Social Planning with the Replicator Dynamics

Anand Siththaranjan and Claire J. Tomlin

***Abstract*— Approaches to social planning tend to assume that the behavior of agents is at an equilibrium, yet in practice people's behavior gradually adapts to their experiences. In this work, a model of social planning under the replicator dynamics is studied. This model allows for a social planner to control the learning process of agents by influencing the relative fitness of different strategies. The desiderata that such a social planner would ideally achieve – exponential stability and budget-balance – are described. Existence of a solution for any full-support distribution, as well as an analysis of its properties, are shown constructively by leveraging classical tools from geometric control theory. Though the solution is optimal in an environment without transfer costs, this may not generally hold otherwise. We formulate a relevant optimal control problem to model this setting, and determine performance guarantees based in our original solution.**

I. Introduction

When agents' behavior evolves according to a learning process, how can social planners implement desirable outcomes? The replicator dynamics provides a model based in natural selection of how populations co-evolve due to their relative fitness. This model has been used by researchers to study congestion games in traffic and network applications, where a continuum of agents is treated as flows through a network, interactions between biological species, search with positive externalities, and more [1]. In particular, agents adapt their behavior by myopically imitating observed outcomes better than their own with probability proportional to the degree to which such an outcome was superior. For example, consider a setting in which agents in a traffic network aim to get to their destination quickly. Agents with the same destination may choose to take different paths to get there. Upon arrival, they can observe the paths taken by others and their duration. When deciding which path they take the next day, an agent chooses to take faster paths observed the previous day with higher probability.

In this work, we study a complete information two-player symmetric game where agents update their strategies via the replicator dynamics, and designers provide monetary transfers as incentives to reach an optimal distribution of strategies. We assume agents have quasi-linear utilities, which dictate how a planner interacts with the agents. We first formulate this model as a control system based in the replicator dynamics, and study solutions to this system whereby a planner can condition the transfers at a specific time on the distribution of strategies taken only at that time. Our goal is to identify a mechanism by which the planner may implement a given distribution of strategies as the limit point of this control system. This is while maintaining the properties of exponential stability and budget-balance. Though we assume a two player game structure, our results generalize straightforwardly to population games.

Our main theorem is that when aiming to stabilize to **any** distribution of strategies with full support[1], there always exists an exponentially stabilizing and budget-balanced feedback controller. That is, there is a mechanism that induces a controller which satisfies these desirable properties. We further derive standard bounds on the time-to-reach, robustness and optimality of this mechanism.

A. Related Work

In evolutionary game theory, the main area of relevant study is *evolutionary implementation*. This literature primarily studies implementing efficient outcomes with incomplete information, such as through Pigouvian pricing [2]. Such literature has studied logit dynamics [2], aggregative games [3], [4], and more. Though we focus on the deterministic replicator dynamics with quasi-linear utility under complete information, rather than showing implementation of only Pareto efficient outcomes under certain classes of relevant games, we show nearly all outcomes can be implemented in arbitrary games. Furthermore, we provide rates of exponential convergence that can be chosen by the designer, and don't make assumptions about properties of an agent's non-monetary utility.

Similar to us, other works have leveraged traditional control-theoretic tools. For example, adaptive gain controllers can be used to incentivize cooperative behaviour in 2x2 games [5] under the replicator dynamics. Further work has generally explored equilibrium selection in such games [6], specifically those dominant strategy and anti-coordination structures. We differ in focusing on general games and target distributions beyond equilibria, however we assume complete information about the payoffs whereas these other approaches require less information due to their limitation to certain games. Geometric techniques, which are key to feedback linearization, have also been applied to the replicator dynamics [7]. However they focus on a different control structure that affects the magnitude of the fitness equations. Leveraging feedback control to influence population dynamics can be viewed as enriching the causal map from the social state to payoffs as in a payoff dynamic model (PDM) [8]. We do not however consider richer environments with delays for example, that is possible to model

Authors are with the University of California, Berkeley, Department of Electrical Engineering and Computer Science (address: 2594 Hearst Ave, Berkeley, CA 94720, USA) (e-mail: {anandsranjan,tomlin}@berkeley.edu).

[1]All strategies are played with positive probability.

with PDMs.

A line of literature at the intersection of computer science and game theory that is similar to our work is called *adaptive incentive design* [9]. This work studies how planners can dynamically design incentives to guide their decision-making when agents employ some unknown learning process inspired by machine learning approaches. We differ in that we employ additional constraints such as budget-balance under different dynamics. Further work in this direction studies a very similar setting to our work [10], [11], where they leverage two-timescale dynamics and the convergence of best-response dynamics in potential games to achieve similar results. Again, our work differs in that we consider different dynamics, arbitrary games, and budget-balanced controllers. Along with [10], [12], we apply our solution to traffic tolling. However, these works leverage logit and best-response dynamics in their analysis.

