

Incentive Designs for Learning Agents to Stabilize Coupled Exogenous Systems

Jair Certório, *Student Member, IEEE*, Nuno C. Martins, *Senior Member, IEEE*,
Richard J. La and Murat Arcak, *Fellow, IEEE*.

Abstract—We consider a large population of learning agents noncooperatively selecting strategies from a common set, influencing the dynamics of an exogenous system (ES) we seek to stabilize at a desired equilibrium. Our approach is to design a dynamic payoff mechanism capable of shaping the population’s strategy profile, thus affecting the ES’s state, by offering incentives for specific strategies within budget limits. Employing system-theoretic passivity concepts, we establish conditions under which a payoff mechanism can be systematically constructed to ensure the global asymptotic stability of the ES’s equilibrium. In comparison to previous approaches originally studied in the context of the so-called epidemic population games, the method proposed here allows for more realistic epidemic models and other types of ESs, such as predator-prey dynamics. The stability of the equilibrium is established with the support of a Lyapunov function, which provides useful bounds on the transient states.

I. INTRODUCTION

Systems whose behavior depends on the strategic choices of many agents can be studied through the lens of evolutionary game theory, in particular when considering large populations of nondescript agents that repeatedly revise their strategies. Examples of systems with such dependency on the aggregate choice of a large population include models of traffic congestion [1], optimal power dispatch [2], distributed task allocation [3], building temperature control [4], and epidemic mitigation [5].

The coupling between the *evolutionary dynamics*, which models the population’s strategic choices, and the *exogenous system* (ES) dynamics, which captures the state of the system affected by the decisions of the population, makes the task of designing stabilizing policies challenging. This is especially true when we aim to design policies that not only improve the behavior of the system, but also provide performance guarantees that hold at any given time.

In our work, we generalize the design concept from [5] to a larger class of ESs which, when agents stop revising their strategies, have a Lyapunov function and satisfy some mild assumptions. We design incentives that guarantee the

convergence of the population state to an equilibrium, which can be selected independently of the payoff mechanism. While our results require similar assumptions on the behavior of the population to those of [5], unlike [5], [6] we obtain a bound on the instantaneous cost of implementing our incentives.

In addition, we show that the proposed payoff mechanism is compatible with many ESs, such as the epidemic models studied by [5], [6], [7] and the Leslie-Gower model for studying the interaction between populations of hosts and parasites [8]. As an application, we use the proposed framework to devise incentives for a modified epidemic model studied in [6], while considering disease transmission rates that depend nonlinearly on the agents’ choices.

A. Contributions

The goal of our study is to develop a new framework for designing a dynamic payoff mechanism that guarantees the convergence of both the population and the ES, whose dynamics are influenced by the strategic choices of the agents in the population, to a desirable equilibrium that can be selected independently of the payoff. The incentive we design has a bound on the instantaneous reward offered to the population, which is not guaranteed in the previous studies that use δ -passivity for designing policies that mitigate epidemics.

The assumptions we introduce on the learning rule employed by the agents are similar to those of [5]. However, we relax one of the assumptions on the learning rule, which is not easy to verify, and replace it with another assumption that is easier to check. The proposed mechanism works on previously studied ESs [5], [6], [7], but is not limited to epidemic models; our framework is more general and can be applied to any system satisfying the conditions identified in this paper, even when the learning rule is unknown to the policy maker. As an illustrative example, we show in §VI-A that the Leslie-Gower system studied in [8] satisfies these conditions.

II. RELATED WORKS

Earlier studies showed that, for potential games, bounded-rational learning rules with the *positive correlation* (PC) property, where agents revise their strategies in a way that increases their payoffs, guarantee the convergence of the population state to *Nash equilibria* (NEs) [9]. In addition, Hofbauer and Sandholm [10] established that, for the class of *contractive games*, many evolutionary dynamics lead the

The work of Certório and Martins was supported by the AFOSR grant FA95502310467, and the NSF grants 2135561 and 2139713. The work of Arcak was supported by the NSF grant CNS-2135791

Jair Certório, Nuno C. Martins, and Richard J. La are with the Department of Electrical and Computer Engineering and Institute for Systems Research, University of Maryland, College Park, College Park, MD 20740 USA (e-mail: certorio, nmartins, hyongla@umd.edu).

Murat Arcak is with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA 94720 USA (e-mail: arcak@eecs.berkeley.edu).

population state to an NE. A recent study demonstrated that for certain potential and strictly contractive games, the population state converges to an NE, even when the revision rates depend explicitly on the current strategies of the agents [11], [12]. For a survey of earlier studies and the applications of population games we refer the reader to [13], [14], and the references therein.

Motivated by the class of contractive games, Fox and Shamma [15] showed that certain learning rules, such as *impartial pairwise comparison* and *excess-payoff target* rules, exhibit a form of passivity which they named δ -passivity. The concept of δ -passivity was generalized in [16] to admit a large class of dynamical payoff mechanism and in [1] that introduced δ -dissipativity. In [17], the authors determined a sufficient condition for the interconnections of δ -dissipative dynamical systems to also be δ -dissipative.

Adapting tools from robust control, Mabrok and Shamma [18] studied the passivity properties of higher-order games and determined necessary conditions for evolutionary dynamics to be stable for all higher-order passive games. Their work also proved that replicator dynamics is lossless [18], which was later shown to be not δ -passive [19].

The population game framework has been used in many problems. For example, it has been used for distributed optimization [20] and distributed NE seeking [21]. Obando et al. [4] studied a temperature control problem, where the population models the heating power to be distributed in a building and is coupled to a thermal model of the building. We refer the reader to a survey [14] for additional examples.

To the best of our knowledge, [5] is the first study that used δ -passivity as a design tool: a dynamic payoff mechanism was designed to lessen the impact of an epidemic subject to a limit on the long-term budget available to the decision maker. This work was extended to cases with nonnegligible disease mortality rates [6] and to scenarios with noisy payoffs to agents [22]. The same framework was also used to consider two-population scenarios [7].

