

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 vary according to the 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 population's 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 use the Lyapunov function to construct anytime upper bounds for the size of the population's infectious fraction. 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, modeled by simple learning rules (protocols or rules for short) used by the population's agents, and do not need detailed knowledge of these rules. Yet, they assume that the agents know exactly 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 $\{1, \dots, n\}$, $n \geq 2$, 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 perturbed best response 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-C).

The design methods from [1], [2] are unsuitable here because they depend on an assumption of no noise. Our approach, while also based on system-theoretic passivity as in [1], [2], requires substantial changes in the stability proofs and in the characterization of the set of endemic equilibria with minimized infection prevalence, which here depends on the noise distribution. For an important class of PBR rules that we will later describe, this distribution hinges on a single parameter, which we will be able to learn from data.

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

$$p(t) := r(t) - c, \quad (1)$$

where c is the vector whose ℓ -th entry c_ℓ is the inherent cost of the ℓ -th strategy, and $r(t)$ is a reward vector meant to incentivize the adoption of safer (costlier) strategies, where $r_\ell(t)$ is the ℓ -th strategy's reward.

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

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

where $x(t)$ is the so-called *population state* taking values in the standard simplex \mathbb{X} defined below and whose ℓ -th entry

Corresponding author: N. C. Martins. The work of Park was supported by funding from King Abdullah University of Science and Technology (KAUST). This work was supported by AFOSR Grant FA9550-19-1-0315 and NSF Grant 2135561.

Shinkyu Park is with the CEMSE, at King Abdullah University of Science and Technology. (e-mail: shinkyu.park@kaust.edu.sa).

Jair Certorio, Nuno C. Martins, and Richard J. La are with the Dept. of ECE and ISR, at the University of Maryland, at College Park. (e-mails: certorio@umd.edu, nmartins@umd.edu, hyongla@umd.edu).

$x_\ell(t)$ represents the proportion of the population adopting the ℓ -th strategy at time t .

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

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:

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

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

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

A Lipschitz continuous map $\mathcal{T} : \mathbb{X} \times \mathbb{R}^n \rightarrow [0, 1]^{n \times n}$ is referred to as *the learning rule (or revision protocol)* and models the agents' strategy revision preferences. In [6, Part II] and [7, §13.3-13.5], there is a comprehensive discussion on learning rule types.

B. Perturbed Best Response Learning Rule

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 assume

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

where v_1, \dots, v_n are random variables admitting a positive joint probability density function over \mathbb{R}^n . In particular, when v_1, \dots, v_n are i.i.d. Gaussian random variables, (4) is referred to as the probit learning rule [9]. In the following example, we discuss another learning rule that has been widely studied in the literature.

Example 1: A well-known example of a PBR learning rule is the logit learning rule [5] specified as

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

where $\mu > 0$ quantifies the noise intensity. Namely, v_1, \dots, 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. Interestingly, $C(p) = (C_1(p), \dots, C_n(p))$ is the gradient of the softmax function $\mu \ln(e^{p_1/\mu} + \dots + e^{p_n/\mu})$. In the noise-free limit as μ tends to zero, C will tend to the best response learning rule.

According to [4], we can represent the choice function as¹

$$C(p) = (C_1(p), \dots, C_n(p)) = \arg \max_{z \in \text{int}(\mathbb{X})} (z'p - Q(z)), \quad (5)$$

¹Since Q satisfies (6), according to the analysis presented in [4], the maximization in (5) admits a unique solution.

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

$$\tilde{z}' \nabla^2 Q(z) \tilde{z} > 0, \quad z \in \mathbb{X}, \tilde{z} \in T\mathbb{X} \setminus \{0\} \quad (6a)$$

$$\lim_{z_{\min} \rightarrow 0} \|\nabla Q(z)\|_2 = \infty, \quad \text{where } z_{\min} = \min_{1 \leq i \leq n} z_i, \quad (6b)$$

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 (5), 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 $x^* = C(p^*)$.

C. Epidemic Population Game

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

$$\dot{I}(t) = (\mathcal{B}(t)(1 - I(t) - R(t)) - \sigma)I(t), \quad (\text{EPGa})$$

$$\dot{R}(t) = \gamma I(t) - \omega R(t), \quad (\text{EPGb})$$

$$\dot{q}(t) = G(I(t), R(t), x(t), q(t)), \quad (\text{EPGc})$$

$$r(t) = H(I(t), R(t), x(t), q(t)), \quad (\text{EPGd})$$

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. Specifically, they are the numbers of infectious, recovered, and susceptible individuals at time t divided by the population size. Here, (EPGa,b) is a normalized SIRS model with $\sigma := \gamma + \theta$ and $\omega := \psi + \theta$, where γ and ψ denote the daily recovery rate and the daily rate at which recovered individuals become susceptible (due to waning immunity), respectively, and θ is the daily birth rate. We assume that newborns are susceptible. In this study, for the simplicity of our discussion we do not explicitly model the deaths due to the disease or natural deaths (unrelated to the disease). We refer a reader interested in learning how disease-induced or natural deaths can be handled to the study in [2].

Finally, (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 $m \geq 1$.

D. Problem Formulation and Paper Structure

The strategies' inherent costs are assumed to decrease with transmission rates, and we order the entries of $\vec{\beta}$ and c as:

$$\vec{\beta}_i < \vec{\beta}_{i+1} \text{ and } c_i > c_{i+1}, \quad 1 \leq i \leq n-1.$$

We assume that $\vec{\beta}_1 > \sigma$, i.e., a transmission rate less than or equal to σ would be unfeasible or too onerous.

Henceforth, c and $\vec{\beta}$ satisfying the conditions above are assumed given and fixed. We will use \tilde{c} 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 .

$$\tilde{c}_i := c_i - c_n, \quad 1 \leq i \leq n. \quad (7)$$

Definition 1: Given a cost budget c^* in $(0, \tilde{c}_1)$, we determine the optimal endemic transmission rate β^* as:

$$\beta^* := \min_{r \in \mathbb{R}_{\geq 0}^n} \{ \beta' C(r - \tilde{c}) \mid r' C(r - \tilde{c}) \leq c^* \}, \quad (8)$$

where C is the choice function defined in (5). Recall that $C(r - \tilde{c})$ is the population state at an equilibrium of the PBR EDM when r is the (fixed) reward vector assigned to the population. [1, Remark 2] discusses the significance of this type of cost constraint.

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

$$\lim_{t \rightarrow \infty} (I, R, \mathcal{B})(t) = (I^*, R^*, \beta^*), \quad (P1)$$

$$\limsup_{t \rightarrow \infty} r'(t)x(t) \leq c^*, \quad (P2)$$

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 system formed by (EDM) and (EPG). Here, the nontrivial endemic equilibrium for (EPGa,b) is:

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

We will seek G and H for which a Lyapunov function for the closed-loop system 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 [10], [11] 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 for 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 present examples 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.

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 [12], 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. [13] and Sontag [11] 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. [14] 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. [10] 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 [15]–[21]. For instance, [15] 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. [21] 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], [22]–[25]. Other studies also considered epidemics with multiple populations [26], [27]; 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 an edge weight indicates the contact or interactions rates across different populations. We refer an interested reader to [3], [22] 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 [26]. 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 [28].

Kuniya and Muroya [27] 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,

²For (P2) to be a meaningful constraint, without loss of generality, we assume that every limit point of $r(t)$ has non-negative entries.