II. MODEL

Let $\mathcal{S}$ be a **finite set of strategies** that can be played by each agent in a two-agent symmetric game. We denote $\Delta(\mathcal{S})$ as the set of probability distributions on $\mathcal{S}$, where $p \in \Delta(\mathcal{S})$ is such that $p(s)$ is the probability of $s \in \mathcal{S}$. Distributions over strategies with full support are elements of the relative interior, denoted relint$(\Delta(\mathcal{S}))$, i.e. relint$(\Delta(\mathcal{S})) = \{p \in \Delta(\mathcal{S}) : p(s) > 0 \ \forall s \in \mathcal{S}\}$. Denote the size of the set by $|\mathcal{S}| = \sum_{s\in\mathcal{S}} 1$. Let $u(s, s')$ be the utility achieved by $s \in \mathcal{S}$ when playing against $s' \in \mathcal{S}$. Define $\theta_t \in \Delta(\mathcal{S})$ as the distribution of agents that play a strategy at time $t \geq 0$. For example, $\theta_t(s) = 1$ implies that all agents are playing s at time t. Note that $\forall t$, $\theta_t(s) \geq 0 \ \forall s$ and $\sum_{s\in\mathcal{S}} \theta_t(s) = 1$. At any given time t, define $u_t(s) = \mathbb{E}_{s'\sim\theta_t}[u(s, s')]$. This is the expected utility of playing strategy s against a population distribution θ_t, where we suppress reference to θ_t when it is clear. We define $\bar{u}_t = \mathbb{E}_{s\sim\theta_t}[u_t(s)]$ as the average utility across strategies at distribution θ_t. These allow us to define the continuous-time evolution of a population distribution $\theta_t(s)$ by the replicator dynamics with respect to some initial distribution $\hat{\theta}_0 \in \Delta(\mathcal{S})$: $\dot{\hat{\theta}}_t(s) = \hat{\theta}_t(s)[u_t(s) - \bar{u}_t]$, where $\dot{\hat{\theta}}_t(s)$ is interpreted as the time derivative of $\hat{\theta}_t(s)$. We say that θ is a (symmetric) Nash equilibrium if it is a fixed point of the best-response map. That is, $\theta \in \arg\max_{\hat{\theta}\in\Delta(\mathcal{S})} \mathbb{E}_{s\sim\hat{\theta},s'\sim\theta}[u(s, s')]$. The relationship between Nash equilibria and limit points of the replicator dynamics is well-known, and more details can be found in [13].

To extend the replicator dynamics to the setting where a designer can affect utilities, we introduce a *pricing strategy*, or controller, $k : \Delta(\mathcal{S}) \times \mathcal{S} \to \mathbb{R}$ where $k(\theta, s)$ for $s \in \mathcal{S}$ is the transfers given to agents that play strategy s when the distribution of strategies in the population is θ. Let $\mathcal{K}$ be the set of controllers. We assume preferences over outcomes in the game and transfers given by $k \in \mathcal{K}$ are determined by $\hat{u}$: $\hat{u}(s, s', \theta) = u(s, s') + k(\theta, s)$. Inherent in this is that agents have quasilinear preferences, with transfers entering linearly. This is a standard assumed form of an agent's utility function. Due to the following, $\hat{u}(s) = u(s) + k(\theta, s)$, we can redefine our evolutionary dynamics as

$$\begin{aligned}\dot{\theta}_t(s) &= \theta_t(s)\left[u(s) + k(\theta_t, s) - \bar{\hat{u}}_t\right] \\ &= f_s(\theta_t) + g_s(\theta_t)k_t(s) - \bar{\hat{u}}_t\end{aligned}$$

where $k_t(s) = k(\theta_t, s)$ for simplicity and $\bar{\hat{u}}_t = \bar{u} + \sum_{s\in\mathcal{S}} \theta_t(s)k_t(s)$. Note that by this structure, the dynamics of population s, $\dot{\theta}_t(s)$, is directly affected by the controller for other strategies, ie. $\{k_t(s')\}_{s'\neq s}$. This is intuitively reasonable: our control model provides an extrinsic utility to different strategy populations, hence the relative fitness of each strategy may differ depending on the perturbed utility of other strategies. As such, the control system is nearly of a typical form: for each $s \in \mathcal{S}$. the first portion $f_x(\theta_t) + g_s(\theta_t)k_t(s)$ is in control affine form, but the second part $\bar{u}'_t = \bar{u} + \sum_{s'\in\mathcal{S}} \theta_t(s')k_t(s')$ is not.

A. Anonymity

Our model is one of population dynamics and thus we have no further information on individual agents. A benefit of this is that a mechanism cannot discriminate based on identity, rather only on the actions taken. This is favorable from a privacy perspective, since the planner does not require both identity and action information. We could have considered a model with identity, and thus we can consider a mechanism that conditions based on identity and action. This makes the problem trivial: we can merely sort agents into groups and punish them with large negative transfers if they do not take a prescribed action, which we choose such that it induces the targeted distribution of behaviour. In equilibrium, no transfers are made and thus it is a budget-balanced mechanism. We forgo the possibility of this by only being able to affect to give transfers to agents by observing the action they take, irrespective of their identity. As such, we would not be able to directly punish them for not taking a prescribed action as we cannot observe their identity.

III. DESIDERATA FOR MECHANISMS

In this work, we are interested in mechanisms that output a controller that *implements* some distribution over actions. Formally, a mechanism is a function $M : \mathcal{A} \to \mathcal{K}$ where $\mathcal{A} \subseteq \Delta(\mathcal{S})$. We aim to identify the largest possible domain for a mechanism while still respecting certain desiderata. Before describing the desired qualities of a mechanism, we define a property of a target distribution $p \in \Delta(\mathcal{S})$:

Definition 1: $p \in \Delta(\mathcal{S})$ is **asymptotically implementable** if there exists a controller k that asymptotically stabilizes to p.

Here asymptotic stability is defined as follows:

Definition 2: A controller $k_t(\cdot)$ **asymptotically stabilizes** to p if $\lim_{t\to\infty} \theta_t(s) = p(s) \ \forall s \in \mathcal{S}$, where θ_t is induced by k_t from any $\theta_0 \in$ relint$(\Delta(\mathcal{S}))$, and Lyapunov stable around θ_0.

Similarly we say p is **ϵ-asymptotically implementable** if there is $p' \in B_\epsilon(p)$ that is asymptotically implementable.