Our work extends the design method in [5] to a larger class of ESs, which includes epidemic models as examples. Our assumptions are similar to those of previous studies, but the proposed dynamic payoff mechanism has a provable bound on the incentives provided to the agents. We also determine conditions under which a class of ESs coupled to a population of learning agents can be stabilized to a desired equilibrium.

III. POPULATION GAMES AND LEARNING RULES

We consider a population of a large number of nondescript agents, in which each agent follows a single strategy at any given time and can repeatedly revise its strategy, based on the payoffs to available strategies at the revision times. We assume that the agents have a common set of n strategies available to them. The instantaneous payoff obtained by following the i -th strategy at time t is given by $p_i(t)$, and the payoff vector offered to the population at time t is denoted as $p(t) := (p_i(t) \mid i \in [n])$, where $[n] := \{1, \dots, n\}$. The payoff perceived by agents at time t is the difference between

the rewards offered by the policy maker at time t , which is denoted by $r(t)$, and the vector that contains the intrinsic costs of the n strategies, which we denote by c . Thus, the payoff vector at time t is given by

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

These assumptions render the tools of population games well-suited for analyzing the strategic interactions among the agents. The population state at time t is denoted by $x(t) := (x_i(t) \mid i \in [n])$, with $x_i(t)$ being the proportion of the population following the i -th strategy at time t . The vector $x(t)$ takes values in the standard simplex

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

In the large-population limit, for $t \geq 0$, the population state x evolves according to the Evolutionary Dynamics Model (EDM)

$$\dot{x}(t) = V(x(t), p(t)), \quad (\text{EDM})$$

where $V : \mathbb{X} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$. The i -th element of $V(x, p)$ is

$$V_i(x, p) := \sum_{j=1}^n (x_j \tau_{ji}(x, p) - x_i \tau_{ij}(x, p)), \quad (2)$$

with a learning rule (also referred to as a revision protocol) τ that is a Lipschitz continuous map $\tau : \mathbb{X} \times \mathbb{R}^n \rightarrow [0, \bar{\tau}]^{n \times n}$, upper bounded by $\bar{\tau} > 0$.

We consider the EDM coupled to an ES, whose state at time t is denoted by $y(t)$. The ES state $y(t)$ takes values in $\mathbb{Y} \subset \mathbb{R}^m$ and evolves according to

$$\dot{y}(t) = f(y(t); x(t)), \quad (3)$$

where $f : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}^m$ is locally Lipschitz continuous and the population state $x(t)$ acts as time-varying parameters of the ES. We assume that, for any $x \in \mathbb{X}$, if $x(t) \equiv x$ then (3) has a unique equilibrium denoted as $y^*(x)$.

The dynamics of rewards $r(t)$ offered to the agents is described by the following:

$$\begin{aligned} \dot{q}(t) &= G(y(t), x(t), q(t)), \\ r(t) &= H(y(t), x(t), q(t)), \\ q(0) &= q_0, \end{aligned} \quad (4)$$

where G and H form a dynamic payoff mechanism to be designed by the policy maker.

Definition 1 A learning rule τ is said to satisfy the positive correlation (PC) condition if the following holds: for all $(x, p) \in \mathbb{X} \times \mathbb{R}^n$,

$$V(x, p) \neq 0 \Rightarrow p^\top V(x, p) > 0.$$

Definition 2 A learning rule τ is Nash Stationary (NS) if, given the best response map $\mathcal{M} : \mathbb{R}^n \rightarrow 2^{\mathbb{X}}$, where

$$\mathcal{M}(p) := \arg \max_{x \in \mathbb{X}} p^\top x, \quad p \in \mathbb{R}^n,$$

the following holds:

$$V(x, p) = 0 \Leftrightarrow x \in \mathcal{M}(p), p \in \mathbb{R}^n.$$

Definition 3 An EDM is δ -passive if there exist (i) a differentiable function $\mathcal{S} : \mathbb{X} \times \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$ and (ii) a Lipschitz continuous function $\mathcal{P} : \mathbb{X} \times \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$, which satisfy the following inequality for all x, p and u in $\mathbb{X}, \mathbb{R}^n$ and $\mathbb{R}^n$, respectively:

$$\frac{\partial \mathcal{S}}{\partial x}(x, p)V(x, p) + \frac{\partial \mathcal{S}}{\partial p}(x, p)u \leq -\mathcal{P}(x, p) + u^\top V(x, p) \quad (5)$$

where $\mathcal{S}$ and $\mathcal{P}$ must also satisfy the equivalences below:

$$\begin{aligned} \mathcal{S}(x, p) = 0 &\Leftrightarrow V(x, p) = 0, \\ \mathcal{P}(x, p) = 0 &\Leftrightarrow V(x, p) = 0. \end{aligned}$$

Since the EDM is determined by the learning rule, we say that the learning rule is δ -passive if the resulting EDM is δ -passive. Two well-known classes of learning rules that satisfy PC and NS conditions and lead to δ -passive evolutionary dynamics are the *separable excess payoff target* and the *impartial pairwise comparison* learning rules.

Example 1 A learning rule τ is said to be of the *separable excess payoff target type* [23] if, for each j in $[n]$, there is some $\rho_j : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ such that

$$\tau_{ij}(x, p) = \rho_j(p_j - x^\top p) \text{ for all } i \in [n],$$

and ρ_j satisfies $\rho_j(v) = 0$ for $v \leq 0$ and $\rho_j(v) > 0$ for $v > 0$.

Example 2 A learning rule is said to be of the *impartial pairwise comparison type* [24] if, for each j in $[n]$, there is some $\rho : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ such that

$$\tau_{ij}(x, p) = \rho_j(p_j - p_i) \text{ for all } i \in [n],$$

and ρ_j satisfies $\rho_j(v) = 0$ for $v \leq 0$ and $\rho_j(v) > 0$ for $v > 0$.

Lastly, we introduce a lemma that will be useful for proving our main result (Theorem 1) in the following section.

