22)$$

where $\bar{q}(\mu)$ is a solution, which depends on $\mu$, to $\bar{\beta} = \vec{\beta}'C^\mu(\bar{q}(\mu)\vec{\beta} + \bar{r} - \tilde{c}(\bar{I}))$. Ideally, we want the evaluation of (22) to be computationally tractable so that the planner can assess the upper bound (22) over a wide range of $\bar{r}, \bar{\beta}$ and select the one that minimizes the endemic transmission rate $\bar{\beta}$ subject to the budget constraint.

In the following proposition, we derive technical conditions under which the upper bound (22) can be computed using a single run of Newton's method. For this purpose, given $\bar{\beta}$, we fix $\bar{r} = \tilde{c}(\bar{I})$ with $\bar{I} = \eta(1 - \sigma/\bar{\beta})$ and assume $\tilde{c}'(\bar{I})C^1(0) < c^*$.

*Proposition 3:* Suppose $\mu$ lies within the interval $[\mu_L, \mu_U]$ and $\bar{\beta}$ satisfies $\bar{\beta} < \vec{\beta}'C^1(0)$. For fixed $\bar{\beta}$, the average stationary cost $(\bar{q}\vec{\beta} + \bar{r})'C^\mu(\bar{q}\vec{\beta} + \bar{r} - \tilde{c}(\bar{I}))$ is upper bounded by

$$\mu_U\lambda(\bar{\beta} - \bar{\beta}_n) + \tilde{c}'(\bar{I})\arg\max_{z \in \text{int}(\mathbb{X})}(z'(\lambda\vec{\beta}) - \bar{Q}(z)). \qquad (23)$$

The parameter $\lambda$ is a negative real number satisfying $\bar{\beta} = \vec{\beta}'\arg\max_{z \in \text{int}(\mathbb{X})}(z'(\lambda\vec{\beta}) - \bar{Q}(z))$, where

$$\vec{\beta}'\arg\max_{z \in \text{int}(\mathbb{X})}(z'(\lambda\vec{\beta}) - \bar{Q}(z))$$

is an increasing function of $\lambda$.

The proof is provided in Appendix F. By Proposition 3, the unique solution $\lambda$ in (23) can be found using Newton's method. Based on the results stated in the proposition, the planner can compute an appropriate $\bar{\beta}$ as follows: Starting from $\bar{\beta} = \vec{\beta}'C^1(0)$, which incurs the average stationary cost of $\tilde{c}'(\bar{I})C^1(0) < c^*$, the planner can repeatedly assess (23) as
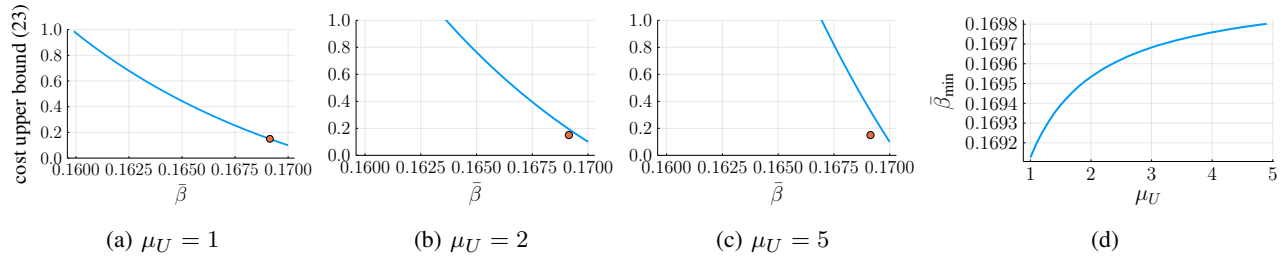
Fig. 2: Plots (a)–(c) illustrate how the cost upper bound (23) varies depending on $\bar{\beta}$ for each fixed $\mu_U$: (a) $\mu_U = 1$, (b) $\mu_U = 2$, and (c) $\mu_U = 5$. The red circle indicates the optimal endemic transmission rate $\beta^* = 0.1691$ and the budget $c^* = 0.15$. Plot (d) depicts the smallest $\bar{\beta}_{\min}$ under which the cost upper bound does not exceed the budget $c^* = 0.15$ for $\mu_U \in [1, 5]$, i.e., $\bar{\beta}_{\min} = \min\{\bar{\beta} > 0 \,|\, \mu_U \lambda(\bar{\beta} - \bar{\beta}_n) + \tilde{c}'(\bar{I}) \arg\max_{z \in \mathrm{int}(\mathbb{X})} (z'(\lambda\vec{\beta}) - \bar{Q}(z)) \leq c^*\}$.

it decreases the value of $\bar{\beta}$ until it finds the lowest $\bar{\beta}$ whose associated bound (23) satisfies the budget constraint.

Fig. 2 illustrates how the upper bound (23) varies depending on the values of $\mu_U$ and $\bar{\beta}$. The choice function is assumed to be logit with parameter $\mu = 1$, and the parameters $\vec{\beta}, \tilde{c}(I)$, and $c^*$ are given by $\vec{\beta} = (0.15, 0.19)$, $\tilde{c}(I) = (0.2038 - 0.2I, 0)$, and $c^* = 0.15$, respectively. From the plots in Fig. 2, we observe that as the bound on $\mu$ becomes loose, so does the estimated upper bound of the average stationary cost for the given budget. In particular, we can see from Fig. 2a that if the upper bound of $\mu$ is tight, i.e., $\mu_U = \mu$, then so is the bound (23) for the given $c^* = 0.15$.

## IV. SIMULATIONS

We now present simulations that illustrate our main results. In particular, we use a deployment scenario, detailed below, to demonstrate how the parameter learning approach for the choice function, along with the upper bound of the average stationary cost (as discussed in §III-C), can be applied by the planner to infer the parameter $\mu$ of the choice function and to determine appropriate $\bar{r}, \bar{\beta}$ for (EPG).

1) At the beginning of a pandemic, given a budget of $c^*$ and a known upper bound of $\mu$, the planner computes the minimum $\bar{\beta}$ such that the corresponding cost upper bound (23) satisfies the budget constraint with $\bar{r} = \tilde{c}(\bar{I})$, where $\bar{I} = \eta(1 - \sigma/\bar{\beta})$.

2) By surveying randomly selected agents on their strategy choices with the reward vector $r = \tilde{r} + \tilde{c}(I)$, where $\tilde{r}$ is a constant vector and $I$ denotes the current infectious population, the planner computes the sample mean of the survey outcomes. Using Proposition 2, it then estimates the upper and lower bounds of $\mu$ for a given confidence level.

3) After estimating the value of $\mu$ satisfying any required accuracy and confidence level, the planner computes the optimal solution $r^*, \beta^*$ for (9) and revises $\bar{r}, \bar{\beta}$ accordingly. Using the revised $\bar{r}, \bar{\beta}$ and estimated $\mu$, the planner can establish the anytime bound on the infectious population $I(t)$ based on (14).