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 [29]–[35]. An earlier work relevant to the perturbed decision-making is [31] 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* [36], of such a game with perturbed payoffs.

Rosenthal [32] 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 [33] 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 experiment data, indicating the need for developing a parameter estimation method. More in-depth treatment of the logit learning rule is presented in [30]. 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 [37].

Other relevant publications include [34] which presents a model that captures the noise in agent decision-making to study the robustness of the *stochastic stability*. [35] 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

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

where the payoff vector $p(t)$ is defined by the reward $r(t)$ and the intrinsic cost \tilde{c} as in (1) and (7). We adopt the notion of δ -passivity from [38] to analyze the stability of (PBR EDM) when it is interconnected with (EPG). (PBR EDM) is δ -passive in that there is a continuously differentiable function $\mathcal{S} : \mathbb{X} \times \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$ given by

$$\mathcal{S}(x, p) = \max_{z \in \text{int}(\mathbb{X})} (z'p - Q(z)) - (x'p - Q(x)), \quad (9)$$

which satisfies

$$\begin{aligned}\nabla'_x \mathcal{S}(x, p) \mathcal{V}(x, p) + \nabla'_p \mathcal{S}(x, p) u &\leq u' \mathcal{V}(x, p), \\ \forall x \in \mathbb{X}, p \in \mathbb{R}^n, u \in \mathbb{R}^n.\end{aligned}\quad (10)$$

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 δ -storage function of (PBR EDM).

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

$$\begin{aligned}G(I, R, x, q) &= (\hat{I} - I) + \eta(\ln I - \ln \hat{I}) + v^2(\bar{\beta} - \mathcal{B}) \\ &\quad + \frac{\mathcal{B}}{\gamma}(R - \hat{R})(1 - \eta - R)\end{aligned}\quad (11a)$$

$$H(I, R, x, q) = q\bar{\beta} + \bar{r}, \quad (11b)$$

where $\hat{I} = \eta(1 - \sigma/\mathcal{B})$, $\hat{R} = (1 - \eta)(1 - \sigma/\mathcal{B})$, $\mathcal{B} = \beta'x$, and $\eta = \omega/(\omega + \gamma)$. The constants $\bar{\beta} \in (\beta_1, \bar{\beta}_n)$ and $\bar{r} \in \mathbb{R}_{\geq 0}^n$ 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 v 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}) . Consequently, a higher value of v leads to a faster convergence of \mathcal{B} to $\bar{\beta}$, but may potentially slow down the convergence of (I, R) to its reference (\hat{I}, \hat{R}) . As we explain in §III-B, v plays an important role in establishing an anytime bound on the infectious fraction $I(t)$. In particular, using simulations, we illustrate that v can be adjusted to regulate the overshoot and convergence rate of (EPG).

The following is a remark on the choice of G and H .

Remark 1: The design of G can be improved by scaling it with a constant κ as follows.

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

The higher the value of κ , the faster the dynamics for the variable $q(t)$. As a result, when κ is sufficiently high, the planner can more swiftly adjust the value of $q(t)$ to contain the spread of epidemics, as we will illustrate in Fig. 2.

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.

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

Since adding an equal value to each entry of the payoff vector p does not affect the results of this work, we use (11b) for concise presentation, unless otherwise specified.

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

We begin by presenting the main convergence result for the feedback interconnection of (EPG) and (PBR EDM).

Theorem 1: Consider the feedback interconnection of (EPG) and (PBR EDM). For any given $\bar{\beta} \in (\bar{\beta}_1, \bar{\beta}_n)$ and $\bar{r} \in \mathbb{R}_{\geq 0}^n$, it holds that

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

The proof is provided in Appendix B.

We note that according to Theorem 1, $(\bar{q}\bar{\beta} + \bar{r})' C(\bar{q}\bar{\beta} + \bar{r} - \tilde{c})$, which represents the limit of $r'(t)x(t)$, denotes the average reward that the social planner needs to spend to maintain the transmission rate $\bar{\beta}$ when the underlying dynamics reach the equilibrium state. In what follows, we investigate the problem of designing optimal $\bar{r}, \bar{\beta}$ that achieve the minimum endemic transmission rate while satisfying the budget constraint (P2). In light of Theorem 1, recall that the budget constraint is given as $(\bar{q}\bar{\beta} + \bar{r})' C(\bar{q}\bar{\beta} + \bar{r} - \tilde{c}) \leq c^*$ where c^* is a positive constant and \bar{q} is the limit of $q(t)$ for given $\bar{r}, \bar{\beta}$.

Consider the optimal choice of $\bar{r} = r^*$ with r^* given as

$$r^* = \arg \min_{r \in \mathbb{R}_{\geq 0}^n} \{ \bar{\beta}' C(r - \tilde{c}) \mid r' C(r - \tilde{c}) \leq c^* \}. \quad (14)$$

According to the definition of (PBR EDM), $C(r - \tilde{c})$ is the population state at the equilibrium when the agents are incentivized with a reward vector r . We remark that for the planner to compute the minimum r^* , it is essential to have knowledge of the function C . In §III-C, we consider a scenario where C is initially unknown to the planner. In this case, C must be estimated using data on the agents' strategy selections which are collected over time.

As a corollary of Theorem 1, we show that when \bar{r} is defined as $\bar{r} = r^*$ and $\bar{\beta}$ is the resulting minimum transmission rate $\bar{\beta} = \beta^* = \bar{\beta}' C(r^* - \tilde{c})$, the reward vector $r(t)$ converges to r^* , and so does the transmission rate $\mathcal{B}(t)$ to the minimum β^* that satisfies the budget constraint.

Corollary 1: Suppose \bar{r} and $\bar{\beta}$ are determined as a solution to (14): $\bar{r} = r^*$ and $\bar{\beta} = \beta^* = \bar{\beta}' C(r^* - \tilde{c})$. For the feedback interconnection of (EPG) and (PBR EDM), it holds that

- 1) $\lim_{t \rightarrow \infty} \mathcal{B}(t) = \beta^*$, and
- 2) $\lim_{t \rightarrow \infty} r(t) = r^*$ while satisfying $r^{*'} C(r^* - \tilde{c}) \leq c^*$.

Remark 2: The proof of Theorem 1 adopts the Lyapunov stability technique which, in conjunction with Corollary 1, implies that the infectious fraction $I(t)$ converges to I^* . In other words, the reward mechanism defined by (EPGc,d) asymptotically attains the minimum achievable infectious fraction while asymptotically satisfying the given budget constraint.

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

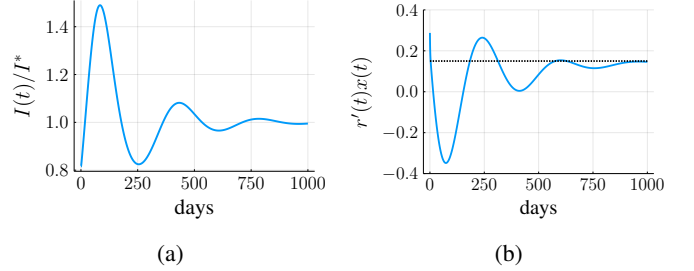


Fig. 1: Simulation results for Example 2 illustrating (a) the infectious fraction $I(t)$ of the population with respect to I^* and (b) the average cost $r'(t)x(t)$.