Lemma 1 For any fixed $q \in \mathbb{R}^n$ and $\bar{x} \in \mathcal{M}(q)$, the only vector $x \in \mathbb{X}$ that satisfies

$$x \in \mathcal{M}(\bar{x} - x + q) \quad (6)$$

is $x = \bar{x}$.

Proof: Since $\bar{x} \in \mathcal{M}(q)$, it is a solution to (6). To see that no other solution exists, rewrite (6) as

$$(y - x)^\top (x - \bar{x} - q) \geq 0 \quad \forall y \in \mathbb{X}.$$

Define $f(x) = \|x - \bar{x} - q\|_2^2$, and note from the inequality above

$$(y - x)^\top \nabla f(x) \geq 0 \quad \forall y \in \mathbb{X}.$$

It then follows from the minimum principle that $x \in \arg \min f(z)$ s.t. $z \in \mathbb{X}$. As this is a convex problem with strictly convex objective, it has a unique solution. Since $\bar{x}$ is a solution as noted above, the lemma follows. ■

IV. MAIN RESULT

Our goal is to design a dynamic payoff mechanism given by the maps G and H , which not only guarantees the convergence of the population state to some $x^* \in \mathbb{X}$ selected by the policy maker, but also gives bounds on the ES state $y(t)$ and the instantaneous cost $r(t)$ incurred by the policy maker. Our convergence result assumes that the population adopts a learning rule that is δ -passive, NS and PC, along with some conditions on the ES in (3). For the incentive design, we do not need to know the learning rule used by the agents as long as it satisfies the properties above.

Although our assumptions are similar to those in the previous works, we do not require the assumption in equation [5, (13)] on the storage function associated with the learning rule, but instead assume that the learning rule is PC, which is simpler to check.

Our approach to designing the dynamic payoff mechanism leverages the δ -passivity of the EDM, which yields a Lyapunov function for the overall system. The Lyapunov function is used to bound $(y, x)(t)$ based on the initial condition $(y, x)(0)$. Moreover, the maps G and H , combined with the bound on $(y, x)(t)$, also enable us to bound the instantaneous rewards provided to the agents, as discussed in §V.

Theorem 1 Consider a payoff vector $p^* \in \mathbb{R}^n$, a population state $x^* \in \mathcal{M}(p^*)$, and positive design parameters k_1, k_2 , and k_3 . Suppose that the exogenous system (3) satisfies the following:

- (i) There is a nonnegative continuously differentiable function $\mathcal{U} : \mathbb{Y} \times \mathbb{X} \rightarrow \mathbb{R}_{\geq 0}$ such that, for any $x \in \mathbb{X}$ and $y \in \mathbb{Y}$, $\frac{\partial \mathcal{U}}{\partial y}(y; x)f(y; x) \leq 0$.
- (ii) For any $\alpha \in \mathbb{R}_{\geq 0}$, $\{(y, x) \in \mathbb{Y} \times \mathbb{X} \mid \mathcal{U}(y; x) \leq \alpha\}$ is compact.
- (iii) For every $x \in \mathbb{X}$, the set $\{y^*(x)\}$ is the largest invariant subset of $\{y \in \mathbb{Y} \mid \frac{\partial \mathcal{U}}{\partial y}(y; x)f(y; x) = 0\}$.
- (iv) The function $\mathcal{U}$ satisfies $\frac{\partial \mathcal{U}}{\partial x}(y^*(x); x) = \mathbf{0}$ for every $x \in \mathbb{X}$.

In addition, assume that

- (v) the learning rule τ is Nash stationary, δ -passive, and positively correlated.

Then, the dynamic payoff mechanism given by

$$\begin{aligned} G(y, x, q) &= -k_1 \nabla_x \mathcal{U}(y; x) \\ &\quad - k_2(x - x^*) - k_3(q - p^*), \\ H(y, x, q) &= c + q, \end{aligned} \quad (7)$$

guarantees that, for any initial condition $(y, x)(0) \in \mathbb{Y} \times \mathbb{X}$ and q_0 in $\mathbb{R}^n$, we have $(y, x, q)(t) \xrightarrow{t \rightarrow \infty} (y^*(x^*), x^*, p^*)$.

Proof: Define the following candidate Lyapunov function for the overall system comprised of (2), (3), and (4).

$$\begin{aligned} \mathcal{L}(y, x, q) &:= k_1 \mathcal{U}(y; x) + k_3(\max_i p_i^* - x^\top p^*) \\ &\quad + \frac{k_2}{2} \|x - x^*\|_2^2 + \mathcal{S}(x, q), \end{aligned} \quad (8)$$

where $\mathcal{S}$ is the storage function of the EDM. Due to our selection of H , we have $p = q$.

We denote by $\dot{\mathcal{L}}(y, x, q)$ the directional derivatives of the function (8) along the vector field defined by (2), (3), and (4) at the point (y, x, q) :

$$\begin{aligned}\dot{\mathcal{L}}(y, x, q) &= k_1(\nabla_y \mathcal{U}(y; x)^\top f(y; x) + \nabla_x \mathcal{U}(y; x)^\top V(x, q)) \\ &\quad - k_3 V(x, q)^\top p^* + k_2(x - x^*)^\top V(x, q) \\ &\quad + \nabla_x \mathcal{S}(x, q)^\top f(y; x) + \nabla_p \mathcal{S}(x, q)^\top G(y, x, q) \\ &\leq k_1 \nabla_y \mathcal{U}(y; x)^\top f(y; x) + V(x, q)^\top G(y, x, q) \\ &\quad + (k_1 \nabla_x \mathcal{U}(y; x) + k_2(x - x^*) - k_3 p^*)^\top V(x, q) \\ &\quad - \mathcal{P}(x, q), \\ &= k_1 \nabla_y \mathcal{U}(y; x)^\top f(y; x) - k_3 q^\top V(x, q) - \mathcal{P}(x, q),\end{aligned}$$

where the inequality follows from (5). Conditions (i) and (v) imply that $\dot{\mathcal{L}}$ is nonpositive, and $\mathcal{L}$ is a nonstrict Lyapunov function.