A. Stabilization

In the usual flavour of control design, we desire a controller that stabilizes to a given population distribution $p \in \mathcal{A}$. We define a desirable form of stability via the following:

Definition 3: Fix $\lambda > 0$. A controller k_t is **λ-exponentially stabilizing** to p if $\exists m > 0$ such that for all $\theta_0 \in \text{relint}(\Delta(\mathcal{S}))$ $\forall t \geq 0$, $\|\theta_t - p\| \leq m\|\theta_0 - p\|e^{-\lambda t}$. We say that M is λ-exponentially stabilizing if for any $p \in \mathcal{A}$, $M(p) = k_t$ is λ-exponentially stabilizing to o.

B. Budget-Balance

The second desired solution quality is that of *budget-balance*:

Definition 4: For a given population $\theta(s)$, k is a **budget-balanced controller** with respect to θ if $\sum_{s\in\mathcal{S}} \theta(s)k(\theta, s) = 0$. We say that M is a **budget-balanced mechanism** if for all $p \in \mathcal{A}$ and θ_t induced by $M(p) = k$ from any $\theta_0 \in \text{relint}(\Delta(\mathcal{S}))$, then k is a budget-balanced controller with respect to θ_t for all $t \geq 0$.
Intuitively, this condition specifies that the net transfers to all agents is zero. If instead it were positive, then it means that the social planner loses revenue, and if negative then it means that the social planner gains revenue. We find the latter non-ideal in our model as we consider a planner that cares only about agents' welfare and not their own revenue. Furthermore, we require this hold at every time $t \geq 0$. This is a stricter criteria that it integrating to zero over all time, as we do not want the planner to ever be in a deficit or surplus as the latter harms agents in the short run, and the former means the planner must take loans, which is not ideal. Nevertheless, we are able to find a desirable solution even with this strong requirement.

We define budget-balanced transfers as those that are *weighted* net-zero as opposed to the usual condition of unweighted net-zero. Specifically, it is weighted by the population distribution. When considering this as some percentage of a finite population, then this weighted sum is equivalent to transfer between agents when transfers must be the same within a strategy population:

Proposition 1: Let $\mathcal{I}_s$ be a non-empty, finite collection of agents with strategy $s \in \mathcal{S}$, and let T_s^i be the transfers associated with player $i \in \mathcal{I}_s$. If transfers are identical across strategy populations, denoted T_s, then the set of transfer $\{T_s^i : i \in \mathcal{I}_s, s \in \mathcal{S}\}$ is budget balanced if and only if $\{T_s : s \in \mathcal{S}\}$ is weighted budget balanced.

Proof: Let $\theta(s) = \frac{|\mathcal{I}_s|}{\sum_{s'\in\mathcal{S}} |\mathcal{I}_{s'}|}$. Observe the following: $\sum_{i\in\mathcal{I}_s} T_s^i = |\mathcal{I}_s|T_s = \theta(s)\left[\sum_{s'\in\mathcal{S}} |\mathcal{I}_{s'}|\right] T_s$. Then the budget-balancing condition is as follows: $0 = \sum_{s\in\mathcal{S}} \sum_{i\in\mathcal{I}_s} T_s^i = \sum_{s\in\mathcal{S}} \theta(s)T_s \left[\sum_{s'\in\mathcal{S}} |\mathcal{I}_{s'}|\right] = \sum_{s\in\mathcal{S}} \theta(s)T_s$. Thus transfers are budget balancing if and only if $\{T_s\}_{s\in\mathcal{S}}$ is weighted budget-balancing. ■

One can observe that the criteria of budget-balance can always be satisfied without affecting the replicator dynamics. We can do so by redistributing any profits or losses gained by the social planner equally across all agents, which does not change the dynamics. As such, any stability properties of the system are maintained under such a redistribution since the dynamics only depend on the relative payoffs and are thus unchanged. However when transfers are costly, such a redistribution could be costly. We study this in Section V.

IV. IMPLEMENTABILITY

In this section, we are concerned with what distributions are implementable:

Definition 5: $\mathcal{A} \subseteq \mathcal{S}$ is **λ-implementable** if there exists a mechanism M with domain $\mathcal{A}$ such that $p \in \mathcal{A}$ and M is λ-exponentially stabilizing and budget-balanced.
Further we say that the **time guarantee** for a mechanism M with respect to ϵ and p is a (t, ϵ) such that for all $t' > t$, $\theta_{t'} \in B_\epsilon(p)$ for θ_t induced by $M(p)$. Intuitively, a (t, ϵ) time guarantee means that we can find a controller that allows the the distribution of strategies to reach within ϵ of our target p in t units of time.

We also provide a notion of approximate implementability:

Definition 6: $\mathcal{A} \subseteq \mathcal{S}$ is **(λ, ϵ)-implementable** if there exists $\mathcal{B} \subseteq \mathcal{S}$ λ-implementable such that for all $p \in \mathcal{A}$, there exists $p' \in \mathcal{B} \cap B_\epsilon(p)$. $\mathcal{A} \subseteq \mathcal{S}$ is **λ-approximately implementable** if it is (λ, ϵ)-implementable for all $\epsilon > 0$.

We are interested in finding a maximal set $\mathcal{A}$ that is implementable for a given λ. Our main result is the following:

Theorem 1: Fix $\lambda > 0$. Then $\mathcal{A} = \text{relint}(\Delta(\mathcal{S}))$ is λ-implementable. Furthermore, there exists an implementing mechanism M such that for $p \in \mathcal{A}$ and all $\epsilon > 0$ sufficiently small, there exists a $\bar{t}_\epsilon(\lambda) = \mathcal{O}(\frac{1}{\lambda}(1 - \ln(\epsilon)))$ such that $(\bar{t}_\epsilon(\lambda), p)$ is a time guarantee for M with respect to ϵ and p.