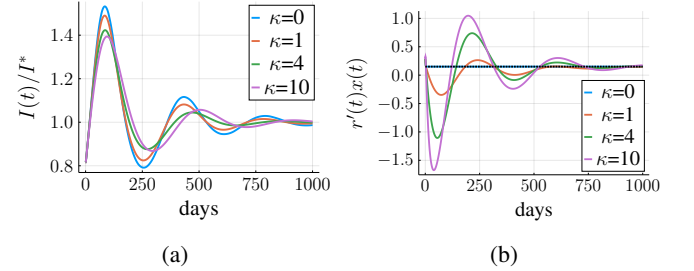


Fig. 2: Simulation results for Example 2 illustrating (a) the infectious fraction $I(t)$ of the population with respect to I^* and (b) the average cost $r'(t)x(t)$, evaluated with different values of κ for (12). Note that $\kappa = 0$ corresponds to the case where the planner is using a static payoff mechanism and the reward is fixed to $r(t) = r^*$. We note that when $\kappa = 0$, the average cost trajectory rapidly converges to its limit point, which is represented by the black dotted horizontal line in (b).

(EPG), where $\gamma = 0.1$ (infectiousness period ~ 10 days), $\sigma = \gamma$, $\omega = 0.005$ (immunity period ~ 200 days), and $v = 3$. With the transmission rates $\bar{\beta} = (0.15, 0.19)$, intrinsic cost $c = (0.2, 0)$, and budget $c^* = 0.15$, we obtain $r^* \approx (0.287, 0)$ and $\beta^* = 0.1691$ from (14) under which $x(t)$ and $(I(t), R(t))$ converge to $x^* = C(r^* - \tilde{c}) \approx (0.522, 0.478)$ 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 $r'(t)x(t)$. We can observe that $I(t)/I^*$ converges to 1 in Fig. 1a, indicating that the infectious fraction $I(t)$ approaches the endemic equilibrium I^* , and $r'(t)x(t)$ converges to c^* , marked by the dotted horizontal line in Fig. 1b. In addition, Fig. 2 illustrates the scenario where the vector field G of (EPGc) is scaled by a constant κ as we explained in Remark 1. As we can observe from the figure, increasing κ will reduce the overshoot in $I(t)/I^*$ (see Fig. 2a) at the expense of larger instantaneous cost (see Fig. 2b).

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

$$\pi_v(\alpha) = \bar{I}^{-1} \sup \{ \mathcal{B}^{-1} \mathcal{I} \mid \mathcal{S}(\mathcal{I}, \mathcal{R}, \mathcal{B}) \leq \alpha \}, \quad (15)$$

where $\bar{I} = \eta(1 - \sigma/\bar{\beta})$ for fixed $\bar{\beta} \in (\bar{\beta}_1, \bar{\beta}_n)$ and $\mathcal{S}(\mathcal{I}, \mathcal{R}, \mathcal{B})$ is a function defined in (24), which we adopt to establish the convergence result in Theorem 1, and \mathcal{I} and \mathcal{R} are given as $\mathcal{I} = \mathcal{B}I$ and $\mathcal{R} = \mathcal{B}R$, respectively.

Recall the δ -storage function $\mathcal{S}(x, p)$ of (PBR EDM) defined in (9). We note that according to (25), the function $\mathcal{S}(x(t), p(t)) + \mathcal{S}(\mathcal{I}(t), \mathcal{R}(t), \mathcal{B}(t))$ is decreasing in t and, hence, it holds that

$$\begin{aligned} \mathcal{S}(\mathcal{I}(t), \mathcal{R}(t), \mathcal{B}(t)) &\leq \mathcal{S}(x(0), p(0)) \\ &\quad + \mathcal{S}(\mathcal{I}(0), \mathcal{R}(0), \mathcal{B}(0)), \quad \forall t \geq 0. \end{aligned} \quad (16)$$

By selecting $\alpha = \mathcal{S}(x(0), p(0)) + \mathcal{S}(\mathcal{I}(0), \mathcal{R}(0), \mathcal{B}(0))$ and using (15), we can establish that $I(t) \leq \bar{I}_{\pi_v}(\alpha)$ holds for all $t \geq 0$. Similar to the remark we made in §III-A, evaluating the anytime bound using (15) requires the planner to have knowledge about the choice function C . This is necessary to compute the δ -storage function \mathcal{S} .

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 β^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, $\mathcal{B}(0) = \beta^o$, $p(0) = q(0)\bar{\beta} + \bar{r} - \tilde{c}$, and $\mathcal{S}(\mathcal{I}(0), \mathcal{R}(0), \mathcal{B}(0)) = v^2\bar{\beta}^2/2$ hold, where $\bar{\beta} = \beta^o - \beta$. Also, we can evaluate the δ -storage function \mathcal{S} of (PBR EDM) at $t = 0$ as

$$\begin{aligned} \mathcal{S}(x(0), p(0)) &= \max_{z \in \text{int}(\mathbb{X})} (z'(q(0)\bar{\beta} + \bar{r} - \tilde{c}) - Q(z)) \\ &\quad - (x'(0)(q(0)\bar{\beta} + \bar{r} - \tilde{c}) - Q(x(0))) \\ &= \max_{z \in \text{int}(\mathbb{X})} (z'(q(0)\bar{\beta} + \bar{r} - \tilde{c}) - Q(z)) \\ &\quad - \max_{z \in \text{int}(\mathbb{X})} (z'(q(0)\bar{\beta} + r^o - \tilde{c}) - Q(z)) \\ &\quad - x'(0)(\bar{r} - r^o). \end{aligned} \quad (17)$$

To derive the latter equality, we use the fact that before the revision to \bar{r} and $\bar{\beta}$, the state was at the endemic equilibrium. Consequently, $\mathcal{S}(x(0), q(0)\bar{\beta} + r^o - \tilde{c}) = 0$ holds. We can express $\mathcal{S}(x(0), p(0))$ as a continuous function $B_S(\tilde{r})$ of \tilde{r} satisfying $B_S(0) = 0$, where $\tilde{r} = r^o - \bar{r}$.³

In conjunction with (15) and (16), by selecting $\alpha = \mathcal{S}(x(0), p(0)) + \mathcal{S}(\mathcal{I}(0), \mathcal{R}(0), \mathcal{B}(0)) = B_S(\tilde{r}) + v^2\bar{\beta}^2/2$, we can establish the anytime bound on $I(t)$ as

$$I(t) \leq \bar{I}_{\pi_v}(B_S(\tilde{r}) + v^2\bar{\beta}^2/2), \quad \forall t \geq 0. \quad (18)$$

Therefore, the bound (18) informs 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. By the continuity of $\pi_v(B_S(\tilde{r}) + v^2\bar{\beta}^2/2)$ in \tilde{r} and $\bar{\beta}$, the bound also suggests how to choose \bar{r} and $\bar{\beta}$ while satisfying a given constraint on the overshoot.

³The function B_S is continuous since it is differentiable [4].

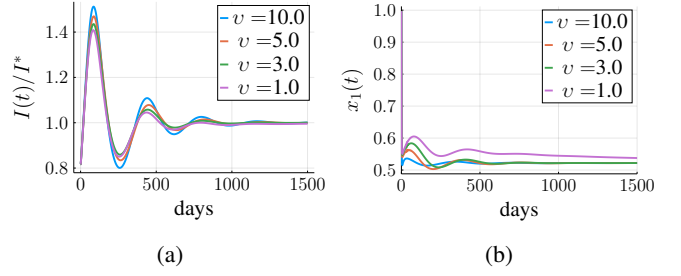


Fig. 3: Simulation results for Example 3 illustrating (a) the infectious fraction $I(t)$ of the population with respect to I^* and (b) the portion $x_1(t)$ of the population selecting strategy 1, evaluated with different values of v .