Consider that the agents adopt the logit learning rule with $\mu = 1$ and the parameters of (EPG) are given by $\sigma = \gamma = 0.1$, $\omega = 0.005$, $\kappa = 1$, and $\upsilon = 3$. With the transmission rates $\vec{\beta} = (0.15, 0.19)$, intrinsic cost $c(I) = (0.2038 - 0.2I, 0)$, and budget $c^* = 1$, we obtain $r^* \approx (1.3248, 0)$ and $\beta^* \approx 0.1598$

from (9) under which $I(t)$ converges to $I^* \approx 0.0178$. To define initial conditions for (EPG) and (PBR EDM), we assume that the majority of the agents use the costlier strategy at $t = 0$, i.e., $x(0) = (0.997, 0.003)$, and $I(0) = 0.0159$, $R(0) = 0.318$, and $q(0) = 0$.

Suppose that the planner is aware of the range $(0, 5]$ for the parameter $\mu$, but not its exact value. By following the method explained in Proposition 3, the planner selects $\bar{r} = \tilde{c}(\bar{I})$ and $\bar{\beta} \approx 0.167$ under which the upper bound (23) satisfies the budget constraint. The planner utilizes (21) with $\epsilon = 0.05$ to establish the lower and upper bounds of $\mu$ and to determine the number of data points required to achieve a confidence level of 0.95 as it gathers poll data from 1,000 randomly selected agents every 30 days. Let $t_0 = 240$ be the day on which the estimated $\mu$ satisfies the estimation accuracy requirement. Once it does, the planner computes the estimate of $\mu$ based on the collected poll data, and the parameters $\bar{r}, \bar{\beta}$ are updated using the solution $r^*, \beta^*$ to (9). We consider the following two scenarios for the update of $\bar{r}, \bar{\beta}$.

In Scenario 1, the planner assigns $\bar{r} = r^*$ and $\bar{\beta} = \beta^*$ on the day $t_0$, whereas in Scenario 2, the planner selects $\bar{\beta} = \beta^*$ on the day $t_0$ and gradually changes $\bar{r}$ toward $r^*$. To implement the second scenario, after the day $t_0$, the planner evaluates

$$\alpha = \mathcal{S}(x(t), p(t)) + \mathscr{S}(\mathcal{I}(t), \mathcal{R}(t), \mathcal{B}(t)),$$
$$t = t_0 + 30, t_0 + 60, \cdots$$

at every 30 days and updates $\bar{r}$ to the value that is closest to $r^*$ while ensuring that $\alpha$ remains below 0.0004. According to (14), this corresponds to a 15% overshoot in $I(t)/I^*$. Fig. 3 illustrates the simulation results for the two scenarios. As depicted in Fig. 3a, the overshoot in the ratio $I(t)/I^*$ is smaller when the parameters $\bar{r}, \bar{\beta}$ are judiciously tuned after $t_0$ using the anytime bound (14). But, this reduction in the second overshoot, which appears after $t_0$, requires penalizing the population (i.e., negative rewards) as can be observed in Fig. 3b. Thus, maintaining non-negative rewards would demand increasing the rewards according to (13).

## V. CONCLUSIONS AND FUTURE RESEARCH PLANS

We studied the problem of designing a dynamic payoff mechanism in epidemic population games. Unlike existing studies, such as [1], [2], our work analyzes the dynamics underlying the games where the agents' decision making is
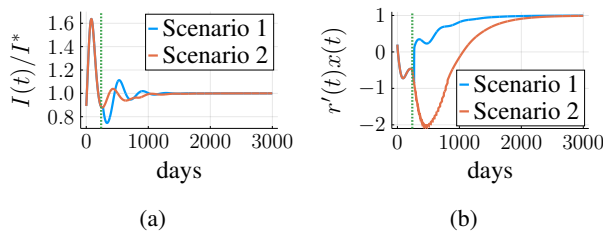
This article has been accepted for publication in IEEE Transactions on Automatic Control. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAC.2025.3554601

9



(a)

(b)

Fig. 3: Simulation results for Scenario 1 and Scenario 2 illustrating (a) the infectious fraction $I(t)$ of the population with respect to $I^*$ and (b) the average cost $r'(t)x(t)$, where the parameters $\bar{r}, \bar{\beta}$ of (EPG) are updated using the optimal $r^*, \beta^*$ beginning from the day $t_0 = 240$, as indicated by the green dotted vertical line.

subject to perturbation. We adopted (PBR EDM) to formalize this perturbed decision-making process and established stability of the feedback interconnection between (EPG) and (PBR EDM). Notably, our main results demonstrate the global convergence of the average transmission rate to its achievable minimum, subject to the budget constraint, thereby minimizing the infectious population in the long run. Additionally, we established an anytime bound on the infectious population using the Lyapunov stability method. We also discussed scenarios where the social planner needs to estimate the parameter of the agents' learning rule. As a direction for future research, we plan to analyze the transient behavior of the average cost trajectory. In the current study, we focused on designing (EPG) for the minimization of the average transmission rate while satisfying the asymptotic budget constraint. However, this consideration does not provide sufficient insight into the variation of the transient average cost.

[1] N. C. Martins, J. Certório, R. J. La, Epidemic population games and evolutionary dynamics, Automatica 153 (2023) 111016.

[2] J. Certório, N. C. Martins, R. J. La, Epidemic population games with nonnegligible disease death rate, IEEE Control Systems Letters 6 (2022) 3229–3234.

[3] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, A. Vespignani, Epidemic processes in complex networks, Rev. Mod. Phys. 87 (2015) 925–979.

[4] J. Hofbauer, W. H. Sandholm, On the global convergence of stochastic fictitious play, Econometrica 70 (6) (2002) 2265–2294.

[5] J. Hofbauer, W. H. Sandholm, Evolution in games with randomly disturbed payoffs, Journal of Economic Theory 132 (1) (2007) 47–69.

[6] W. H. Sandholm, Population games and evolutionary dynamics, MIT Press, 2010.

[7] W. Sandholm, Population games and deterministic evolutionary dynamics, in: Handbook of Game Theory with Economic Applications, Vol. 4, Elsevier, 2015, Ch. 13, pp. 703–778.

[8] J. Certorio, R. J. La, N. C. Martins, Epidemic population games for policy design: Two populations with viral reservoir case study, in: 2023 62nd IEEE Conference on Decision and Control (CDC), 2023, pp. 7667–7674.

[9] P. Godara, S. Herminghaus, K. M. Heidemann, A control theory approach to optimal pandemic mitigation, PLOS ONE 16 (2) (2021) 1–16.

[10] E. D. Sontag, An explicit formula for minimizing the infected peak in an SIR epidemic model when using a fixed number of complete lockdowns, International Journal of Robust and Nonlinear Control 33 (9) (2021) 4708–4731.

[11] A. Mao, L. Dworkin, S. Suri, D. J. Watts, Resilient cooperators stabilize long-run cooperation in the finitely repeated prisoner's dilemma, Nature Communications 8 (1) (2017) 13800.

[12] F. Di Lauro, I. Z. Kiss, J. C. Miller, Optimal timing of one-shot interventions for epidemic control, PLOS Computational Biology 17 (3) (2021) 1–25.

[13] M. A. Al-Radhawi, M. Sadeghi, E. D. Sontag, Long-term regulation of prolonged epidemic outbreaks in large populations via adaptive control: A singular perturbation approach, IEEE Control Systems Letters 6 (2022) 578–583.