All proofs can be found in the Appendix. We construct our controller to be one similar to feedback linearization, that is $k_t(s) = \frac{v_t(s) - f_s(\theta_t)}{g_s(\theta_t)}$. Hence we can say that the maximal domain for a mechanism M must contain $\text{relint}(\Delta(\mathcal{S}))$, as well as any pure strict Nash equilibria, and for approximate implementability the maximal domain is exactly $\Delta(\mathcal{S})$.

The construction of our mechanism relies on knowledge of the underlying utility function, hence any misspecification can cause asymptotic or exponential stability to no longer hold. However due to the design of our controller, when we know that preferences are quasilinear and we have bounded error in our specification of the utility function u, we nevertheless achieve approximate and exponential stability results by leveraging the robustness of exponential stability to bounded and proportional additive uncertainty[2]:

Theorem 2: Let $\hat{u}$ be a misspecified model for the utility function, such that the true unknown utility function u is within $\epsilon > 0$ of $\hat{u}$, i.e. $\|u - \hat{u}\|_\infty \leq \epsilon$.

1) For all $\epsilon' > \frac{\epsilon}{\lambda}$ and $\lambda > 0$, p is ϵ'-asymptotically implementable.
2) For all $\lambda > \epsilon$, if $p \in \text{relint}(\Delta(\mathcal{S}))$ then p is $(\lambda - \epsilon)$-implementable.

Proof: Let $\delta : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$ be such that $u = \hat{u} + \delta$. Though we do not know u, we know that that $\|\delta\| \leq \epsilon$ by assumption. We can

[2]See Proposition 5 in the Appendix for a more general characterization.

rewrite our dynamical system as follows: $\dot{\theta}_t(s) = \theta_t(s)[\hat{u}_t(s)+\delta_t(s)+k_t(s)-\mathbb{E}_{s'\sim\theta_t}[\hat{u}_t(s')+k_t(s')]-\bar{\delta}_t] = \theta_t(s)\left[\hat{u}_t(s)+k_t(s)-\mathbb{E}_{s'\sim\theta_t}\left[\hat{u}_t(s')+k_t(s')\right]\right] + N_t(s) = y(\theta_t,k_t,s) + N_t(s)$, where $\delta_t(s) = \mathbb{E}_{s'\sim\theta_t}\left[\delta(s,s')\right]$ and $\bar{\delta}_t = \mathbb{E}_{s\sim\theta_t}\left[\delta_t(s)\right]$, $y(\theta_t,k_t,s) = \theta_t(s)\left[\hat{u}_t(s)+k_t(s)-\mathbb{E}_{s'\sim\theta_t}\left[\hat{u}_t(s')+k_t(s')\right]\right]$, and $N_t(s)=\theta_t(s)(\delta_t(s)-\bar{\delta}_t)$. Observe the following bound on $N_t(s)$: $\|N_t(s)\| = \|(I_s-\theta_t)^\top D\theta_t\| \leq 1\cdot\epsilon\cdot\|\theta_t\|$ where $I_s\in\mathbb{R}^{|\mathcal{S}|}$ is a basis vector such that it has 1 as its s-th entry and zero otherwise, and D is the matrix representation of δ.

Observe that our controller, as it is feedback linearizing, induces the following closed-loop dynamics: $\dot{\theta}_t(s) = -\lambda\theta_t(s)+N_t(s)$. Note that the there being finitely many strategies makes it clear that the expectation, and thus the controller, is smooth in the state. Given this, using $V(x)=\frac{1}{2}x^\top x$ as our Lyapunov function, which clearly satisfy the conditions of the converse Lyapunov theorem [14], gives us exponential stability of $\dot{\theta}_t(s) = -\lambda\theta_t(s)$. Note that $c_1=c_2=\frac{1}{2}$, $c_3=\lambda$, and $c_4=1$.

Note that since the unknown error term δ is bounded, enters linearly into the known dynamics, and our known dynamics can be stabilized to an arbitrarily small ball, we can leverage Proposition 5 (see Appendix). In particular, choose $p'\in\text{relint}(\Delta(\mathcal{S}))$ such that $\|p-p'\|<\epsilon'-\frac{\epsilon}{\lambda}$. Consider the controller with respect to p', and since our technique of feedback linearization induces the above exponentially stable linear system, we have that $\kappa=\frac{\epsilon c_4}{c_3}\sqrt{\frac{c_2}{c_1}}=\frac{\epsilon}{\lambda}$. Thus $\theta_t\to B_\epsilon(p')$ and thus is within $\theta_t\to B_{\epsilon'}(p)$.

To see the second claim, observe that by assumption, $\epsilon<\lambda=\frac{c_3}{c_4}$, hence by Proposition 5, we have that the rate of stabilization is $\lambda-\epsilon$ and that the overshoot constant $m=\sqrt{\frac{c_2}{c_1}}=1$. ■

This result shows that for a fixed λ, we can always achieve approximate asymptotic stability wherein we reach ϵ-ball around our target. Thus our worst case error due to misspecification is linear in the misspecification error. Furthermore, if we choose λ sufficiently large, then we are still able to exponentially stabilize at a slower rate as if our model were correctly specified. In cases where we have constraints or transfer cost considerations, we may be unable to choose a large enough λ, and thus the former result gives us a necessary approximation guarantee.

V. Tradeoffs in Implementation

In this section we analyze the tradeoffs induced by stabilization when transfers are costly. This is motivated by the idea that large transfers, which represent payments to or from agents, may not be favourable as they impose a large monetary burden on agents or the planner. Consider the task of stabilizing to an ϵ-ball around a distribution $p\in\Delta(\mathcal{S})$. Recall that by applying our controller, we can stabilize to a distribution in $\mathcal{O}(\frac{1}{\lambda})$ time for $\lambda>0$. In that time, we can bound the cumulative value of a quadratic cost as follows:

Proposition 2: Fix $\lambda>0$, $p\in\Delta(\mathcal{S})$, and $\epsilon>0$. Then the cost, as defined by $\texttt{cost}:\mathcal{K}\to\mathbb{R}_+$, of our feedback controller can be bounded as follows: $\texttt{cost}(k_t) = \int_0^{\bar{t}_\epsilon(\lambda)}\sum_{s\in\mathcal{S}}\|k_\tau(s)\|^2\cdot d\tau \leq \mathcal{O}(\lambda)$.