Example 3: Similar to Example 2, consider that the agents adopt the logit learning rule with $\mu = 1$ and the parameters of (EPG) are given by $\sigma = \gamma = 0.1$ and $\omega = 0.005$ along with the transmission rates $\bar{\beta} = (0.15, 0.19)$ and intrinsic cost $c = (0.2, 0)$ associated with two available strategies. Suppose that in its prior use, the social planner implemented (EPG) with a loose budget constraint, where $r^o = (6.018, 0)$ and $\beta^o = 0.15012$, and the closed-loop model reached an equilibrium where the majority of the agents select the costlier strategy, i.e., $x_1(0) \approx 0.997$, and the infectious and recovered portions of the population are given as $I(0) \approx 0.0159$ and $R(0) \approx 0.317$, respectively.

Now suppose the planner imposes a tighter budget constraint with $c^* = 0.15$ and assigns $\bar{r} = r^* \approx (0.287, 0)$, $\bar{\beta} = \beta^* = 0.1691$, where r^* and β^* are a solution to (14). As illustrated in Fig. 3, we execute multiple simulations with several different values of v . The simulation results indicate that the oscillation in $I(t)/I^*$ becomes lower as we decrease the value of v (see Fig. 3a). This agrees with what can be predicted by the anytime bound (18). On the other hand, for smaller values of v , we observe slower convergence of the population state to its equilibrium (see Fig. 3b).

C. Learning One-Parameter Choice Function C^μ

In §III-A and §III-B, we discussed computing the optimal stationary reward vector r^* and endemic transmission rate β^* using the minimization (14), and establishing the anytime bound (18). 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 [32], [35] demonstrating the predictive power of such a class of models.⁴ To proceed, we express C as

$$C(p) = C^\mu(p) = \arg \max_{z \in \text{int}(\mathbb{X})} (z'p - \mu\bar{Q}(z)), \quad (19)$$

⁴Also, 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.

where $\bar{Q} : \text{int}(\mathbb{X}) \rightarrow \mathbb{R}$ is a known deterministic payoff perturbation whereas $\mu > 0$ is an unknown parameter that needs to be estimated. The logit learning rule, which is one of the most widely studied models in the game theory literature [4], [5], [30], is a specific instance. The analytical results presented in this section can be readily extended to the scenarios where the planner employs multiple one-parameter candidate models, and determines the best model among the candidates and its parameter.

The parameter μ 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, \dots, 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 $v_i = \mu\bar{v}_i$ for all i in $\{1, \dots, n\}$, where $\bar{v}_1, \dots, \bar{v}_n$ are random variables associated with \bar{Q} . Consequently, μ can be interpreted as the scaling factor for the random variables \bar{v}_i . Hence, the larger the parameter μ 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 $\bar{v}_1, \dots, \bar{v}_n$.

In the following proposition, we examine an important property of the choice function C^μ that is useful in the estimation of μ .

Proposition 1: Let C^μ be a one-parameter choice function (19). For any fixed reward vector $r \in \mathbb{R}_{\geq 0}^n$ and intrinsic cost $\bar{c} \in \mathbb{R}_{\geq 0}^n$, where not all entries of $r - \bar{c}$ are identical, $(r - \bar{c})'C^\mu(r - \bar{c})$ is a decreasing function of μ .

The proof is provided in Appendix C. Consequently, the choice function C^μ satisfies $(r - \bar{c})'C^{\hat{\mu}}(r - \bar{c}) = (r - \bar{c})'C^\mu(r - \bar{c})$ if and only if $\hat{\mu} = \mu$ holds. Hence, when $(r, C^\mu(r - \bar{c}))$ is given as data for the estimation of μ , the social planner can learn unique μ . In what follows, we discuss how the planner can devise a parameter estimation scheme based on Proposition 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 does not know the parameter μ , can employ (EPG) with an initial selection of \bar{r} and $\bar{\beta}$ to curb the spread of the pandemic. In the meantime, the planner can collect data on the agents' decision on strategy selections to estimate μ . For the data collection, we consider a survey method in which randomly selected agents are asked about which strategies they would choose given a fixed reward vector r , which is not necessarily the same as \bar{r} adopted for (EPG). Based on the survey data, the planner estimates bounds on $(r - \bar{c})'C^\mu(r - \bar{c})$ and, consequently, those on μ using Proposition 1.

For the survey, we choose r that satisfies

$$\max_{1 \leq i \leq n} (r_i - \bar{c}_i) - \min_{1 \leq i \leq n} (r_i - \bar{c}_i) = 2. \quad (20)$$

With such fixed r , let \mathbf{R} be a discrete random variable whose probability is defined as $\mathbb{P}[\mathbf{R} = r_i - \bar{c}_i] = C_i^\mu(r - \bar{c})$. Note that $\mathbb{E}[\mathbf{R}] = (r - \bar{c})'C^\mu(r - \bar{c})$ holds. After surveying K randomly selected agents, using the answers collected, we can compute K realizations $\{R^{(i)}\}_{i=1}^K$ of \mathbf{R} . Then, by applying

Chebyshev's inequality, we can establish

$$\begin{aligned} \mathbb{P}\left(\left|\mathbb{E}[\mathbf{R}] - \frac{1}{K} \sum_{i=1}^K R^{(i)}\right| \leq \epsilon\right) &\geq 1 - \frac{\text{Var}[\mathbf{R}]}{\epsilon^2 K} \\ &\geq 1 - \frac{1}{\epsilon^2 K}, \end{aligned} \quad (21)$$

where, to establish the last inequality, we use (20) and Popoviciu's inequality. The inequality (21) explains how to find both an upper and a lower bounds of $(r - \bar{c})'C^\mu(r - \bar{c})$ using the sample mean, with an arbitrarily high level of confidence. Consequently, using Proposition 1, we can compute bounds on μ . Later, once a sufficient amount of data has been collected to infer μ with any required accuracy, the planner can optimize the choice of \bar{r} and $\bar{\beta}$ using (14).

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

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

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

where $\bar{q}(\mu)$ is a solution, which depends on μ , to $\bar{\beta}'C^\mu(\bar{q}(\mu)\bar{\beta} + \bar{r} - \bar{c})$.

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, we fix $\bar{r} = \bar{c}$ and assume $\bar{c}'C^1(0) < c^*$.

Proposition 2: Suppose μ lies within the interval $[\mu_L, \mu_U]$ and $\bar{\beta}$ satisfies $\bar{\beta} < \bar{\beta}'C^1(0)$. For fixed $\bar{\beta}$, the average stationary cost $(\bar{q}\bar{\beta} + \bar{r})'C^\mu(\bar{q}\bar{\beta} + \bar{r} - \bar{c})$ is upper bounded by

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

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

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

is an increasing function of λ .

The proof is provided in Appendix D. By Proposition 2, the unique solution λ 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 $\bar{\beta} = \bar{\beta}'C^1(0)$, which incurs the average stationary cost of $\bar{c}'C^1(0) < c^*$, the planner can repeatedly assess (23) as it decreases the value of $\bar{\beta}$ until it finds the lowest $\bar{\beta}$ whose associated bound (23) satisfies the budget constraint.