[14] M. A. Amaral, M. M. de Oliveira, M. A. Javarone, An epidemiological model with voluntary quarantine strategies governed by evolutionary game dynamics, Chaos, Solitons & Fractals 143 (2021) 110616.

[15] C. T. Bauch, D. J. D. Earn, Vaccination and the theory of games, Proceedings of the National Academy of Sciences 101 (36) (2004) 13391–13394.

[16] A. d'Onofrio, P. Manfredi, P. Poletti, The impact of vaccine side effects on the natural history of immunization programmes: An imitation-game approach, Journal of Theoretical Biology 273 (1) (2011) 63–71.

[17] A. R. Hota, S. Sundaram, Game-theoretic vaccination against networked SIS epidemics and impacts of human decision-making, IEEE Transactions on Control of Network Systems 6 (4) (2019) 1461–1472.

[18] K. M. A. Kabir, J. Tanimoto, Evolutionary game theory modelling to represent the behavioural dynamics of economic shutdowns and shield immunity in the COVID-19 pandemic, Royal Society Open Science 7 (9) (2020) 201095.

[19] U. Maitra, A. R. Hota, V. Srivastava, SIS epidemic propagation under strategic non-myopic protection: A dynamic population game approach, IEEE Control Systems Letters 7 (2023) 1578–1583.

[20] H. Khazaei, K. Paarporn, A. Garcia, C. Eksin, Disease spread coupled with evolutionary social distancing dynamics can lead to growing oscillations, in: 2021 60th IEEE Conference on Decision and Control (CDC), 2021, pp. 4280–4286.

[21] C. Nowzari, V. M. Preciado, G. J. Pappas, Analysis and control of epidemics: A survey of spreading processes on complex networks, IEEE Control Systems Magazine 36 (1) (2016) 26–46.

[22] W. Mei, S. Mohagheghi, S. Zampieri, F. Bullo, On the dynamics of deterministic epidemic propagation over networks, Annual Reviews in Control 44 (2017) 116–128.

[23] P. E. Paré, C. L. Beck, T. Başar, Modeling, estimation, and analysis of epidemics over networks: An overview, Annual Reviews in Control 50 (2020) 345–360.

[24] K. Paarporn, C. Eksin, J. S. Weitz, J. S. Shamma, Networked SIS epidemics with awareness, IEEE Transactions on Computational Social Systems 4 (3) (2017) 93–103.

[25] M. Alutto, L. Cianfanelli, G. Como, F. Fagnani, Multiple peaks in network SIR epidemic models, in: 2022 IEEE 61st Conference on Decision and Control (CDC), 2022, pp. 5614–5619.

[26] T. Kuniya, Y. Muroya, Global stability of a multi-group SIS epidemic model for population migration, Discrete and Continuous Dynamical Systems - B 19 (4) (2014) 1105–1118.

[27] L. Stella, A. P. Martínez, D. Bauso, P. Colaneri, The role of asymptomatic infections in the COVID-19 epidemic via complex networks and stability analysis, SIAM Journal on Control and Optimization 60 (2) (2022) S119–S144.

[28] H.-C. Chen, J. W. Friedman, J.-F. Thisse, Boundedly rational nash equilibrium: A probabilistic choice approach, Games and Economic Behavior 18 (1) (1997) 32–54.

[29] S. P. Anderson, J. K. Goeree, C. A. Holt, The logit equilibrium: A perspective on intuitive behavioral anomalies, Southern Economic Journal 69 (1) (2002) 21–47.

[30] J. C. Harsanyi, Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points, International Journal of Game Theory 2 (1) (1973) 1–23.

[31] R. W. Rosenthal, A bounded-rationality approach to the study of noncooperative games, International Journal of Game Theory 18 (3) (1989) 273–292.

[32] R. D. McKelvey, T. R. Palfrey, Quantal response equilibria for normal form games, Games and Economic Behavior 10 (1) (1995) 6–38.

[33] L. E. Blume, How noise matters, Games and Economic Behavior 44 (2) (2003) 251–271.

[34] J. K. Goeree, C. A. Holt, An experimental study of costly coordination, Games and Economic Behavior 51 (2) (2005) 349–364, special Issue in Honor of Richard D. McKelvey.

[35] J. C. Harsanyi, Games with incomplete information played by "Bayesian" players, I-III. Part II. Bayesian equilibrium points, Management Science 14 (5) (1968) 320–334.

[36] C. M. Capra, J. K. Goeree, R. Gomez, C. A. Holt, Anomalous behavior in a traveler's dilemma?, The American Economic Review 89 (3) (1999) 678–690.

This article has been accepted for publication in IEEE Transactions on Automatic Control. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAC.2025.3554601
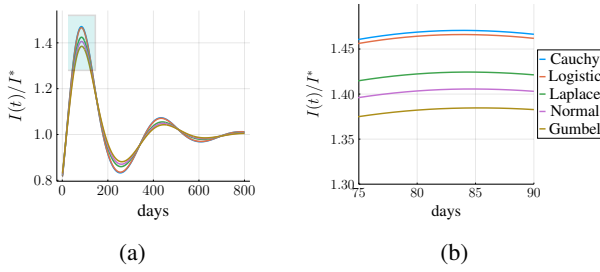
10



Fig. 4: The ratio $I(t)/I^*$ for (EPG) evaluated with different noise distributions for the choice function (4). The square area highlighted in (a) is presented at an enlarged scale in (b).

[37] S. Park, N. C. Martins, J. S. Shamma, From population games to payoff dynamics models: A passivity-based approac, in: 2019 IEEE 58th Conference on Decision and Control (CDC), 2019, pp. 6584–6601.

## ACKNOWLEDGEMENT

## APPENDIX

### A. Evaluation of (EPG) under Different Learning Rules for (PBR EDM)

Fig. 4 depicts the infectious fraction of the population when the probability distributions of the random variables $v_1, \cdots, v_n$ for (4) are defined by the Cauchy, logistic, Laplace, normal, and generalized extreme value (Gumbel) distributions. From the figure, we can observe that the trajectory of the infectious fraction does not vary substantially, even when the choice function is defined by different noise distributions.

### B. Proof of Theorem 1

To prove the theorem, we first introduce the following lemma.

*Lemma 2:* The function $\vec{\beta}'C(q\vec{\beta} + \bar{r} - \tilde{c}(\bar{I}))$ is increasing with respect to $q$.