By condition (ii) and $\dot{\mathcal{L}} \leq 0$, $\{(y, x, q)(t) | t \geq 0\}$ is bounded. Since (7) is a bounded-input bounded-output linear system with state p and bounded input $-k_1 \nabla_x \mathcal{U}(y; x) - k_2(x - x^*) + k_3 p^*$, which is a continuous function of the trajectory $\{(y, x, q)(t) | t \geq 0\}$, we obtain that $\{(y, x, q)(t) | t \geq 0\}$ is bounded.

Let $\mathbb{E}$ denote the largest invariant subset within $\{(y, x, q) | \dot{\mathcal{L}}(y, x, q) = 0\}$. From the LaSalle-Krasovskii invariance principle, $(y, x, q)(t)$ converges to the ω -limit set $L^+ \subset \mathbb{E}$, which is compact and invariant with respect to (2), (3), and (4) [25, Lemma 4.1].

For any $(y, x, q)(0) \in \mathbb{E}$, the trajectory will satisfy that, for all $t \geq 0$, $x(t) = x(0)$ and $y(t) = y^*(x(0))$ due to the fact that $\dot{\mathcal{L}}(y(t), x(t), q(t)) = 0$ for all $t \geq 0$ implies $\dot{x}(t) = 0$ for all $t \geq 0$, and condition (iii) in Theorem 1. This, together with condition (iv), leads to

$$\dot{q}(t) = -k_2(x(0) - x^*) - k_3(q(t) - p^*)$$

so that $q(t) \xrightarrow{t \rightarrow \infty} \frac{k_2}{k_3}(x^* - x(0)) + p^*$ and

$$(y, x, q)(t) \xrightarrow{t \rightarrow \infty} (y^*(x(0)), x(0), \frac{k_2}{k_3}(x^* - x(0)) + p^*).$$

Since the learning rule is assumed NS, we must have

$$x(0) \in \mathcal{M}\left(x^* - x(0) + \frac{k_3}{k_2} p^*\right),$$

and by Lemma 1 we have $x(0) = x^*$ and $L^+ = \{(y^*(x^*), x^*, p^*)\}$. ■

V. BOUNDS

As proven in Theorem 1, if the ES in (3) satisfies (i)-(iv) and the learning rule satisfies (v), the payoff mechanism described in (7) can be used to stabilize the ES and the population to a desired equilibrium. Also, we are able to determine bounds for the state and the rewards $r(t)$ offered to the population.

If $p(t) = \mathbf{0}$ at some t , the storage function of the EDM is equal to zero at t , as any $x(t) \in \mathbb{X}$ is a best response.

In particular, if $p(0) = p^* = \mathbf{0}$, both $\mathcal{S}(x(0), p(0))$ and $\max(p^*) - x(0)^\top p^*$ are equal to zero, and we have

$$\begin{aligned}\mathcal{L}_0 &:= \mathcal{L}(y(0), x(0), \mathbf{0}) \\ &= k_1 \mathcal{U}(y(0); x(0)) + \frac{k_2}{2} \|x(0) - x^*\|_2^2.\end{aligned}$$

Furthermore, because $\mathcal{L}$ is decreasing along trajectories and $\mathcal{S}(x, \mathbf{0}) = 0$ for any $x \in \mathbb{X}$, we obtain that, for any $t \geq 0$,

$$\mathcal{L}(y(t), x(t), \mathbf{0}) \leq \mathcal{L}(y(t), x(t), p(t)) \leq \mathcal{L}(y(0), x(0), \mathbf{0}). \quad (9)$$

This in turn can be used to bound not only $y(t)$ and $x(t)$ but also the policy maker's instantaneous cost $\bar{c}(t) := x(t)^\top r(t)$: for any $t \geq 0$,

$$\bar{c}(t) \leq \max\{g(y, x) | x \in \mathbb{X}, y \in \mathbb{Y}, \mathcal{L}(y, x, \mathbf{0}) \leq \mathcal{L}_0\}, \quad (10)$$

$$g(y, x) := \|c + G(y, x, \mathbf{0})/k_3\|_\infty,$$

where the terms $G(y, x, \mathbf{0})$, $\mathcal{L}(y, x, \mathbf{0})$ and $\mathcal{L}_0$ are affected by the choice of the parameters k_1 , k_2 , and k_3 .

VI. EXAMPLES

In §VI-A and §VI-B we present two systems that fit our framework as the ES in (3). They meet conditions (i)-(iv) of Theorem 1, and when coupled to a population that employs a learning rule satisfying condition (v), the dynamic payoff mechanism described in Theorem 1 stabilizes the overall system to a desired equilibrium $y^*(x)$. In §VI-C we consider a modification of [6] to exemplify how our theorem can be leveraged for design: We first select a target equilibrium population state x^* that minimizes the disease transmission rate subject to a budget constraint, and choose the parameters k_1 , k_2 , and k_3 so that the peak size of the infected population is guaranteed to be below a given threshold. We then present simulation results using several different learning rules.

Our examples focus on systems that are naturally coupled to a population of agents and are affected by the strategic choices of the agents. The system considered in [5], [6] is a compartmental model of an epidemic disease, and the population state affects the transmission rate of the disease. Similarly, [7] considers an epidemic model with two interacting populations, with the transmission rates of each population being affected by its agents' current strategies. Korobeinikov [8] studied a Host-Parasite model, and by finding a nonstrict Lyapunov function he proved the convergence to the unique equilibrium of the model. We modify this model so that some of its parameters change according to the population state. We choose a desirable equilibrium $(y^*(x^*), x^*)$ of this modified model to reduce the number of parasites at the equilibrium, and then use (7) to stabilize the equilibrium.