Fig. 4 illustrates how the upper bound (23) varies depending on the values of μ_U and $\bar{\beta}$. The choice function is assumed to be logit with parameter $\mu = 1$, and the parameters $\bar{\beta}, \bar{c}$,

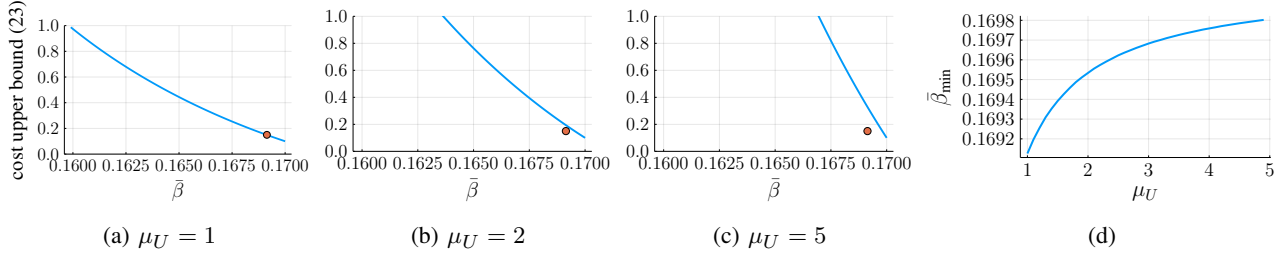


Fig. 4: Plots (a)–(c) illustrate how the cost upper bound (23) varies depending on $\bar{\beta}$ for each fixed μ_U : (a) $\mu_U = 1$, (b) $\mu_U = 2$, and (c) $\mu_U = 5$. The red circle indicates the optimal endemic transmission rate $\beta^* = 0.1691$ and the budget $c^* = 0.15$. Plot (d) depicts 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., $\bar{\beta}_{\min} = \min\{\bar{\beta} > 0 \mid \mu_U \lambda(\bar{\beta} - \bar{\beta}_n) + \tilde{c}' \arg \max_{z \in \text{int}(\mathbb{X})} (z'(\lambda\bar{\beta}) - \bar{Q}(z)) \leq c^*\}$.

and c^* are given by $\bar{\beta} = (0.15, 0.19)$, $\tilde{c} = (0.2, 0)$, and $c^* = 0.15$, respectively. From the plots in Fig. 4, we observe that as the bound on μ becomes loose, so does the estimated upper bound of the average stationary cost for the given budget. In particular, we can see from Fig. 4a that if the upper bound of μ 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 μ 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^* , the planner assigns $\bar{r} = \tilde{c}$ and, using a bound of μ known a priori, computes the lowest $\bar{\beta}$ for which its corresponding cost upper bound (23) satisfies the budget constraint.
- 2) By surveying randomly selected agents on their strategy choice with a constant reward vector r , the planner computes the sample mean of the survey outcomes. Using (21) and Proposition 1, it estimates the upper and lower bounds of μ .
- 3) After estimating the value of μ satisfying any required accuracy, the planner computes the optimal solution r^*, β^* to (14) and revises $\bar{r}, \bar{\beta}$ using the solution. Based on (15), with revised $\bar{r}, \bar{\beta}$ and estimated μ , the planner can establish the anytime bound on the infectious fraction $I(t)$ of the population.

Consider that the agents adopt the logit learning rule with $\mu = 1$ and the parameters of (EPG) are given by $\sigma = \gamma = 0.1$, $\omega = 0.005$, and $v = 3$. With the transmission rates $\bar{\beta} = (0.15, 0.19)$, intrinsic cost $c = (0.2, 0)$, and budget $c^* = 1$, we obtain $r^* \approx (1.3248, 0)$ and $\beta^* \approx 0.1598$ from (14) 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 μ , but not its exact value. By following the method explained in Proposition 2, the planner selects $\bar{r} = \tilde{c}$ and $\bar{\beta} \approx 0.167$ under which the upper bound (23) satisfies the budget constraint. The planner uses (21) to determine when the average of samples is within 0.05 of the true value of $\mathbb{E}[\mathbf{R}]$ with probability 0.95 as it accumulates poll data from 1000 randomly selected agents every 30 days. Let $t_0 = 240$ be the day on which the estimated μ satisfies the estimation accuracy requirement. Once it does, the planner computes the estimate of μ based on the collected poll data, and the parameters $\bar{r}, \bar{\beta}$ are updated using the solution r^*, β^* to (14). 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

$$\alpha = \mathcal{S}(x(t), p(t)) + \mathcal{J}(\mathcal{I}(t), \mathcal{R}(t), \mathcal{B}(t)),$$

$$t = t_0 + 30, t_0 + 60, \dots$$

at every 30 days and updates \bar{r} to the value that is closest to r^* while ensuring that α remains below 0.0004. According to (15), this corresponds to a 15% overshoot in $I(t)/I^*$. Fig. 5 illustrates the simulation results for the two scenarios. As depicted in Fig. 5a, 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 (15). 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. 5b. 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 subject to perturbation. For instance, such perturbation is relevant when modeling the strategy selection of agents based on (inaccurate) estimates of net rewards. 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

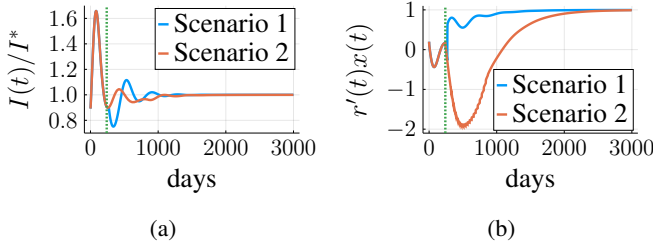


Fig. 5: Simulation results for Scenario 1 and Scenario 2 illustrating (a) the infectious fraction $I(t)$ of the population with respect to I^* and (b) the average cost $r'(t)x(t)$, where the parameters $\bar{r}, \bar{\beta}$ of (EPG) are updated using the optimal r^*, β^* beginning from the day $t_0 = 240$, as indicated by the green dotted vertical line.

the global convergence of the average transmission rate to its achievable minimum, subject to the budget constraint, thereby minimizing the infectious fraction of the population in the long run. Additionally, we established an anytime bound on the infectious fraction of the population using the Lyapunov stability method. We also discussed scenarios where the social planner needs to estimate the parameter of the agents' learning rule. The simulation results provided throughout the paper illustrate our theoretical findings and highlight their efficacy in curbing the spread of epidemic, as demonstrated in illustrative deployment scenarios, explained in §IV.

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. As we illustrated in Fig. 2, modulating the parameter κ in (12) can be a potential solution to regulate the overshoot in the average cost trajectory. Along same lines, we also plan to explore higher-order design of (EPGc) for $q(t)$. In our current work, the first-order dynamic model defined by (11a) was sufficient to establish the global convergence results, subject to the asymptotic budget constraint. However, with regards to the transient behavior of the average cost, adopting a higher-order model could be beneficial in controlling the variation in the cost trajectory.

- [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.
- [9] S. Arigapudi, Exit from equilibrium in coordination games under probit choice, *Games and Economic Behavior* 122 (2020) 168–202.
- [10] P. Godara, S. Herminghaus, K. M. Heidemann, A control theory approach to optimal pandemic mitigation, *PLOS ONE* 16 (2) (2021) 1–16.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] 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.