*Proof:* To proceed, we establish that $\frac{\partial}{\partial q}\vec{\beta}'C(q\vec{\beta} + \bar{r} - \tilde{c}(\bar{I})) > 0, \ \forall q \in \mathbb{R}$. Note that

$$\frac{\partial}{\partial q}\vec{\beta}'C(q\vec{\beta} + \bar{r} - \tilde{c}(\bar{I}))$$
$$= \frac{\partial}{\partial q}\vec{\beta}' \arg\max_{z \in \text{int}(\mathbb{X})}(z'(q\vec{\beta} + \bar{r} - \tilde{c}(\bar{I})) - Q(z))$$
$$= \vec{\beta}'\nabla_p \arg\max_{z \in \text{int}(\mathbb{X})}(z'p - Q(z))\Big|_{p=q\vec{\beta}+\bar{r}-\tilde{c}(\bar{I})}\vec{\beta}. \quad (24)$$

By the same arguments used in the proof of [4, Theorem 2.1], we conclude that

$$\vec{\beta}'\nabla_p \arg\max_{z \in \text{int}(\mathbb{X})}(z'p - Q(z))\Big|_{p=q\vec{\beta}+\bar{r}-\tilde{c}(\bar{I})}\vec{\beta} > 0. \quad (25)$$

Therefore, $\vec{\beta}'C(q\vec{\beta} + \bar{r} - \tilde{c}(\bar{I}))$ is an increasing function of $q$ satisfying

$$\lim_{q \to \infty}\vec{\beta}'C(q\vec{\beta} + \bar{r} - \tilde{c}(\bar{I})) = \vec{\beta}_n \quad (26a)$$
$$\lim_{q \to -\infty}\vec{\beta}'C(q\vec{\beta} + \bar{r} - \tilde{c}(\bar{I})) = \vec{\beta}_1. \quad (26b)$$

We provide a three-part proof in which we establish $\lim_{t\to\infty}\|\dot{x}(t)\|_2 = \lim_{t\to\infty}|I(t) - \hat{I}(t)| = \lim_{t\to\infty}|R(t) - \hat{R}(t)| = 0$ in **Part 2** and $\lim_{t\to\infty}q(t) = \bar{q}$ and $\lim_{t\to\infty}\mathcal{B}(t) = \bar{\beta}$ in **Part 3**, where $\bar{q}$ is a unique solution to $\bar{\beta} = \vec{\beta}'C(\bar{q}\vec{\beta} + \bar{r} - \tilde{c}(\bar{I}))$. Consequently, in conjunction with $\lim_{t\to\infty}\|x(t) - C(q(t)\vec{\beta}+\bar{r}-\tilde{c}(\bar{I}))\|_2 = \lim_{t\to\infty}\|\dot{x}(t)\|_2 = 0$, we can conclude that the two statements of the theorem are valid.

***Part 1.*** $\mathcal{I}(t), \hat{\mathcal{I}}(t)$ *are strictly positive for all* $t \geq 0$: Adopting the same Lyapunov function $\mathscr{S}$ from [1], we can state

$$\mathscr{S}(\mathcal{I}, \mathcal{R}, \mathcal{B})$$
$$= (\mathcal{I} - \hat{\mathcal{I}}) + \hat{\mathcal{I}}\ln\frac{\hat{\mathcal{I}}}{\mathcal{I}} + \frac{1}{2\gamma}(\mathcal{R} - \hat{\mathcal{R}})^2 + \frac{v^2}{2}(\mathcal{B} - \bar{\beta})^2$$
$$= (\mathcal{I} - \eta(\mathcal{B} - \sigma)) + \eta(\mathcal{B} - \sigma)\ln\frac{\eta(\mathcal{B} - \sigma)}{\mathcal{I}}$$
$$+ \frac{1}{2\gamma}(\mathcal{R} - (1-\eta)(\mathcal{B}-\sigma))^2 + \frac{v^2}{2}(\mathcal{B} - \bar{\beta})^2, \quad (27)$$

where we use $\hat{\mathcal{I}} = \mathcal{B}\hat{I} = \eta(\mathcal{B} - \sigma)$ and $\hat{\mathcal{R}} = \mathcal{B}\hat{R} = (1-\eta)(\mathcal{B} - \sigma)$. Note that using (EPG) and (PBR EDM), we can derive the following relation.

$$\kappa^{-1}\frac{d}{dt}\mathcal{S}(x(t), p(t)) + \frac{d}{dt}\mathscr{S}(\mathcal{I}(t), \mathcal{R}(t), \mathcal{B}(t))$$
$$= \kappa^{-1}\nabla'_x\mathcal{S}(x(t), p(t))\mathcal{V}(x(t), p(t))$$
$$- (\mathcal{I}(t) - \hat{\mathcal{I}}(t))^2 - \frac{\omega}{\gamma}(\mathcal{R}(t) - \hat{\mathcal{R}}(t))^2 \leq 0, \quad (28)$$

which implies, since $\mathcal{S}$ and $\mathscr{S}$ are non-negative functions, $\mathscr{S}$ is a bounded function – both below and above. In conjunction with the fact that $\hat{\mathcal{I}} = \eta(\mathcal{B} - \sigma) \geq \eta(\mathcal{B}_{\min} - \sigma) > 0$, we conclude that $\mathcal{I}(t)$ is strictly positive for all $t \geq 0$, i.e., there is $\delta > 0$ for which $\mathcal{I}(t) \geq \delta, \ \forall t \geq 0$ holds. Otherwise, $\eta(\mathcal{B}(t) - \sigma)\ln\frac{\eta(\mathcal{B}(t)-\sigma)}{\mathcal{I}(t)}$ tends to infinity and so does $\mathscr{S}(\mathcal{I}(t), \mathcal{R}(t), \mathcal{B}(t))$ as $\mathcal{I}(t)$ approaches zero, which contradicts the fact that $\mathscr{S}$ is a bounded function.

We remark that from (EPGc) and the definition of the mapping $G$, since both $\mathcal{I}(t)$ and $\hat{\mathcal{I}}(t)$ are strictly positive for all $t \geq 0$ and also upper bounded according to their respective definitions, we can infer that $\dot{q}(t)$ is bounded.

***Part 2.*** *It holds that* $\lim_{t\to\infty}\|\dot{x}(t)\|_2 = \lim_{t\to\infty}|I(t) - \hat{I}(t)| = \lim_{t\to\infty}|R(t) - \hat{R}(t)| = 0$: For notational convenience, let us define

$$\breve{\mathscr{S}}(\mathcal{I}, \mathcal{R}, \mathcal{B}) = (\mathcal{I} - \hat{\mathcal{I}}) + \hat{\mathcal{I}}\ln\frac{\hat{\mathcal{I}}}{\mathcal{I}} + \frac{1}{2\gamma}(\mathcal{R} - \hat{\mathcal{R}})^2. \quad (29)$$

Note that $0 \leq \breve{\mathscr{S}}(\mathcal{I}, \mathcal{R}, \mathcal{B}) \leq \mathscr{S}(\mathcal{I}, \mathcal{R}, \mathcal{B})$. According to (28) and (29), we can derive

$$\kappa^{-1}\mathcal{S}(x(t), p(t)) + \breve{\mathscr{S}}(\mathcal{I}(t), \mathcal{R}(t), \mathcal{B}(t)) - \alpha$$
$$\leq \kappa^{-1}\mathcal{S}(x(t), p(t)) + \mathscr{S}(\mathcal{I}(t), \mathcal{R}(t), \mathcal{B}(t)) - \alpha$$
$$= \int_0^t \left(\kappa^{-1}\nabla'_x\mathcal{S}(x(\tau), p(\tau))\mathcal{V}(x(\tau), p(\tau))\right.$$
$$\left. - (\mathcal{I}(\tau) - \hat{\mathcal{I}}(\tau))^2 - \frac{\omega}{\gamma}(\mathcal{R}(\tau) - \hat{\mathcal{R}}(\tau))^2\right) d\tau, \ \forall t \geq 0, \quad (30)$$

where $\alpha = \kappa^{-1}\mathcal{S}(x(0), p(0)) + \mathscr{S}(\mathcal{I}(0), \mathcal{R}(0), \mathcal{B}(0))$. We claim that

$$\lim_{t\to\infty}\left(\kappa^{-1}\mathcal{S}(x(t), p(t)) + \breve{\mathscr{S}}(\mathcal{I}(t), \mathcal{R}(t), \mathcal{B}(t))\right) = 0, \quad (31)$$

which implies

$$\lim_{t\to\infty}\mathcal{S}(x(t),p(t))=0$$

$$\lim_{t\to\infty}|I(t)-\hat{I}(t)|\le\lim_{t\to\infty}\frac{1}{\bar{\beta}_1}|\mathcal{I}(t)-\hat{\mathcal{I}}(t)|=0$$

$$\lim_{t\to\infty}|R(t)-\hat{R}(t)|\le\lim_{t\to\infty}\frac{1}{\bar{\beta}_1}|\mathcal{R}(t)-\hat{\mathcal{R}}(t)|=0.$$