This emphasizes that in the worst case, the cost of our controller is increasing linearly in λ, whereas the time to (approximate) stabilization is decreasing in λ at rate proportional to $\frac{1}{\lambda}$. How to balance these tradeoffs is application-specific, and in the next section, we consider one approach to do so.

A. Optimal Control of Population Dynamics

To capture the tradeoff between stabilization and control costs, we use the following functional that mimics stabilization problems commonly seen in control applications: for $k\in\mathcal{K}$ a controller, p some target distribution, and $\rho>0$, define $J[k] = \int_0^\infty\sum_{s\in\mathcal{S}}\left(\|\theta_\tau(s)-p(s)\|^2+\rho\|k_\tau(s)\|^2\right)\cdot d\tau$. Hence our optimization problem is $\inf_{k(s):[0,\infty)\to\mathbb{R}} J[k]$ subject to $\sum_{s\in\mathcal{S}}\theta_t(s)k_t(s) = 0\ \forall t\geq 0$, and $\dot{\theta}_t(s) = f_s(\theta_t)+g_s(\theta_t)k_t(s)\ \forall s\in\mathcal{S}$. Note that the solution has finite value only if k_t asymptotically stabilizes the system to p.

Though not optimal, the use of feedback linearization can help to provide sub-optimality guarantees. If no structure is imposed on p, then in general control costs in the long run may be infinite. For example, this may occur if we're trying to stabilize to a non-equilibrium point since a constant control input would have to be applied for all time. If p were locally exponentially stable, then we have stronger results. An example of a population game with the above property is the following:

Example 1: Consider the coordination game given by strategies $\{A,B\}$ and the symmetric utility function: $u(x,y)=2$ if $x=y=A$, 1 if $x=y=B$, and 0 otherwise. The two pure strategy NE are (A,A) and (B,B), and the mixed strategy NE is $(\frac{1}{3},\frac{2}{3})$. Letting $x_t=\theta_t(A)$, we have the following replicator dynamics: $\dot{x}=-3x^2(x-\frac{1}{3})(x-1)$. It is straightforward to find that the system is locally exponentially stable around $[0,\frac{1}{3})$.

Theorem 3: Let p be a locally stable Nash s.t $B_\epsilon(p)\cap\Delta(\mathcal{S})$ is a subset of its region of attraction (ROA) with exponential rate $\gamma>0$. Let $\lambda>0$. Fix $\theta_0\in\text{relint}\Delta(\mathcal{S})$. Then we have that $\exists\delta(p,x_0)\in(0,1)$ and $c,\eta>0$[3] such that $\min_{k(s):[0,\infty)\to\mathbb{R}} J[k]$ is upper bounded by

$$\underbrace{\mathcal{O}(\frac{\eta^{\frac{\gamma}{\lambda}}}{\gamma})}_{\text{State cost in ROA}} + \underbrace{\mathcal{O}(\frac{1}{\lambda})}_{\text{State cost outside ROA}} + \rho\cdot\underbrace{\mathcal{O}(\lambda)}_{\text{Control cost}}$$

This result can be used by planners as a means of informing what values of λ and p are reasonable by giving an upper bound on the overall cost as a function of these variables. This is in contrast to repeatedly running simulations for different combinations of these variables, thus providing a useful tool for quick analysis. For example, in the setting of traffic tolling, the trade-off between fast stabilization to socially optimal traffic flows and the size of transfers can be analyzed more easily.

[3]Constant factors and lower order terms are omitted in the expression for clarity.

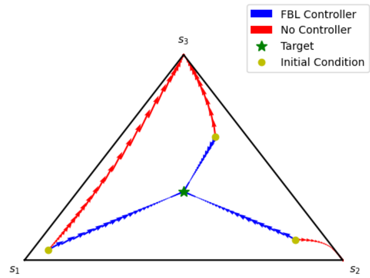

Fig. 1: The initial conditions are $(\theta_0(s_i))_{i=1}^3 \in \{(0.9, 0.05, 0.05), (0.1, 0.8, 0.1), (0.1, 0.3, 0.6)\}$.

VI. Application: Non-Atomic Routing Game

In this section, we study our proposed controller in the setting of a *non-atomic routing game*. Consider a continuum of agents with total demand 1, and the same origin and destination. Their strategy $s \in S$ is a choice of route, with associated cost given by the function $l : [0, 1] \to \mathbb{R}$ where $l_s(\theta)$ is the time taken to reach the destination (from the origin) when the *flow* of traffic is distributed across routes according to θ. This is no longer a normal-form game, however can now be expressed as a population games by considering $\hat{F}_s(\theta) = -l_s(\theta)$. Though our results are written in for two player games, it straightforwardly applies to population games by considering $\dot{\theta}_t(s) = \theta_t(s)\left[F_s(\theta_t) - \theta_t \cdot F(\theta_t)\right]$ where $F_s(\theta) = \hat{F}_s(\theta) + k(s, \theta)$.

Consider the following example[4]. The set of strategies will be denoted $\mathcal{S} = \{s_i\}_{i=1}^3$, and the origin and destination are denoted O and D respectively. $\mathcal{S}$ represent a set of parallel routes between O and D. The corresponding costs will be given by $l_{s_i}(\theta) = -(i \cdot \theta(s_i)^2 + i)$. In Figure 1, we simulate the replicator dynamics for various initial conditions under our controller as well as with no controller, with the goal of stabilizing to $\theta^* \equiv \frac{1}{3}$. Observe that under our controller, the dynamics of the system approach the target in a straight line. This is a consequence of our system being feedback linearized, thus allowing for such a linear change in state.