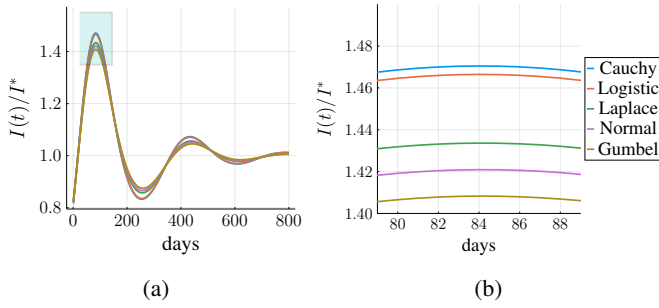


Fig. 6: The ratio $I(t)/I^*$ for (EPG) evaluated with different noise distributions for the choice function (4). The square area highlighted in (a) is presented at an enlarged scale in (b).

- [32] R. W. Rosenthal, A bounded-rationality approach to the study of noncooperative games, *International Journal of Game Theory* 18 (3) (1989) 273–292.
- [33] R. D. McKelvey, T. R. Palfrey, Quantal response equilibria for normal form games, *Games and Economic Behavior* 10 (1) (1995) 6–38.
- [34] L. E. Blume, How noise matters, *Games and Economic Behavior* 44 (2) (2003) 251–271.
- [35] 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.
- [36] 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.
- [37] 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.
- [38] S. Park, N. C. Martins, J. S. Shamma, From population games to payoff dynamics models: A passivity-based approach, in: 2019 IEEE 58th Conference on Decision and Control (CDC), 2019, pp. 6584–6601.

APPENDIX

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

Fig. 6 depicts the infectious fraction of the population when the probability distributions of the random variables v_1, \dots, 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

We provide a three-part proof in which we establish $\lim_{t \rightarrow \infty} \|\dot{x}(t)\|_2 = \lim_{t \rightarrow \infty} |I(t) - \hat{I}(t)| = \lim_{t \rightarrow \infty} |R(t) - \hat{R}(t)| = 0$ in **Part 2** and $\lim_{t \rightarrow \infty} q(t) = \bar{q}$ and $\lim_{t \rightarrow \infty} \mathcal{B}(t) = \bar{\beta}$ in **Part 3**, where \bar{q} is a unique solution to $\bar{\beta} = \bar{\beta}' C(\bar{q} \bar{\beta} + \bar{r} - \bar{c})$. Consequently, in conjunction with $\lim_{t \rightarrow \infty} \|x(t) - C(q(t) \bar{\beta} + \bar{r} - \bar{c})\|_2 = \lim_{t \rightarrow \infty} \|\dot{x}(t)\|_2 = 0$, 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 \mathcal{S} from [1], we can state

$$\begin{aligned} \mathcal{S}(\mathcal{I}, \mathcal{R}, \mathcal{B}) &= (\mathcal{I} - \hat{\mathcal{I}}) + \hat{\mathcal{I}} \ln \frac{\hat{\mathcal{I}}}{\mathcal{I}} + \frac{1}{2\gamma} (\mathcal{R} - \hat{\mathcal{R}})^2 + \frac{v^2}{2} (\mathcal{B} - \bar{\beta})^2 \\ &= (\mathcal{I} - \eta(\mathcal{B} - \sigma)) + \eta(\mathcal{B} - \sigma) \ln \frac{\eta(\mathcal{B} - \sigma)}{\mathcal{I}} \\ &\quad + \frac{1}{2\gamma} (\mathcal{R} - (1 - \eta)(\mathcal{B} - \sigma))^2 + \frac{v^2}{2} (\mathcal{B} - \bar{\beta})^2, \end{aligned} \quad (24)$$

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

$$\begin{aligned} \frac{d}{dt} \mathcal{S}(x(t), p(t)) + \frac{d}{dt} \mathcal{S}(\mathcal{I}(t), \mathcal{R}(t), \mathcal{B}(t)) &= \nabla'_x \mathcal{S}(x(t), p(t)) \mathcal{V}(x(t), p(t)) \\ &\quad - (\mathcal{I}(t) - \hat{\mathcal{I}}(t))^2 - \frac{\omega}{\gamma} (\mathcal{R}(t) - \hat{\mathcal{R}}(t))^2 \leq 0, \end{aligned} \quad (25)$$

which implies, since \mathcal{S} and \mathcal{S} are non-negative functions, \mathcal{S} is a bounded function – both below and above. In conjunction with the fact that $\hat{\mathcal{I}} = \eta(\mathcal{B} - \sigma) \geq \eta(\bar{\beta}_1 - \sigma) > 0$, 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 $\mathcal{S}(\mathcal{I}(t), \mathcal{R}(t), \mathcal{B}(t))$ as $\mathcal{I}(t)$ approaches zero, which contradicts the fact that \mathcal{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 $\lim_{t \rightarrow \infty} \|\dot{x}(t)\|_2 = \lim_{t \rightarrow \infty} |I(t) - \hat{I}(t)| = \lim_{t \rightarrow \infty} |R(t) - \hat{R}(t)| = 0$: For notational convenience, let us define

$$\check{\mathcal{S}}(\mathcal{I}, \mathcal{R}, \mathcal{B}) = (\mathcal{I} - \hat{\mathcal{I}}) + \hat{\mathcal{I}} \ln \frac{\hat{\mathcal{I}}}{\mathcal{I}} + \frac{1}{2\gamma} (\mathcal{R} - \hat{\mathcal{R}})^2. \quad (26)$$

Note that $0 \leq \check{\mathcal{S}}(\mathcal{I}, \mathcal{R}, \mathcal{B}) \leq \mathcal{S}(\mathcal{I}, \mathcal{R}, \mathcal{B})$. According to (25) and (26), we can derive

$$\begin{aligned} \mathcal{S}(x(t), p(t)) + \check{\mathcal{S}}(\mathcal{I}(t), \mathcal{R}(t), \mathcal{B}(t)) - \alpha &\leq \mathcal{S}(x(t), p(t)) + \mathcal{S}(\mathcal{I}(t), \mathcal{R}(t), \mathcal{B}(t)) - \alpha \\ &= \int_0^t (\nabla'_x \mathcal{S}(x(\tau), p(\tau)) \mathcal{V}(x(\tau), p(\tau)) \\ &\quad - (\mathcal{I}(\tau) - \hat{\mathcal{I}}(\tau))^2 - \frac{\omega}{\gamma} (\mathcal{R}(\tau) - \hat{\mathcal{R}}(\tau))^2) d\tau, \quad \forall t \geq 0, \end{aligned} \quad (27)$$

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

$$\lim_{t \rightarrow \infty} \left(\mathcal{S}(x(t), p(t)) + \check{\mathcal{S}}(\mathcal{I}(t), \mathcal{R}(t), \mathcal{B}(t)) \right) = 0, \quad (28)$$

which implies

$$\begin{aligned} \lim_{t \rightarrow \infty} \mathcal{S}(x(t), p(t)) &= 0 \\ \lim_{t \rightarrow \infty} |I(t) - \hat{I}(t)| &\leq \lim_{t \rightarrow \infty} \frac{1}{\beta_1} |I(t) - \hat{I}(t)| = 0 \\ \lim_{t \rightarrow \infty} |R(t) - \hat{R}(t)| &\leq \lim_{t \rightarrow \infty} \frac{1}{\beta_1} |R(t) - \hat{R}(t)| = 0. \end{aligned}$$

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

$$\tilde{z}'\nabla Q(y) = \tilde{z}'r \iff y = \arg \max_{z \in \text{int}(\mathbb{X})} (z'r - Q(z)) \quad (29)$$

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

$$\begin{aligned} \mathcal{S}(x(t), p(t)) &= (y(t) - x(t))'\nabla Q(y(t)) \\ &\quad - (Q(y(t)) - Q(x(t))). \end{aligned} \quad (30)$$

Therefore, by the strict convexity (6a) of Q and (30), $\lim_{t \rightarrow \infty} \mathcal{S}(x(t), p(t)) = 0$ implies that $\lim_{t \rightarrow \infty} \|\dot{x}(t)\|_2 = 0$.