A. Leslie-Gower predator-prey model

Korobeinikov [8] studies the Leslie-Gower model that captures the interaction of populations of hosts and parasites. Let $O(t)$ and $P(t)$ denote the number of hosts and parasites, respectively, at time t . The population sizes evolve according to the following differential equations:

$$\dot{O}(t) = (z_1 - a_1 P(t) - b_1 O(t)) O(t), \quad (11a)$$

$$\dot{P}(t) = (z_2 - a_2 P(t)/O(t)) P(t), \quad (11b)$$

where z_1, z_2, a_1 , and a_2 are positive, and b_1 is nonnegative. The intrinsic population growth rates of the hosts and parasites are z_1 and z_2 , respectively. The parameter b_1 relates to a growth limit on the hosts without parasites, while a_1 relates to a decrease of hosts due the parasites, and a_2 relates to a population limit on the parasites due to the number of hosts. The unique co-existing equilibrium, where $O^*, P^* > 0$, is

$$O^* = \frac{z_1}{a_1 z_2 + a_2 b_1} \quad \text{and} \quad P^* = \frac{z_1 z_2}{a_1 z_2 + a_2 b_1}.$$

The following is a nonstrict Lyapunov function of (11) on $(0, \infty)^2$:

$$U(O, P) := \tilde{U}((O, P); (O^*, P^*)), \quad (12)$$

where

$$\tilde{U}(v; w) := \log\left(\frac{v_1}{w_1}\right) + \frac{w_1}{v_1} + \frac{a_1 w_1}{a_2} \left(\log\left(\frac{v_2}{w_2}\right) + \frac{w_2}{v_2} \right),$$

with $v, w \in \mathbb{R}_{>0}^2$. The directional derivatives of (12) along the vector field defined by (11) is

$$\dot{U}(O, P) = -\frac{a_1}{P}(P - P^*)^2 - \frac{b_1}{O}(O - O^*)^2$$

with positive a_1 and nonnegative b_1 . This confirms that (12) is a Lyapunov function of (11).

Suppose that z_1 and z_2 are functions of the population state, i.e., $z_1, z_2 : \mathbb{X} \rightarrow \mathbb{R}_{>0}$, and $b_1 : \mathbb{X} \rightarrow \mathbb{R}_{\geq 0}$. Such scenarios could arise when the agents are the farmers who breed and raise livestock, which are the hosts affected by parasites. The strategic choices of the agents could include, for example, how many animals to breed or which measures to take to reduce the spread of parasites, e.g., diagnosing, isolating, and treating infected hosts. In this case, the Leslie-Gower model as the ES in (3) is described by

$$f(y; x) := \begin{bmatrix} (z_1(x) - a_1 y_2 - b_1(x) y_1) y_1 \\ (z_2(x) - a_2 y_2 / y_1) y_2 \end{bmatrix}$$

with the equilibrium

$$y^*(x) := \left[\frac{\frac{z_1(x)}{a_1 z_2(x) + a_2 b_1(x)}}{\frac{z_1(x) z_2(x)}{a_1 z_2(x) + a_2 b_1(x)}} \right] \in \mathbb{Y} = (0, \infty)^2.$$

Suppose z_1, z_2 , and b_1 are continuously differentiable so that the equilibrium map y^* is also continuously differentiable.

Define

$$\mathcal{U}(y; x) := \tilde{U}(y; y^*(x)) - \frac{a_1}{a_2} y_1^*(x).$$

Note that (a) $\mathcal{U}(y; x) > 0$ for any $y \neq y^*(x)$, and (b) for any $x \in \mathbb{X}$, if $x(t) \equiv x$, then $\dot{\mathcal{U}}(y(t); x(t)) = \dot{U}(y_1(t), y_2(t))$. Thus, it satisfies conditions (i) and (iii). As a continuous function, the sublevel sets of $\mathcal{U}$ are closed, and we can verify that they are also bounded so that condition (ii) is satisfied. Lastly, $\mathcal{U}$ satisfies (iv):

$$\begin{aligned} & \frac{\partial \mathcal{U}}{\partial x}(y^*(x); x) \\ &= \left(\frac{\partial \tilde{U}}{\partial w}(y^*(x); y^*(x)) - \frac{a_1}{a_2} \begin{bmatrix} 1 \\ 0 \end{bmatrix}^\top \right) \frac{\partial y^*}{\partial x}(x) = \mathbf{0}, \end{aligned}$$

because

$$\frac{\partial \tilde{U}}{\partial w}(y^*(x); y^*(x)) = \frac{a_1}{a_2} \begin{bmatrix} 1 \\ 0 \end{bmatrix}^\top.$$

B. Epidemic Population Games (EPG)

Previous studies on EPG [5], [6], [7] examined epidemic compartmental models coupled to a population, with the epidemic model being the ES. Here we show that the model used in [6] satisfies the conditions in Theorem 1, even for a more general dependency of the transmission rates on the agents' strategies than that considered in [6]. The epidemic model satisfies conditions (i)-(iv) and, when coupled to a population employing a learning rule that satisfies (v), we can use Theorem 1 to drive them to a desirable equilibrium. A similar analysis shows that the epidemic models in [5], [7] also satisfy conditions (i)-(iv).

We first briefly describe the normalized susceptible-infectious-recovered-susceptible (SIRS) model, which is the ES we aim to stabilize. Let $I(t)$, and $R(t)$ denote the proportions of infected agents and recovered agents, respectively, in the population at time t . Suppose that $N(t)$ is the population size at time t . The population size changes according to $\dot{N}(t) = (g - \delta I(t))N(t)$, where $g := \theta - \zeta$ is the difference between the birth rate θ and the natural death rate ζ , and $\delta > 0$ is the disease death rate. The disease recovery rate and the rate at which a recovered individual becomes susceptible again due to waning immunity are denoted by γ and ψ , respectively.

Since the model is normalized, $I(t) \in [0, 1]$ and $R(t) \in [0, 1 - I(t)]$ at any time t . The ES state $y(t)$ is given by $(I(t), R(t))$ and evolves according to

$$\dot{I}(t) = (\mathcal{B}(t)S(t) + \delta I(t) - \sigma)I(t), \quad (13a)$$

$$\dot{R}(t) = \gamma I(t) - \omega R(t) + \delta R(t)I(t), \quad (13b)$$

where $\mathcal{B}(t)$ is the average transmission rate at time t , $\bar{\sigma} := \gamma + \zeta + \delta$, $\sigma := g + \bar{\sigma}$, $\bar{\omega} := \psi + \zeta$, $\omega := g + \bar{\omega}$, and $\bar{\sigma}^{-1}$ is the mean infectious period for an affected individual (till recovery or death). The adopted time unit is one day, and newborns are assumed susceptible. As in [6], we assume $\delta > 0$ but moderate such that $\delta < \min\{\omega, \gamma\}$. Also, $\mathcal{B}(t) > \sigma$ for all $t \geq 0$ so that there is a unique endemic equilibrium.