Furthermore, according to the analysis used in [4, Theorem 2.1],

$$\tilde{z}'\nabla Q(y)=\tilde{z}'r\iff y=\arg\max_{z\in\text{int}(\mathbb{X})}(z'r-Q(z))\qquad(32)$$

holds for all $r\in\mathbb{R}^n$, $y\in\text{int}(\mathbb{X})$, and $\tilde{z}\in T\mathbb{X}$. Let $y(t)=\arg\max_{z\in\text{int}(\mathbb{X})}(z'p(t)-Q(z))$. Then, we can derive

$$\mathcal{S}(x(t),p(t))=(y(t)-x(t))'\nabla Q(y(t))$$
$$-(Q(y(t))-Q(x(t))).\qquad(33)$$

Therefore, by the strict convexity (7a) of $Q$ and (33), $\lim_{t\to\infty}\mathcal{S}(x(t),p(t))=0$ implies that $\lim_{t\to\infty}\|\dot{x}(t)\|_2=0$.

In what follows, we justify the claim. Let us define the following set $\mathbb{O}_\epsilon$:

$$\mathbb{O}_\epsilon=\left\{t>0\ \middle|\ \kappa^{-1}\mathcal{S}(x(t),p(t))+\breve{\mathscr{S}}(\mathcal{I}(t),\mathcal{R}(t),\mathcal{B}(t))>\frac{\epsilon}{2}\right\}.$$

Since $\mathcal{S}$ and $\breve{\mathscr{S}}$ are differentiable, $\mathbb{O}_\epsilon$ is an open set and can be represented as an infinite union of disjoint open intervals, i.e., $\mathbb{O}_\epsilon=\cup_{i=1}^\infty\mathbb{I}_i$ where $\mathbb{I}_i=(a_i,b_i)$ is an open interval. Note that

$$\kappa^{-1}\mathcal{S}(x(a_i),p(a_i))+\breve{\mathscr{S}}(\mathcal{I}(a_i),\mathcal{R}(a_i),\mathcal{B}(a_i))=\frac{\epsilon}{2}.\qquad(34)$$

The $\delta$-storage function of (PBR EDM) satisfies

$$\mathcal{S}(x,p)=y'p-Q(y)-(x'p-Q(x))$$
$$=p'(y-x)-(Q(y)-Q(x))$$
$$\overset{(i)}{\le}(p-\nabla Q(x))'(y-x)$$
$$=-\nabla_x'\mathcal{S}(x,p)\mathcal{V}(x,p),\qquad(35)$$

where $y=\arg\max_{z\in\text{int}(\mathbb{X})}(z'p-Q(z))$. For $(i)$, we use the convexity (7a) of $Q$ to establish $Q(y)-Q(x)\ge\nabla'Q(x)(y-x)$. Also, by **Part 1**, since $\mathcal{I},\hat{\mathcal{I}}$ are strictly positive, i.e., there is $\delta>0$ for which $\mathcal{I},\hat{\mathcal{I}}\ge\delta$ holds, we can find a positive constant $k$ satisfying the inequality given by

$$k|\mathcal{I}-\hat{\mathcal{I}}|\ge(\mathcal{I}-\hat{\mathcal{I}})+\hat{\mathcal{I}}\ln\frac{\hat{\mathcal{I}}}{\mathcal{I}}.\qquad(36)$$

For instance, (36) holds with $k$ defined as

$$k=1+\eta(\vec{\beta}_n-\sigma)\max_{\mathcal{I},\hat{\mathcal{I}}\ge\delta}\frac{\ln\hat{\mathcal{I}}-\ln\mathcal{I}}{\hat{\mathcal{I}}-\mathcal{I}}.$$

Consequently, by (35) and (36), for every $t\in\mathbb{O}_\epsilon$, there is a constant $\delta_\epsilon>0$ for which

$$\kappa^{-1}\nabla_x'\mathcal{S}(x(t),p(t))\mathcal{V}(x(t),p(t))$$
$$-(\mathcal{I}(t)-\hat{\mathcal{I}}(t))^2-\frac{\omega}{\gamma}(\mathcal{R}(t)-\hat{\mathcal{R}}(t))^2<-\delta_\epsilon.\qquad(37)$$

Since $\kappa^{-1}\mathcal{S}(x(t),p(t))+\breve{\mathscr{S}}(\mathcal{I}(t),\mathcal{R}(t),\mathcal{B}(t))$ is a non-negative function of $t$, from (30) and (37), we can derive

$$-\alpha\le\int_0^\infty\left(\kappa^{-1}\nabla_x'\mathcal{S}(x(\tau),p(\tau))\mathcal{V}(x(\tau),p(\tau))\right.$$
$$\left.-(\mathcal{I}(\tau)-\hat{\mathcal{I}}(\tau))^2-\frac{\omega}{\gamma}(\mathcal{R}(\tau)-\hat{\mathcal{R}}(\tau))^2\right)\mathrm{d}\tau$$
$$\le\int_{\mathbb{O}_\epsilon}\left(\kappa^{-1}\nabla_x'\mathcal{S}(x(\tau),p(\tau))\mathcal{V}(x(\tau),p(\tau))\right.$$
$$\left.-(\mathcal{I}(\tau)-\hat{\mathcal{I}}(\tau))^2-\frac{\omega}{\gamma}(\mathcal{R}(\tau)-\hat{\mathcal{R}}(\tau))^2\right)\mathrm{d}\tau$$
$$\le-\delta_\epsilon\mathcal{L}(\mathbb{O}_\epsilon),\qquad(38)$$

where $\mathcal{L}(\mathbb{O}_\epsilon)$ is the Lebesgue measure of $\mathbb{O}_\epsilon$. Therefore, the set $\mathbb{O}_\epsilon$ has finite Lebesgue measure which implies that $\lim_{i\to\infty}|b_i-a_i|=0$.