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

$$\mathbb{O}_\epsilon = \left\{ t > 0 \mid \mathcal{S}(x(t), p(t)) + \check{\mathcal{J}}(\mathcal{I}(t), \mathcal{R}(t), \mathcal{B}(t)) > \frac{\epsilon}{2} \right\}.$$

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

$$\mathcal{S}(x(a_i), p(a_i)) + \check{\mathcal{J}}(\mathcal{I}(a_i), \mathcal{R}(a_i), \mathcal{B}(a_i)) = \frac{\epsilon}{2}. \quad (31)$$

The δ -storage function of (PBR EDM) satisfies

$$\begin{aligned} \mathcal{S}(x, p) &= y'p - Q(y) - (x'p - Q(x)) \\ &= p'(y - x) - (Q(y) - Q(x)) \\ &\stackrel{(i)}{\leq} (p - \nabla Q(x))'(y - x) \\ &= -\nabla'_x \mathcal{S}(x, p) \mathcal{V}(x, p), \end{aligned} \quad (32)$$

where $y = \arg \max_{z \in \text{int}(\mathbb{X})} (z'p - Q(z))$. For (i), we use the convexity (6a) of Q to establish $Q(y) - Q(x) \geq \nabla'Q(x)(y - x)$. Also, by **Part 1**, since $\hat{\mathcal{I}}, \mathcal{I}$ are strictly positive, i.e., there is $\delta > 0$ for which $\mathcal{I}, \hat{\mathcal{I}} \geq \delta$ holds, we can find a positive constant k satisfying the inequality given by

$$k|\mathcal{I} - \hat{\mathcal{I}}| \geq (\mathcal{I} - \hat{\mathcal{I}}) + \hat{\mathcal{I}} \ln \frac{\hat{\mathcal{I}}}{\mathcal{I}}. \quad (33)$$

For instance, (33) holds with k defined as

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

Consequently, by (32) and (33), for every $t \in \mathbb{O}_\epsilon$, there is a constant $\delta_\epsilon > 0$ for which

$$\begin{aligned} \nabla'_x \mathcal{S}(x(t), p(t)) \mathcal{V}(x(t), p(t)) \\ - (\mathcal{I}(t) - \hat{\mathcal{I}}(t))^2 - \frac{\omega}{\gamma} (\mathcal{R}(t) - \hat{\mathcal{R}}(t))^2 < -\delta_\epsilon. \end{aligned} \quad (34)$$

Since $\mathcal{S}(x(t), p(t)) + \check{\mathcal{J}}(\mathcal{I}(t), \mathcal{R}(t), \mathcal{B}(t))$ is a non-negative function of t , from (27) and (34), we can derive

$$\begin{aligned} -\alpha &\leq \int_0^\infty (\nabla'_x \mathcal{S}(x(\tau), p(\tau)) \mathcal{V}(x(\tau), p(\tau)) \\ &\quad - (\mathcal{I}(\tau) - \hat{\mathcal{I}}(\tau))^2 - \frac{\omega}{\gamma} (\mathcal{R}(\tau) - \hat{\mathcal{R}}(\tau))^2) d\tau \\ &\leq \int_{\mathbb{O}_\epsilon} (\nabla'_x \mathcal{S}(x(\tau), p(\tau)) \mathcal{V}(x(\tau), p(\tau)) \\ &\quad - (\mathcal{I}(\tau) - \hat{\mathcal{I}}(\tau))^2 - \frac{\omega}{\gamma} (\mathcal{R}(\tau) - \hat{\mathcal{R}}(\tau))^2) d\tau \\ &\leq -\delta_\epsilon \mathcal{L}(\mathbb{O}_\epsilon), \end{aligned} \quad (35)$$

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

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

$$\mathcal{S}(x(t), p(t)) + \check{\mathcal{J}}(\mathcal{I}(t), \mathcal{R}(t), \mathcal{B}(t)) < \epsilon, \quad \forall t \geq T_\epsilon. \quad (36)$$

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

$$\max_{t \in \mathbb{J}_j} \left(\mathcal{S}(x(t), p(t)) + \check{\mathcal{J}}(\mathcal{I}(t), \mathcal{R}(t), \mathcal{B}(t)) \right) \geq \epsilon \quad (37)$$

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

$$\begin{aligned} \mathcal{S}(x(\bar{t}_j), p(\bar{t}_j)) + \check{\mathcal{J}}(\mathcal{I}(\bar{t}_j), \mathcal{R}(\bar{t}_j), \mathcal{B}(\bar{t}_j)) \\ - \mathcal{S}(x(a_j), p(a_j)) - \check{\mathcal{J}}(\mathcal{I}(a_j), \mathcal{R}(a_j), \mathcal{B}(a_j)) \\ = \int_{a_j}^{\bar{t}_j} \frac{d}{d\tau} \left(\mathcal{S}(x(\tau), p(\tau)) + \check{\mathcal{J}}(\mathcal{I}(\tau), \mathcal{R}(\tau), \mathcal{B}(\tau)) \right) d\tau \\ \stackrel{(i)}{\leq} - \int_{a_j}^{\bar{t}_j} v^2 \dot{\mathcal{B}}(\tau) (\mathcal{B}(\tau) - \bar{\beta}) d\tau \\ \leq \int_{a_j}^{\bar{t}_j} v^2 \|\vec{\beta}\|_2 \|\dot{x}(\tau)\|_2 (\|\vec{\beta}\|_2 \|x(\tau)\|_2 + \bar{\beta}) d\tau \end{aligned} \quad (38)$$

To obtain (i), we use (25) and the definition of $\check{\mathcal{J}}$. Since both \dot{x} and x are bounded, we can find a constant M for which the following inequality holds.

$$\begin{aligned} \mathcal{S}(x(\bar{t}_j), p(\bar{t}_j)) + \check{\mathcal{J}}(\mathcal{I}(\bar{t}_j), \mathcal{R}(\bar{t}_j), \mathcal{B}(\bar{t}_j)) \\ - \mathcal{S}(x(a_j), p(a_j)) - \check{\mathcal{J}}(\mathcal{I}(a_j), \mathcal{R}(a_j), \mathcal{B}(a_j)) \\ \leq M(\bar{t}_j - a_j) \end{aligned} \quad (39)$$

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

$$\begin{aligned} \mathcal{S}(x(\bar{t}_j), p(\bar{t}_j)) + \check{\mathcal{J}}(\mathcal{I}(\bar{t}_j), \mathcal{R}(\bar{t}_j), \mathcal{B}(\bar{t}_j)) \\ < \mathcal{S}(x(a_j), p(a_j)) + \check{\mathcal{J}}(\mathcal{I}(a_j), \mathcal{R}(a_j), \mathcal{B}(a_j)) + \frac{\epsilon}{2} = \epsilon, \end{aligned}$$

which contradicts (37). This completes **Part 2**.