For fixed $\mathcal{B}(t) \equiv B > \sigma$, the endemic equilibrium of (13) is given by the differentiable functions of B :

$$\begin{aligned} I_B^* &:= \frac{b_B - \sqrt{\Delta}}{2\delta(B - \delta)}, \quad \text{and} \\ R_B^* &:= (1 - \sigma/B) - (1 - \delta/B)I_B^*, \end{aligned}$$

where $b_B := \gamma B + \omega(B - \delta) + \delta(B - \sigma)$, and the discriminant is $\Delta := b_B^2 B - 4\delta\omega(B - \delta)(B - \sigma)$.

Now, suppose $\mathcal{B}(t) \equiv B(x(t))$, where $B : \mathbb{X} \rightarrow (\sigma, \infty)$ is a continuously differentiable function, and let $\mathbb{Y} := (0, 1] \times [0, 1]$ and $y^*(x) := (I_{B(x)}^*, R_{B(x)}^*)$. For a fixed $x(t) \equiv x$, the following is a strict Lyapunov function for (13) on $\mathbb{Y}$:

$$\mathcal{U}(I, R; x) := \tilde{U}(I, R; B(x)), \quad (15)$$

where

$$\tilde{\mathcal{U}}(I, R; B) := (I - I_B^*) + I_B^* \ln \frac{I}{I_B^*} + \frac{a_B}{2} (R - R_B^*)^2,$$

and $a_B := B/(\gamma + \delta R_B^*)$. The derivative of (15) along trajectories is

$$\begin{aligned} \frac{d}{dt} \tilde{\mathcal{U}}(I(t), R(t); B) = & - (B - \delta)(I(t) - I_B^*)^2 \\ & - a_B(\omega - \delta I(t))(R(t) - R_B^*)^2, \end{aligned}$$

and it is negative for any $(I(t), R(t)) \in \mathbb{Y} \setminus \{y^*(x)\}$.

Since (15) is a strict Lyapunov function when $x(t)$ is constant, it satisfies conditions (i) and (iii). As $\mathcal{U}$ is a continuous function, its sublevel sets are closed. Moreover, because the sublevel sets are also contained in $\mathbb{Y} \times \mathbb{X}$, which is a bounded set, they are also bounded and, hence, condition (ii) holds. Lastly, we can verify condition (iv) as follows.

$$\frac{\partial \mathcal{U}}{\partial x}(I_{B(x)}^*, R_{B(x)}^*; x) = \frac{\partial \tilde{\mathcal{U}}}{\partial B}(I_{B(x)}^*, R_{B(x)}^*; B(x)) \frac{\partial B}{\partial x}(x) = 0,$$

where the second equality follows from $\frac{\partial \tilde{\mathcal{U}}}{\partial B}(I_{B(x)}^*, R_{B(x)}^*; B(x)) = 0$.

C. Designing an Intervention Policy for Epidemics with Nonlinear Infection Rate

We consider a modification of the EPG studied in [6], which was described in §VI-B, as an application of Theorem 1 to a dynamic payoff design problem. We aim to mitigate an epidemic outbreak and reduce the endemic level of infected agents while guaranteeing that the long-term cost of the policy maker does not exceed some available budget c^* .

The study in [6] considers the average transmission rate that depends linearly on the population state, with $\mathcal{B}(t) \equiv \beta^\top x(t)$, where $\beta \in \mathbb{R}_{>0}^n$. Such dependency on $x(t)$ is consistent with the choices of the susceptible agents determining the likelihood of contracting the disease when exposed, e.g., choosing to wear masks or getting vaccinated. In their model, the proportion of susceptible agents following the i -th strategy at time t is $x_i(t)S(t)$, and for those agents the rate of new infections is equal to $\beta_i x_i(t)S(t)I(t)$.

Suppose that we allow the average transmission rate to depend on both the choices of the susceptible agents and those of the infected agents. For example, an infected agent that takes no preventive measures is more likely to transmit the disease than another infected agent that does take preventive measures. In this case, the rate of new infections among susceptible agents following strategy i due to the exposure to infected agents adopting strategy j would be $\beta_{ij} x_i(t)S(t)x_j(t)I(t)$. Therefore, the average transmission rate is given by

$$\mathcal{B}(t) \equiv B(x(t)) := x(t)^\top Q x(t), \quad t \geq 0,$$

with $Q \in \mathbb{R}^{n \times n}$ being a positive matrix with the elements $Q_{ij} = \beta_{ij}$. We assume that the disease is too infectious to be eradicated so that $B(x) \geq \sigma$ for all $x \in \mathbb{X}$.

We aim to select a target population equilibrium x^* that minimizes the transmission rate subject to a budget constraint. In order to find x^* we solve

$$x^* \in \arg \min_{z \in \mathbb{X}} B(z) \quad \text{s.t.} \quad c^\top z - \min_i c_i \leq c^*,$$

where c is the vector of intrinsic costs of the strategies, and c^* is the long-term budget available to the policy maker.

After determining x^* , we select $k_1, k_2, k_3 > 0$ and use the payoff mechanism described in Theorem 1 to lead the population and epidemic states to the selected equilibrium.