To complete the proof of the claim, we argue that for every $\epsilon>0$, there is $T_\epsilon>0$ for which it holds that

$$\kappa^{-1}\mathcal{S}(x(t),p(t))+\breve{\mathscr{S}}(\mathcal{I}(t),\mathcal{R}(t),\mathcal{B}(t))<\epsilon,\ \forall t\ge T_\epsilon.\qquad(39)$$

If the argument does not hold, then for some $\epsilon>0$, we can find an infinite subset $\mathbb{J}$ of $\mathbb{N}$ for which

$$\max_{t\in\bar{\mathbb{I}}_j}\left(\kappa^{-1}\mathcal{S}(x(t),p(t))+\breve{\mathscr{S}}(\mathcal{I}(t),\mathcal{R}(t),\mathcal{B}(t))\right)\ge\epsilon\qquad(40)$$

holds for all $j$ in $\mathbb{J}$, where $\bar{\mathbb{I}}_j=[a_j,b_j]$ is the closure of the open subset $\mathbb{I}_j=(a_j,b_j)$ of $\mathbb{O}_\epsilon$. Let $\bar{t}_j$ be a time instant in $\bar{\mathbb{I}}_j$ attaining $\max_{t\in\bar{\mathbb{I}}_j}(\kappa^{-1}\mathcal{S}(x(t),p(t))+\breve{\mathscr{S}}(\mathcal{I}(t),\mathcal{R}(t),\mathcal{B}(t)))$. We can derive the following relation.

$$\kappa^{-1}\mathcal{S}(x(\bar{t}_j),p(\bar{t}_j))+\breve{\mathscr{S}}(\mathcal{I}(\bar{t}_j),\mathcal{R}(\bar{t}_j),\mathcal{B}(\bar{t}_j))$$
$$-\kappa^{-1}\mathcal{S}(x(a_j),p(a_j))-\breve{\mathscr{S}}(\mathcal{I}(a_j),\mathcal{R}(a_j),\mathcal{B}(a_j))$$
$$=\int_{a_j}^{\bar{t}_j}\frac{\mathrm{d}}{\mathrm{d}\tau}\left(\kappa^{-1}\mathcal{S}(x(\tau),p(\tau))+\breve{\mathscr{S}}(\mathcal{I}(\tau),\mathcal{R}(\tau),\mathcal{B}(\tau))\right)\mathrm{d}\tau$$
$$\overset{(i)}{\le}-\int_{a_j}^{\bar{t}_j}v^2\dot{\mathcal{B}}(\tau)(\mathcal{B}(\tau)-\bar{\beta})\mathrm{d}\tau$$
$$\le\int_{a_j}^{\bar{t}_j}v^2\|\vec{\beta}\|_2\|\dot{x}(\tau)\|_2(\|\vec{\beta}\|_2\|x(\tau)\|_2+\bar{\beta})\mathrm{d}\tau.\qquad(41)$$

To obtain $(i)$, we use (28) and the definition of $\breve{\mathscr{S}}$. Since both $\dot{x}$ and $x$ are bounded, we can find a constant $M$ for which the following inequality holds.

$$\kappa^{-1}\mathcal{S}(x(\bar{t}_j),p(\bar{t}_j))+\breve{\mathscr{S}}(\mathcal{I}(\bar{t}_j),\mathcal{R}(\bar{t}_j),\mathcal{B}(\bar{t}_j))$$
$$-\kappa^{-1}\mathcal{S}(x(a_j),p(a_j))-\breve{\mathscr{S}}(\mathcal{I}(a_j),\mathcal{R}(a_j),\mathcal{B}(a_j))$$
$$\le M(\bar{t}_j-a_j).\qquad(42)$$

On the other hand, since $\lim_{j\to\infty}|\bar{t}_j-a_j|\le\lim_{j\to\infty}|b_j-a_j|=0$, for sufficiently large index $j$ in $\mathbb{J}$, we have that

$$\kappa^{-1}\mathcal{S}(x(\bar{t}_j),p(\bar{t}_j))+\breve{\mathscr{S}}(\mathcal{I}(\bar{t}_j),\mathcal{R}(\bar{t}_j),\mathcal{B}(\bar{t}_j))$$
$$<\kappa^{-1}\mathcal{S}(x(a_j),p(a_j))+\breve{\mathscr{S}}(\mathcal{I}(a_j),\mathcal{R}(a_j),\mathcal{B}(a_j))+\frac{\epsilon}{2}=\epsilon,$$

which contradicts (40). This completes **Part 2**.

**Part 3.** *It holds that $\lim_{t\to\infty}q(t)=\bar{q}$ and $\lim_{t\to\infty}\mathcal{B}(t)=\bar{\beta}$:* Since $p(t)=r(t)-\tilde{c}(I(t))=q(t)\vec{\beta}+\bar{r}-\tilde{c}(\bar{I})$, we can re-write

This article has been accepted for publication in IEEE Transactions on Automatic Control. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAC.2025.3554601

12

$$\dot{q}(t) = \kappa v^2 \big(\bar{\beta} - \vec{\beta}' C(q(t)\vec{\beta} + \bar{r} - \tilde{c}(\bar{I}))\big)$$
$$+ \underbrace{\kappa v^2 \big(\vec{\beta}' C(q(t)\vec{\beta} + \bar{r} - \tilde{c}(\bar{I})) - \mathcal{B}(t)\big) + \kappa(\hat{I}(t) - I(t)) + \kappa\eta(\ln I(t) - \ln\hat{I}(t)) + \kappa\frac{\mathcal{B}(t)}{\gamma}(R(t) - \hat{R}(t))(1 - \eta - R(t))}_{=\epsilon(t)} \quad (43)$$

$$\begin{pmatrix} \dot{\tilde{I}}(t) \\ \dot{\tilde{R}}(t) \\ \dot{\tilde{q}}(t) \\ \dot{\tilde{x}}(t) \end{pmatrix} = \begin{pmatrix} -\eta(\bar{\beta} - \sigma) & -\eta(\bar{\beta} - \sigma) & 0 & \eta\left(1 - \frac{\sigma}{\bar{\beta}}\right)\frac{\sigma}{\bar{\beta}}\vec{\beta}' \\ \gamma & -\omega & 0 & 0 \\ \frac{\kappa\sigma}{\bar{\beta}-\sigma} & \frac{\kappa\sigma}{\gamma}(1-\eta) & 0 & -\frac{\kappa\sigma^2}{\bar{\beta}^2}\left(\frac{\eta}{\bar{\beta}-\sigma} + \frac{(1-\eta)^2}{\gamma}\right)\vec{\beta}' - \kappa v^2\vec{\beta}' \\ 0 & 0 & d & -I_n \end{pmatrix} \begin{pmatrix} \tilde{I}(t) \\ \tilde{R}(t) \\ \tilde{q}(t) \\ \tilde{x}(t) \end{pmatrix} \quad (44)$$

$$\begin{pmatrix} \dot{\tilde{I}}(t) \\ \dot{\tilde{R}}(t) \\ \dot{\tilde{q}}(t) \\ \dot{\tilde{\mathcal{B}}}(t) \end{pmatrix} = \underbrace{\begin{pmatrix} -\eta(\bar{\beta} - \sigma) & -\eta(\bar{\beta} - \sigma) & 0 & \eta\left(1 - \frac{\sigma}{\bar{\beta}}\right)\frac{\sigma}{\bar{\beta}} \\ \gamma & -\omega & 0 & 0 \\ \frac{\kappa\sigma}{\bar{\beta}-\sigma} & \frac{\kappa\sigma}{\gamma}(1-\eta) & 0 & -\frac{\kappa\sigma^2}{\bar{\beta}^2}\left(\frac{\eta}{\bar{\beta}-\sigma} + \frac{(1-\eta)^2}{\gamma}\right) - \kappa v^2 \\ 0 & 0 & \vec{\beta}'d & -1 \end{pmatrix}}_{=A} \begin{pmatrix} \tilde{I}(t) \\ \tilde{R}(t) \\ \tilde{q}(t) \\ \hat{\mathcal{B}}(t) \end{pmatrix}, \quad (45)$$

where $\tilde{I}(t) = I(t) - \bar{I}$, $\tilde{R}(t) = R(t) - \bar{R}$, $\tilde{q}(t) = q(t) - \bar{q}$, $\tilde{x}(t) = x(t) - \bar{x}$, $\tilde{\mathcal{B}}(t) = \vec{\beta}'\tilde{x}(t)$, and $d = (d_1 \cdots d_n)'$ with $d_i = \bar{x}_i(\vec{\beta}_i - \bar{\beta})/\mu$.