VII. Discussion

In this section, we provide a discussion on different aspects of our model, potential limitations of our approach, and future directions of exploration.

a) Payoff Information: Our results heavily rely on the designer having complete information of payoffs, that is the function u. Though ϵ misspecified utility function, as described in Theorem 2, has approximation guarantees linear in ϵ, this may still be impractical in reality when there is little data from which utility functions can be estimated from. Hence the application of our controller works best in environments such as traffic routing, where utility functions can be readily estimated due to a surplus of data.

[4]This examples is similar to that of [10], except we have fewer routes so that we can plot the results on the simplex.

b) Dynamics Model: Our analysis uniquely builds on the replicator dynamics, which is one of many evolutionary dynamics models [1]. There is evidence to support that in some settings, the replicator dynamics model can explain human behvaiour [15], [16], we are limited in our ability to model more general environments such as those in payoff dynamics models [8].

c) Practical Implementation: Our model is formulated in continuous time, hence our feedback controller may suffer from practical issues such as delays when implemented discretely. However, the degree to which this is an issue depends on the application. In our motivating example of traffic tolling, the time difference between successive periods represents the time between two different journeys. It is reasonable to assume that this is sufficiently long. In general, we model behavioral change and thus the rate at which this is believed to occur informs whether a discrete time implementation poses issues.

Appendix

A. Proof of Theorem 1

First assume that $k_t(\cdot)$ is budget-balanced. Applying the tools of feedback linearization, we choose $k_t(s) = \frac{v_t(s) - f_s(\theta_t)}{g_s(\theta_t)}$. This is well-defined through the following:

Proposition 3: Consider $\mathcal{X} = \mathbb{R}$, $\dot{x} = u$, $x_0 \in (0, 1)$ and $x^* \in (0, 1)$. Then $\forall \lambda > 0$, $\exists \delta \in (0, 1)$ and $u : [0, \infty) \to \mathbb{R}$ such that $x_0, x^*, x_t \in [\delta, 1 - \delta]$ $\forall t \geq 0$.

Proof: Clearly $\exists \delta > 0$ s.t $x_0, x^* \in [\delta, 1 - \delta]$. Consider the error dynamics $q = x - x^*$ and controller $u = -\lambda e$. Hence $x_t = e^{-\lambda t} x_0 + (1 - e^{-\lambda t}) x^*$, implying our result. ■

As the control system stays in some compact subset of the relative interior of the simplex, k_t is well-defined always since its application to our model results in the same dynamics $\dot{x} = u$. Now observe that $\sum_s \theta_t(s) k_t(s) = 0 \iff \sum_s v_t(s) = 0$. Note that since feedback linearization decouples a system into many one-dimensional linear systems, if each system is exponentially stabilizing with the same rate, then so is the whole system. That is, of $q_t(s) \to 0$ with exponential rate λ (and $M = 1$) for all $s \in \mathcal{S}$, then $q_t \to 0$ with the same rate (with $M = 1$). Consider $p \in \text{relint}(\Delta(\mathcal{S}))$ and a feedback linearizing controller $k_t(s)$ with respect to some $v_t(s)$. Assuming that the overall controller is budget-balancing, the linearized dynamics are $\dot{\theta}_t(s) = v_t(s)$. Choose $v_t(s) = -q_t(s)$ for $q_t(s) = \theta_t(s) - p(s)$. Furthermore, the controller satisfies budget-balance: $\sum_{s \in \mathcal{S}} v_t(s) = -\sum_{s \in \mathcal{S}} (\theta_t(s) - p(s)) = -\sum_{s \in \mathcal{S}} \theta_t(s) + \sum_{s \in \mathcal{S}} p(s) = 0$. Note that we can use any controller of the form $v_t(s) = -\lambda q_t(s)$ for $\lambda > 0$, thus we can choose the rate of exponential stabilization. Note the following time-guarantee result:

Proposition 4: For all $\epsilon > 0$ and $p \in (\Delta(\mathcal{S}))$, there is a time $\bar{t} \in [0, \infty)$ s.t $\theta_t \in B_\epsilon(p)$ $\forall t > \bar{t}_\epsilon(\lambda) = \frac{1}{2\lambda} \ln \frac{4\|q_0\|^2}{\epsilon^2}$ for some controller k_t.

Proof: Choose $p' \in B_{\frac{\epsilon}{2}}(p) \cap \text{relint}(\Delta(\mathcal{S}))$. Recall that we can choose the rate of exponential stability for any $\lambda > 0$ when implementing a feedback linearizing controller with respect to p'. Note that $\dot{q}_t = -\lambda q_t$ implies that $\|e_t\|^2 = \|q_0\|^2 e^{-\lambda 2t}$. To ensure $\|q_t\| \leq \epsilon$, we can upper bound the term on the right-hand side by $\delta > 0$ first: $\|q_0\|^2 e^{-\lambda 2t} \leq \delta^2 \iff \frac{1}{2\lambda}\ln\frac{\|q_0\|^2}{\delta^2} \leq t$. Set $\bar{t}_\epsilon(\lambda) := \frac{1}{2\lambda}\ln\frac{\|q_0\|^2}{\delta^2}$ to ensure $\|q_t\| \leq \delta$. Choose $\delta := \frac{\epsilon}{2}$ to find that $\|q_t\| \leq \frac{\epsilon}{2}$. Let $\tilde{q}_t = \theta_t - p$, hence $\|\tilde{q}_t\| \leq \|\theta_t - p'\| + \|p' - p\| = \epsilon$. ■