Example 3 Consider a disease with parameters: $\delta = 0.005$, $\zeta = 0$, $\theta = 0.0002$, $\gamma = 0.1$ (mean recovery period ~ 10 days) and $\bar{\omega} = 0.011$ (mean immunity period ~ 91 days). Agents have three available strategies with

$$Q = \begin{bmatrix} 0.13 & 0.18 & 0.2 \\ 0.16 & 0.22 & 0.23 \\ 0.17 & 0.28 & 0.5 \end{bmatrix}, \quad \text{and} \quad c = \begin{bmatrix} 0.2 \\ 0.1 \\ 0 \end{bmatrix}.$$

The initial conditions are $I(0) = 0.019$ and $R(0) = 0.172$ for the epidemic, $x(0) = (1 \ 0 \ 0)^\top$ for the EDM, and $q(0) = (0 \ 0 \ 0)^\top$ for the payoff dynamics. The long-term budget of the policy maker is $c^* = 0.1$, and we select $p^* = \mathbf{0}$, which yields $x^* \approx (1 \ 10 \ 1)/12$ and $y^*(x^*) \approx (5.1\%, 46.9\%)$. Our goal is to design G and H so that $I(t) \leq 10\%$ for all $t \geq 0$.

For simplicity, we select $q_0 = -c$ and $p^* = p(0) = \mathbf{0}$ and $k_3 = 1$, as k_3 does not affect the bound on $I(t)$ if $p^* = \mathbf{0}$. Based on (9), we look for values of k_1 and k_2 that meet our requirement that $I(t) \leq 10\%$ for all $t \geq 0$, by solving

$$\begin{aligned} I_{\max}(k_1, k_2) := & \max_{I, R, x} I \\ & \text{s.t. } x \in \mathbb{X}, \\ & I, R \geq 0, \\ & I + R \leq 1, \\ & \mathcal{L}(y, x, 0) \leq \mathcal{L}(y(0), x(0), 0), \end{aligned} \quad (16)$$

where k_1 and k_2 affect constraint (16).

We solve numerically the optimization above for several values of k_1 and k_2 , and show the results in Fig. 1. The requirement that $I(t) \leq 10\%$ is met for k_1, k_2 in the region on the bottom right of the plot, and we select $k_1 = 2$ and $k_2 = 0.022$. We then use the reward mechanism given by Theorem 1 to guide the population to the desired equilibrium.

The simulation results for several different learning rules that satisfy condition (iv) in Theorem 1 are shown in Fig. 2.¹ It is clear that they all converge and satisfy the requirement that $I(t) \leq 10\%$ for all $t \geq 0$. Had we not used the bound, $I_{\max}(k_1, k_2)$, to determine the parameters of the reward mechanism, the bound on the peak of infections may have been violated, as shown in Fig. 3. In both figures we observe that the instantaneous cost, $\bar{c}(t)$, converges to c^* , which is represented as a dashed black line in the plots.

¹See the simulation code at [github/jcert/incentive-design-coupled-dynamics](https://github.com/jcert/incentive-design-coupled-dynamics) for more details on the learning rules that were used.

VII. CONCLUSION

We studied a large population of learning agents whose strategic choices influence the dynamics of an ES we seek to stabilize at a desired equilibrium. Our framework can be used to design a dynamic payoff mechanism that guarantees the convergence of both the population and the ES (Theorem 1). When the conditions on the ES stated in the theorem are met, the designed incentives can stabilize more general systems than previously considered.

We also presented example systems that satisfy the conditions of our main result (§VI-A and §VI-B) and applied our framework to design incentives that mitigate an epidemic with nonlinear infection rates subject to long-time budget constraints (§VI-C). Unlike the incentives designed in the previous studies of EPG, our payoff mechanism is guaranteed to have a bound on the instantaneous reward offered to the population.

In future research we plan to extend the results by relaxing the assumptions on the ES to be only local and to allow the EDM to depend on the ES states. Another direction we are interested in pursuing is to examine design problems where the payoff of certain strategies can only be partially designed.

REFERENCES

- [1] M. Arcak and N. C. Martins, “Dissipativity Tools for Convergence to Nash Equilibria in Population Games,” *IEEE Control Netw. Syst.*, vol. 8, no. 1, pp. 39–50, Mar. 2021.
- [2] A. Pantoja and N. Quijano, “A population dynamics approach for the dispatch of distributed generators,” *IEEE Transactions on Industrial Electronics*, vol. 58, no. 10, pp. 4559–4567, Oct. 2011.
- [3] S. Park, Y. D. Zhong, and N. E. Leonard, “Multi-Robot Task Allocation Games in Dynamically Changing Environments,” in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. Xi’an, China: IEEE, May 2021, pp. 8678–8684. [Online]. Available: <https://ieeexplore.ieee.org/document/9561809/>
- [4] G. Obando, A. Pantoja, and N. Quijano, “Building temperature control based on population dynamics,” *IEEE transactions on Control Systems Technology*, vol. 22, no. 1, pp. 404–412, Jan. 2014.
- [5] N. C. Martins, J. Certório, and R. J. La, “Epidemic population games and evolutionary dynamics,” *Automatica*, vol. 153, p. 111016, Jul. 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0005109823001711>
- [6] J. Certório, N. C. Martins, and R. J. La, “Epidemic Population Games With Nonnegligible Disease Death Rate,” *IEEE Control Systems Letters*, vol. 6, pp. 3229–3234, 2022.
- [7] J. Certório, R. J. La, and N. C. Martins, “Epidemic Population Games for Policy Design: Two Populations with Viral Reservoir Case Study,” in *Proc. IEEE conf. decis. control (CDC)*, 2023, pp. 7667–7674. [Online]. Available: <https://ieeexplore.ieee.org/document/10383665>
- [8] A. Korobeinikov, “A Lyapunov function for Leslie-Gower predator-prey models,” *Appl. Math. Lett.*, vol. 14, no. 6, pp. 697–699, Aug. 2001. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S089396590180029X>
- [9] W. H. Sandholm, “Potential games with continuous player sets,” *Journal of economic theory*, vol. 97, pp. 81–108, 2001.
- [10] J. Hofbauer and W. H. Sandholm, “Stable games and their dynamics,” *Journal of Economic Theory*, vol. 144, no. 4, pp. 1665–1693.e4, Jul. 2009.
- [11] S. Kara, N. C. Martins, and M. Arcak, “Population games with erlang clocks: Convergence to nash equilibria for pairwise comparison dynamics,” in *Proc. IEEE conf. decis. control (CDC)*, 2022, pp. 7688–7695.
- [12] S. Kara and N. C. Martins, “Excess Payoff Evolutionary Dynamics With Strategy-Dependent Revision Rates: Convergence to Nash Equilibria for Potential Games,” *IEEE Control Systems Letters*, vol. 7, pp. 1009–1014, 2023.