$$\det(A - \lambda I) = \lambda^4 + \big(1 + \omega + \eta(\bar{\beta} - \sigma)\big)\lambda^3 + \left(\omega + (\eta + \omega)(\bar{\beta} - \sigma) + \vec{\beta}'d\left(\frac{\kappa\sigma^2}{\bar{\beta}^2}\left(\frac{\eta}{\bar{\beta} - \sigma} + \frac{(1-\eta)^2}{\gamma}\right) + \kappa v^2\right)\right)\lambda^2$$
$$+ \left(\omega(\bar{\beta} - \sigma) + \kappa(\omega + \eta(\bar{\beta} - \sigma))\vec{\beta}'dv^2 + \omega\vec{\beta}'d\frac{\kappa\sigma^2}{\bar{\beta}^2}\frac{\eta}{\bar{\beta} - \sigma} + \eta(\bar{\beta} - \sigma)\vec{\beta}'d\frac{\kappa\sigma^2}{\bar{\beta}^2}\frac{(1-\eta)^2}{\gamma}\right)\lambda + \kappa\vec{\beta}'d\omega(\bar{\beta} - \sigma)v^2 = 0. \quad (46)$$

(EPGc) as in (43). By **Part 2**, we observe that

$$\lim_{t\to\infty}\big(\vec{\beta}'C(q(t)\vec{\beta} + \bar{r} - \tilde{c}(\bar{I})) - \mathcal{B}(t)\big) = \lim_{t\to\infty}\vec{\beta}'\dot{x}(t) = 0$$
$$\lim_{t\to\infty}|I(t) - \hat{I}(t)| = 0$$
$$\lim_{t\to\infty}|R(t) - \hat{R}(t)| = 0,$$

which, in conjunction with the fact that $\max_{\mathcal{I},\hat{\mathcal{I}}\geq\delta}\frac{\ln\hat{\mathcal{I}}-\ln\mathcal{I}}{\hat{\mathcal{I}}-\mathcal{I}} > 0$ is finite, imply that $\lim_{t\to\infty}\epsilon(t) = 0$, where $\epsilon(t)$ is defined in (43). Hence, if $q(t)$ goes to $\bar{q}$, satisfying $\bar{\beta} = \vec{\beta}'C(\bar{q}\vec{\beta}+\bar{r}-\tilde{c}(\bar{I}))$, as $t$ tends to infinity, then we can infer that $\lim_{t\to\infty}\mathcal{B}(t) = \vec{\beta}'C(\bar{q}\vec{\beta} + \bar{r} - \tilde{c}(\bar{I})) = \bar{\beta}$.

Using Lemma 2, we can infer that there is $\bar{q}$ for which $\bar{\beta} = \vec{\beta}'C(\bar{q}\vec{\beta}+\bar{r}-\tilde{c}(\bar{I}))$ holds and together with $\lim_{t\to\infty}\epsilon(t) = 0$, $q(t)$ converges to $\bar{q}$ as $t$ tends to infinity. In particular, if $\bar{\beta} = \vec{\beta}'C(\bar{r} - \tilde{c}(\bar{I}))$ then $\bar{q} = 0$. This completes the proof. ∎

### C. Proof of Proposition 1

We first prove the Lyapunov stability of the closed-loop model. To this end, we show that

$$\kappa^{-1}\mathcal{S}(x, p) + \mathscr{S}(\mathcal{I}, \mathcal{R}, \mathcal{B}) = 0 \quad (47)$$

holds if and only if

$$I = \bar{I} \quad (48a)$$
$$R = \bar{R} \quad (48b)$$
$$q = \bar{q} \quad (48c)$$
$$x = \bar{x}, \quad (48d)$$

hold, where $\bar{x} = C(\bar{q}\vec{\beta}+\bar{r}-\tilde{c}(\bar{I}))$. Since the "if" part directly follows from the definitions of $\mathcal{S}$ and $\mathscr{S}$, we discuss the proof of the "only if" part.

Eqs. (48a) and (48b) follow from the definition of $\mathscr{S}$. Also, from the definition of $\mathscr{S}$, we can infer that $\mathcal{B} = \bar{\beta}$. Recall that, by the strict convexity (7a) of $Q$ and (33), $\mathcal{S}(x, p) = 0$ implies $C(q\vec{\beta} + \bar{r} - \tilde{c}(\bar{I})) = x$. Consequently, by the fact that $\vec{\beta}'C(q\vec{\beta} + \bar{r} - \tilde{c}(\bar{I}))$ is an increasing function of $q$ as shown in Lemma 2, we can infer that (48c) and (48d) hold. As verified in the proof of Theorem 1, $\frac{d}{dt}(\kappa^{-1}\mathcal{S}(x(t), p(t)) + \mathscr{S}(\mathcal{I}(t), \mathcal{R}(t), \mathcal{B}(t))) \leq 0$ holds along the solution trajectory of the closed-loop model. Therefore, we conclude that the closed-loop model is Lyapunov stable. Additionally, in conjunction with the results of Theorem 1, we further conclude that the closed-loop model is asymptotically stable.

By linearizing the closed-loop model around $(\bar{I}, \bar{R}, \bar{q}, \bar{x})$, we can derive (44). Note that all eigenvalues of (44) have negative real parts if the reduced model (45) also exhibits this property. To conclude local exponential stability, given the model's asymptotic stability, it is sufficient to show that (45) does not have any eigenvalues on the imaginary axis.