Part 3. It holds that $\lim_{t \rightarrow \infty} q(t) = \bar{q}$ and $\lim_{t \rightarrow \infty} \mathcal{B}(t) = \bar{\beta}$: Since $p(t) = r(t) - \tilde{c} = q(t)\vec{\beta} + \bar{r} - \tilde{c}$, we can re-write (EPGc) as in (40). By **Part 2**, we observe that

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

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 $\lim_{t \rightarrow \infty} \epsilon(t) = 0$, where $\epsilon(t)$ is defined in (40). Hence, if $q(t)$ goes to \bar{q} , satisfying $\bar{\beta} = \vec{\beta}' C(\bar{q}\vec{\beta} + \bar{r} - \tilde{c})$, as t tends to infinity, then we can infer that $\lim_{t \rightarrow \infty} \mathcal{B}(t) = \vec{\beta}' C(\bar{q}\vec{\beta} + \bar{r} - \tilde{c}) = \bar{\beta}$.

$$\begin{aligned} \dot{q}(t) = & v^2 (\bar{\beta} - \bar{\beta}' C(q(t)\bar{\beta} + \bar{r} - \bar{c})) \\ & + v^2 (\underbrace{\bar{\beta}' C(q(t)\bar{\beta} + \bar{r} - \bar{c}) - \mathcal{B}(t)}_{=\epsilon(t)} + (\hat{I}(t) - I(t)) + \eta(\ln I(t) - \ln \hat{I}(t)) + \frac{\mathcal{B}(t)}{\gamma} (R(t) - \hat{R}(t))(1 - \eta - R(t)) \end{aligned} \quad (40)$$

Next, we show that $\bar{\beta}' C(q\bar{\beta} + \bar{r} - \bar{c})$ is an increasing function of q . To proceed, we establish that $\frac{\partial}{\partial q} \bar{\beta}' C(q\bar{\beta} + \bar{r} - \bar{c}) > 0$, $\forall q \in \mathbb{R}$. Note that

$$\begin{aligned} & \frac{\partial}{\partial q} \bar{\beta}' C(q\bar{\beta} + \bar{r} - \bar{c}) \\ &= \frac{\partial}{\partial q} \bar{\beta}' \arg \max_{z \in \text{int}(\mathbb{X})} (z'(q\bar{\beta} + \bar{r} - \bar{c}) - Q(z)) \\ &= \bar{\beta}' \nabla_p \arg \max_{z \in \text{int}(\mathbb{X})} (z'p - Q(z)) \Big|_{p=q\bar{\beta} + \bar{r} - \bar{c}} \bar{\beta}. \end{aligned} \quad (41)$$

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

$$\bar{\beta}' \nabla_p \arg \max_{z \in \text{int}(\mathbb{X})} (z'p - Q(z)) \Big|_{p=q\bar{\beta} + \bar{r} - \bar{c}} \bar{\beta} > 0. \quad (42)$$

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

$$\lim_{q \rightarrow \infty} \bar{\beta}' C(q\bar{\beta} + \bar{r} - \bar{c}) = \bar{\beta}_n \quad (43a)$$

$$\lim_{q \rightarrow -\infty} \bar{\beta}' C(q\bar{\beta} + \bar{r} - \bar{c}) = \bar{\beta}_1. \quad (43b)$$

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

C. Proof of Proposition 1

We begin with deriving

$$\begin{aligned} & (r - \bar{c})' C^\mu(r - \bar{c}) \\ &= (r - \bar{c})' \arg \max_{z \in \text{int}(\mathbb{X})} (z'(r - \bar{c}) - \mu \bar{Q}(z)) \\ &= (r - \bar{c})' \arg \max_{z \in \text{int}(\mathbb{X})} (\lambda z'(r - \bar{c}) - \bar{Q}(z)), \quad r \in \mathbb{R}_{\geq 0}^n, \end{aligned} \quad (44)$$

where $\lambda = \mu^{-1}$. To complete the proof, we proceed with showing that (44) is an increasing function of λ . Note that

$$\begin{aligned} & \frac{\partial}{\partial \lambda} (r - \bar{c})' \arg \max_{z \in \text{int}(\mathbb{X})} (\lambda z'(r - \bar{c}) - \bar{Q}(z)) \\ &= (r - \bar{c})' \nabla_p \arg \max_{z \in \text{int}(\mathbb{X})} (z'p - \bar{Q}(z)) \Big|_{p=\lambda(r - \bar{c})} (r - \bar{c}). \end{aligned}$$

By the same arguments used in the proof of [4, Theorem 2.1], since not all entries of $r - \bar{c}$ are identical, we infer that

$$(r - \bar{c})' \nabla_p \arg \max_{z \in \text{int}(\mathbb{X})} (z'p - \bar{Q}(z)) \Big|_{p=\lambda(r - \bar{c})} (r - \bar{c}) > 0.$$

We conclude $\frac{\partial}{\partial \lambda} (r - \bar{c})' \arg \max_{z \in \text{int}(\mathbb{X})} (\lambda z'(r - \bar{c}) - \bar{Q}(z)) > 0$, and (44) is an increasing function of λ and, hence, it is decreasing in μ . ■

D. Proof of Proposition 2

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 $\bar{r} = \bar{c}$, $q(t)$ converges to \bar{q} for which $\bar{\beta} = \bar{\beta}' C^\mu(\bar{q}\bar{\beta})$ holds and, hence, $r(t)$ converges to $\bar{q}\bar{\beta} + \bar{c}$.

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

$$\bar{\beta} = \bar{\beta}' C^\mu(\bar{q}\bar{\beta}) = \bar{\beta}' \arg \max_{z \in \text{int}(\mathbb{X})} (z'(\lambda\bar{\beta}) - \bar{Q}(z)). \quad (45)$$

By the same argument used in the proof of Proposition 1, $\bar{\beta}' \arg \max_{z \in \text{int}(\mathbb{X})} (z'(\lambda\bar{\beta}) - \bar{Q}(z))$ is an increasing function of λ . Also, if $\bar{\beta}$ satisfies $\bar{\beta} < \bar{\beta}' C^1(0)$ then it holds that $\lambda < 0$ and $\bar{q} < 0$.

When (EPG) and (PBR EDM) reach their equilibrium state, the planner would spend

$$\begin{aligned} & (\bar{q}\bar{\beta} + \bar{c} - \bar{q}\bar{\beta}_n \mathbf{1})' \arg \max_{z \in \text{int}(\mathbb{X})} (z'(\bar{q}\bar{\beta}) - \mu \bar{Q}(z)) \\ &= \mu \lambda (\bar{\beta} - \bar{\beta}_n) + \bar{c}' \arg \max_{z \in \text{int}(\mathbb{X})} (z'(\lambda\bar{\beta}) - \bar{Q}(z)), \end{aligned} \quad (46)$$

which is upper bounded by

$$\mu_U \lambda (\bar{\beta} - \bar{\beta}_n) + \bar{c}' \arg \max_{z \in \text{int}(\mathbb{X})} (z'(\lambda\bar{\beta}) - \bar{Q}(z)). \quad (47)$$

This completes the proof. ■

Shinkyu Park (Member, IEEE) received the Ph.D. degree from the University of Maryland, College Park, MD, USA, in 2015, 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. He is a recipient of 2022 AACC O. Hugo Schuck Best Paper Award.

Jair Certoio (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.