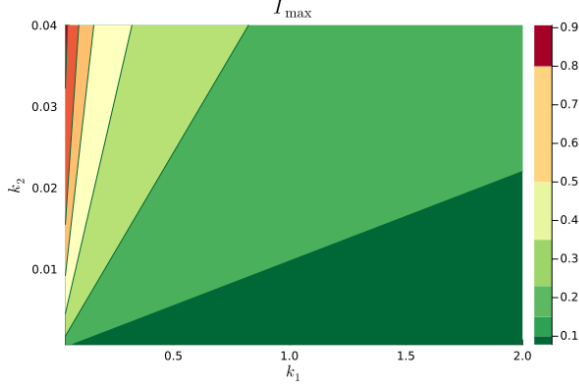


Fig. 1. Bound $I_{\max}(k_1, k_2)$, for the conditions in Example 3, when varying k_1 and k_2 .

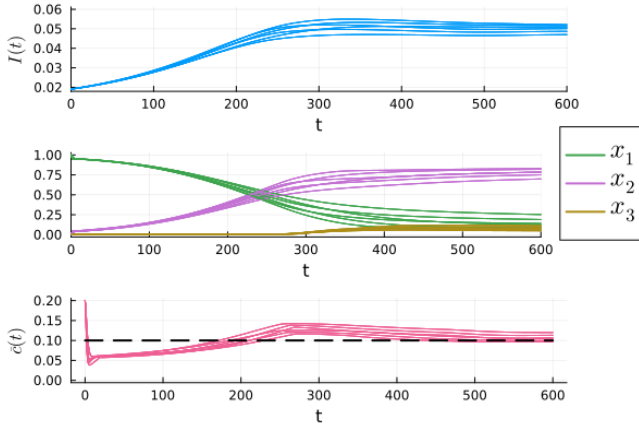


Fig. 2. Simulation of Example 3, for many different learning rules, using $k_1 = 2, k_2 = 0.022, k_3 = 1$.

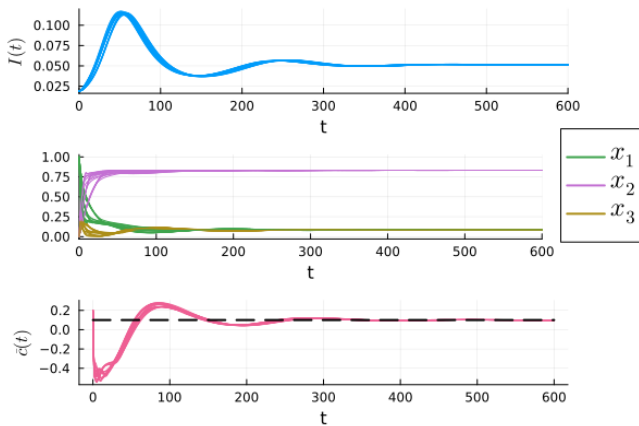


Fig. 3. Simulation of Example 3, for many different learning rules, using $k_1 = k_2 = k_3 = 1$.

- [13] W. H. Sandholm, *Population games and evolutionary dynamics*. MIT Press, 2010.
- [14] N. Quijano, C. Ocampo-Martinez, J. Barreiro-Gomez, G. Obando, A. Pantoja, and E. Mojica-Nava, "The Role of Population Games and Evolutionary Dynamics in Distributed Control Systems: The Advantages of Evolutionary Game Theory," *IEEE Control Syst.*, vol. 37, no. 1, pp. 70–97, 2017. [Online]. Available: <https://ieeexplore.ieee.org/document/7823106/>
- [15] M. J. Fox and J. S. Shamma, "Population games, stable games, and passivity," *Games*, vol. 4, pp. 561–583, 2013.
- [16] S. Park, N. C. Martins, and J. S. Shamma, "Payoff dynamics model and evolutionary dynamics model: Feedback and convergence to equilibria," 2020.
- [17] K. S. Schweidel and M. Arcak, "Compositional Analysis of Inter-connected Systems Using Delta Dissipativity," *IEEE Control Systems Letters*, vol. 6, pp. 662–667, 2022.
- [18] M. A. Mabrok and J. S. Shamma, "Passivity analysis of higher order evolutionary dynamics and population games," in *Proc. IEEE conf. decis. control (CDC)*, 2016, pp. 6129–6134. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/7799211>
- [19] S. Park, J. S. Shamma, and N. C. Martins, "Passivity and evolutionary game dynamics," in *Proc. IEEE conf. decis. control (CDC)*, 2018, pp. 3553–3560.
- [20] J. Martinez-Piazuelo, N. Quijano, and C. Ocampo-Martinez, "A Payoff Dynamics Model for Equality-Constrained Population Games," *IEEE Control Systems Letters*, vol. 6, pp. 530–535, 2022.
- [21] J. Martinez-Piazuelo, C. Ocampo-Martinez, and N. Quijano, "On Distributed Nash Equilibrium Seeking in a Class of Contractive Population Games," *IEEE Control Systems Letters*, pp. 1–1, 2022.
- [22] S. Park, J. Certório, N. C. Martins, and R. J. La, "Epidemic population games and perturbed best response dynamics," 2024.
- [23] W. H. Sandholm, "Excess payoff dynamics and other well-behaved evolutionary dynamics," *Journal of Economic Theory*, vol. 124, no. 2, pp. 149–170, Oct. 2005. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0022053105000554>
- [24] —, "Pairwise comparison dynamics and evolutionary foundations for Nash equilibrium," *Games*, vol. 1, no. 1, pp. 3–17, 2010.
- [25] H. K. Khalil, *Nonlinear systems*. Prentice Hall, 1995.