B. Proof of Proposition 2

We aim to bound the following: $\texttt{cost}(k_t) = \int_0^{\bar{t}_\epsilon(\lambda)} \sum_{s\in\mathcal{S}} \|k_t(s)\|^2 \cdot dt$ First note that $\|k_t(s)\| \leq (\frac{\lambda\|q_t(s)\|}{\delta} + c)^2$, which follows from there existing $\delta > 0$ such that $\theta_t(s) \in [\delta, 1-\delta]$ (see proof of Theorem 1) and $\exists c > 0$ such that $\|u_t(s) - \bar{u}_t\| \leq c$ (see proof of Theorem 3). First note that $\texttt{cost}(k_t) \leq \int_0^{\bar{t}_\epsilon(\lambda)} \sum_{s\in\mathcal{S}} \left[\lambda^2\frac{\|q_t(s)\|^2}{\delta} + 2c\frac{\lambda\|q_t(s)\|}{\delta} + c^2\right] \cdot dt$. We bound each term individually. First, $\int_0^{\bar{t}_\epsilon(\lambda)} \sum_{s\in\mathcal{S}} \lambda^2\frac{\|q_t(s)\|^2}{\delta} \cdot dt \leq \frac{\lambda^2}{\delta}\int_0^{\bar{t}_\epsilon(\lambda)} \|q_0\|^2 e^{-2\lambda t} \cdot dt = \mathcal{O}(\lambda)$, $\int_0^{\bar{t}_\epsilon(\lambda)} \sum_{s\in\mathcal{S}} c^2 = |\mathcal{S}|c^2\bar{t}_\epsilon(\lambda) = \mathcal{O}(\frac{1}{\lambda})$. Secondly, $\int_0^{\bar{t}_\epsilon(\lambda)} \sum_{s\in\mathcal{S}} 2c\frac{\lambda}{\delta}\|q_t(s)\| \cdot dt = 2c\frac{\lambda}{\delta}\sum_{s\in\mathcal{S}}\int_0^{\bar{t}_\epsilon(\lambda)} \|q_t(s)\| \cdot dt \leq 2c\frac{\lambda}{\delta}\sum_{s\in\mathcal{S}} \|q_0(s)\| \int_0^{\bar{t}_\epsilon(\lambda)} e^{-\lambda t} \cdot dt = \mathcal{O}(1)$. Putting this all together, we get that $\texttt{cost}(k_t) \leq \mathcal{O}(\lambda)$.

C. Proof of Theorem 3

Let $\tilde{k}_t(s)$ be the controller from the previous proposition, and $p' \in \text{relint}(\Delta(\mathcal{S}))$ be the target distribution. Define the following controller: $k_t(s) = \tilde{k}_t(s)$ if $t \leq \bar{t}(\lambda)$ and 0 otherwise. We first upper bound the optimal value by the value of the feedback linearizing controller: $\int_0^{\bar{t}(\lambda)} \sum_{s\in\mathcal{S}} \left(\|\theta_\tau(s) - p(s)\|^2 + \rho\|k_\tau(s)\|^2\right) \cdot d\tau + \int_{\bar{t}(\lambda)}^{\infty} \sum_{s\in\mathcal{S}} \left(\|\theta_\tau(s) - p(s)\|^2 + \rho\|k_\tau(s)\|^2\right) \cdot d\tau$. Consider the second component: $\int_{\bar{t}(\lambda)}^{\infty} \sum_{s\in\mathcal{S}} \left(\|\theta_\tau(s) - p(s)\|^2 + \rho\|k_\tau(s)\|^2\right) \cdot d\tau \leq \int_{\bar{t}(\lambda)}^{\infty} \|q_{\bar{t}(\lambda)}\|^2 e^{-2\gamma\tau} \cdot d\tau \leq \epsilon^2\left[-\frac{e^{-2\gamma\tau}}{2\gamma}\right]_{\bar{t}(\lambda)}^{\infty} = \frac{\epsilon^2}{2\gamma}\left[\frac{4\|q_0\|^2}{\epsilon^2}\right]^{-\frac{\gamma}{\lambda}} = \mathcal{O}(\frac{\eta^{\frac{\gamma}{\lambda}}}{\gamma})$, where we set $\eta = \frac{\epsilon^2}{4\|q_0\|^2}$ and leverage the fact that $\theta_\tau \in B_\epsilon(p)$ for all $\tau \geq \bar{t}(\lambda)$ by design of the controller. Now we consider the first portion of the bound, specifically the state cost: $\int_0^{\bar{t}(\lambda)} \sum_{s\in\mathcal{S}} \|\theta_\tau(s) - p(s)\|^2 \cdot d\tau \leq \int_0^{\bar{t}(\lambda)} \left(\|\theta_\tau - p'\| + \|p' - p\|\right)^2 \cdot d\tau \leq \int_0^{\bar{t}(\lambda)} \left(\|\theta_\tau - p'\| + \frac{\epsilon}{2}\right)^2 \cdot d\tau \leq \int_0^{\bar{t}(\lambda)} \left(\|\tilde{q}_\tau\| + \frac{\epsilon}{2}\right)^2 \cdot d\tau \leq \int_0^{\bar{t}(\lambda)} \left(\|\tilde{q}_0\|e^{-\lambda\tau} + \frac{\epsilon}{2}\right)^2 \cdot d\tau$. This final term is $\mathcal{O}(\frac{1}{\lambda})$, and where $\tilde{q}_\tau = \theta_\tau - \tilde{p}$ and we use the property that under the controller we exponentially stabilize to p' with rate λ. To consider the control cost, recall the form of the controller $k_t(s)$ for $s \in \mathcal{S}$: $k_t(s) = -\lambda\frac{\tilde{q}_t(s)}{\theta_t(s)} - u_t(s) + \bar{u}_t$ implies that $\|k_t(s)\| \leq \lambda\frac{\|\tilde{q}_t(s)\|}{\delta} + \|u_t(s) - \bar{u}_t\|$, where $\delta = \min_{s\in\mathcal{S}}\min\{\delta_s, 1-\delta_s\}$ and δ_s is as per the previous lemma (with respect to p' and θ_0). This follows since $|\theta_t(s)| = \theta_t(s) > \delta_s \geq \delta$. To bound the norm of the relative fitness, we have the following: $\|u_t(s) - \bar{u}_t\| \leq c$ where $c = \max_{s,s'\in\mathcal{S}}|u(s,s')|$. Thus we can bound the control cost as follows: $\rho\int_0^{\bar{t}(\lambda)} \sum_{s\in\mathcal{S}} \|k_\tau(s)\|^2 \cdot d\tau \leq \rho\mathcal{O}(\lambda)$, where we use the result from Proposition 2. Together these bounds give $\mathcal{O}(\frac{\eta^{\frac{\gamma}{\lambda}}}{\gamma}) + \mathcal{O}(\frac{1}{\lambda}) + \rho\mathcal{O}(\lambda)$.