Consider the characteristic polynomial (46) for (45). When $v = 0$, (46) has one solution at the origin, and we claim that the remaining three non-zero solutions are not on the imaginary axis. Conversely, when $v > 0$, (46) has no solution at the origin. Since the solutions of (46) continuously depend on $v$, we conclude that all the eigenvalues of (45) are off the imaginary axis for small $v > 0$. To prove the claim, consider

$\lambda = jb$, where $b$ is a nonzero real number. By substituting $\lambda = jb$ in (46) with $\upsilon = 0$, we obtain the following two equations:

$$b^2 = \left(\omega + (\eta + \omega)(\bar{\beta} - \sigma) + \vec{\beta}'d\frac{\kappa\sigma^2}{\bar{\beta}^2}\left(\frac{\eta}{\bar{\beta} - \sigma} + \frac{(1-\eta)^2}{\gamma}\right)\right) \tag{49a}$$

$$\left(1 + \omega + \eta(\bar{\beta} - \sigma)\right)b^2 = \omega(\bar{\beta} - \sigma) + \omega\vec{\beta}'d\frac{\kappa\sigma^2}{\bar{\beta}^2}\frac{\eta}{\bar{\beta} - \sigma}$$
$$+ \eta(\bar{\beta} - \sigma)\vec{\beta}'d\frac{\kappa\sigma^2}{\bar{\beta}^2}\frac{(1-\eta)^2}{\gamma}. \tag{49b}$$

Note that $\vec{\beta}'d \geq 0$. However, from (49a), we can infer that

$$\left(1 + \omega + \eta(\bar{\beta} - \sigma)\right)b^2 > \omega(\bar{\beta} - \sigma) + \omega\vec{\beta}'d\frac{\kappa\sigma^2}{\bar{\beta}^2}\frac{\eta}{\bar{\beta} - \sigma}$$
$$+ \eta(\bar{\beta} - \sigma)\vec{\beta}'d\frac{\kappa\sigma^2}{\bar{\beta}^2}\frac{(1-\eta)^2}{\gamma},$$

which contradicts (49b). Consequently, we conclude no eigenvalues lie on the imaginary axis. ∎

### D. Proof of Lemma 1

Noting that

$$(r - \tilde{c}(I))'C^\mu(r - \tilde{c}(I))$$
$$= (r - \tilde{c}(I))' \arg\max_{z \in \text{int}(\mathbb{X})}(\lambda z'(r - \tilde{c}(I)) - \bar{Q}(z)),$$

where $\lambda = \mu^{-1}$, the proof of this lemma is similar to that of Lemma 2. We omit details for brevity. ∎

### E. Proof of Proposition 2

With $r = \tilde{r} + \tilde{c}(I)$, let $\mathbf{R}$ be a discrete random variable whose probability is defined as $\mathbb{P}(\mathbf{R} = \tilde{r}_i) = C_i^\mu(\tilde{r})$. Note that $\mathbb{E}[\mathbf{R}] = \tilde{r}'C^\mu(\tilde{r})$ holds. After surveying $K$ randomly selected agents, using the answers collected, we can compute $K$ realizations $\{R^{(k)}\}_{k=1}^K$ of $\mathbf{R}$ with $R^{(k)} = \tilde{r}'\xi_k$. Then, by applying Chebyshev's inequality, we can establish

$$\mathbb{P}\left(\left|\mathbb{E}[\mathbf{R}] - \frac{1}{K}\sum_{k=1}^K R^{(k)}\right| \leq \epsilon\right) \geq 1 - 1/\epsilon^2 K, \tag{50}$$

where, to establish the inequality, we use (19) and Popoviciu's inequality. The inequality (50) explains how to find both an upper and a lower bounds of $\tilde{r}'C^\mu(\tilde{r})$ using the sample mean, with an arbitrarily high level of confidence. Consequently, using Lemma 1, we can compute bounds $\mu_L, \mu_R$ on $\mu$ as in (20), leading to (21). ∎

### F. Proof of Proposition 3

Since the budget constraint (P2) stated in **Main Problem** is relevant only when the limit of reward $r(t)$ has all non-negative entries, as explained in Remark 1, we proceed with using (13) for the definition of $H$ in (EPGd). According to Theorem 1, with $\bar{r} = \tilde{c}(\bar{I})$, $q(t)$ converges to $\bar{q}$ for which $\bar{\beta} = \vec{\beta}'C^\mu(\bar{q}\vec{\beta})$ holds and, hence, $r(t)$ converges to $\bar{q}\vec{\beta} + \tilde{c}(\bar{I})$.

By defining $\lambda = \mu^{-1}\bar{q}$, we can express

$$\bar{\beta} = \vec{\beta}'C^\mu(\bar{q}\vec{\beta}) = \vec{\beta}' \arg\max_{z \in \text{int}(\mathbb{X})}(z'(\lambda\vec{\beta}) - \bar{Q}(z)). \tag{51}$$

By the same argument used in the proof of Lemma 1, $\vec{\beta}' \arg\max_{z \in \text{int}(\mathbb{X})}(z'(\lambda\vec{\beta}) - \bar{Q}(z))$ is an increasing function of $\lambda$. Also, if $\bar{\beta}$ satisfies $\bar{\beta} < \vec{\beta}'C^1(0)$ then it holds that $\lambda < 0$ and $\bar{q} < 0$.
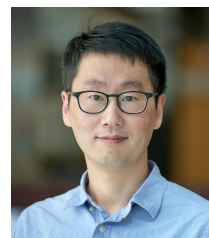
When (EPG) and (PBR EDM) reach their equilibrium state, the planner would spend

$$(\bar{q}\vec{\beta} + \tilde{c}(\bar{I}) - \bar{q}\vec{\beta}_n\mathbf{1})' \arg\max_{z \in \text{int}(\mathbb{X})}(z'(\bar{q}\vec{\beta}) - \mu\bar{Q}(z))$$
$$= \mu\lambda(\bar{\beta} - \vec{\beta}_n) + \tilde{c}'(\bar{I}) \arg\max_{z \in \text{int}(\mathbb{X})}(z'(\lambda\vec{\beta}) - \bar{Q}(z)), \tag{52}$$

which is upper bounded by

$$\mu_U\lambda(\bar{\beta} - \vec{\beta}_n) + \tilde{c}'(\bar{I}) \arg\max_{z \in \text{int}(\mathbb{X})}(z'(\lambda\vec{\beta}) - \bar{Q}(z)). \tag{53}$$

This completes the proof. ∎

**Shinkyu Park** (Member, IEEE) received the Ph.D. degree from the University of Maryland, College Park, MD, USA, in 2015, all in electrical engineering. From 2016 to 2019, he was a Postdoctoral Associate at Massachusetts Institute of Technology, and from 2019 to 2021, he was appointed as an Associate Research Scholar at Princeton University. He is currently the Assistant Professor of Electrical and Computer Engineering at King Abdullah University of Science and Technology.

**Jair Certorio** (Student Member, IEEE), received his B.S. degree in Electrical Engineering from Universidade Estadual de Maringá, Brazil, in 2018. He is currently working towards a Ph.D. degree in Electrical and Computer Engineering from the University of Maryland, College Park.

**Nuno C. Martins** (Senior Member, IEEE) received a Ph.D. degree in Electrical Engineering and Computer Science with a minor in Mathematics from Massachusetts Institute of Technology (MIT), Cambridge. He is Professor of Electrical Engineering in the Electrical and Computer Engineering Department of the University of Maryland at College Park, where he also holds a joint appointment with the Institute for Systems Research. He was Director of the Maryland Robotics Center from 2012 until 2014. He received a 2006 AACC O. Hugo Schuck Award, a National Science Foundation CAREER Award in 2007 and a 2008 IEEE CSS Axelby Award. He currently serves as an associate editor for the IEEE Transactions on Control of Network Systems.

**Richard J. La** received his PhD from the University of California, Berkeley in 2000. He is on the faculty of the department of electrical and computer engineering at the University of Maryland, where he is currently a Professor.