D. Robustness of Exponential Stability

Note the following result [14]: $\dot{x} = f(x)$ is exponentially stable if and only if there exists $c_1, c_2, c_3, c_4 > 0$ such that $c_1\|x\|^2 \leq V(x) \leq c_2\|x\|^2$, $L_f V(x) \leq -c_3\|x\|^2$, and $\left\|\frac{\partial V}{\partial x}(x)\right\| \leq c_4\|x\|$, where $L_f V(x)$ is the Lie derivative with respect to f, i.e $L_f V(x) = \dot{V}(x) = \frac{\partial V}{\partial x}(x) \cdot f(x)$.

Proposition 5: If $\dot{x} = f(x)$ is globally exponentially stable with respect to the origin on $B_1(0)$, then for $\dot{x} = f(x) + q(x)$ such that $\|q(x)\| \leq \epsilon$ for all $x \in \mathcal{X}$, x_t asymptotically stabilizes to $B_\kappa(0)$ where $\kappa = \frac{\epsilon c_4}{c_3}\sqrt{\frac{c_2}{c_1}}$. If $\|q(x)\| \leq \epsilon \cdot \|x\|$, where $\epsilon < \frac{c_3}{c_4}$, then x_t exponentially stabilizes to the origin at rate $\frac{\lambda}{2c_2}$.

Proof: Lemma 5.2 in [17] with $\delta = \epsilon$, $r = 1$ and $\theta < 1$ sufficiently large gives us that $B_\kappa(0)$ is asymptotically stabilized to. If $\|q(x)\| \leq \epsilon \cdot \|x\|$, then $\dot{V}(x) \leq (\epsilon \cdot c_4 - c_3)\|x\|^2$. Let $\lambda = \epsilon \cdot c_4 - c_3 < 0$. Then $\dot{V}(x) \leq \frac{\lambda}{c_2}V(x)$, hence $V(x_t) \leq e^{\frac{\lambda}{c_2}t}V(x_0)$. Using the upper annd lower bounds on V, we can see that $\|x\| \leq \sqrt{\frac{c_2}{c_1}}e^{\frac{\lambda}{2c_2}t}\|x_0\|$. ■

References

[1] W. H. Sandholm, *Population games and evolutionary dynamics*. MIT press, 2010.

[2] ——, "Pigouvian pricing and stochastic evolutionary implementation," *Journal of Economic Theory*, vol. 132, no. 1, pp. 367–382, 2007.

[3] ——, "Negative externalities and evolutionary implementation," *The Review of Economic Studies*, vol. 72, no. 3, pp. 885–915, 2005.

[4] R. Lahkar and S. Mukherjee, "Evolutionary implementation in aggregative games," *Mathematical Social Sciences*, vol. 109, pp. 137–151, 2021.

[5] L. Zino, M. Ye, A. Rizzo, and G. C. Calafiore, "On adaptive-gain control of replicator dynamics in population games," in *2023 62nd IEEE Conference on Decision and Control (CDC)*. IEEE, Dec. 2023, p. 485–490. [Online]. Available: http://dx.doi.org/10.1109/CDC49753.2023.10383983

[6] L. Zino, M. Ye, G. C. Calafiore, and A. Rizzo, "Equilibrium selection in replicator equations using adaptive-gain control," *IEEE Transactions on Automatic Control*, 2025.

[7] V. Raju and P. S. Krishnaprasad, "Lie algebra structure of fitness and replicator control," 2020. [Online]. Available: https://arxiv.org/abs/2005.09792

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

[9] L. J. Ratliff and T. Fiez, "Adaptive incentive design," *CoRR*, vol. abs/1806.05749, 2018. [Online]. Available: http://arxiv.org/abs/1806.05749

[10] C. Maheshwari, K. Kulkarni, M. Wu, and S. Sastry, "Dynamic tolling for inducing socially optimal traffic loads," 2021.

[11] ——, "Inducing social optimality in games via adaptive incentive design," 2022.

[12] G. Como and R. Maggistro, "Distributed dynamic pricing of multiscale transportation networks," *IEEE Transactions on Automatic Control*, vol. 67, no. 4, pp. 1625–1638, 2021.

[13] R. Cressman and Y. Tao, "The replicator equation and other game dynamics," *Proceedings of the National Academy of Sciences*, vol. 111, pp. 10 810 – 10 817, 2014.

[14] S. Sastry, *Nonlinear systems: analysis, stability, and control*. Springer Science & Business Media, 2013, vol. 10.

[15] M. Hoffman, S. Suetens, U. Gneezy, and M. A. Nowak, "An experimental investigation of evolutionary dynamics in the rock-paper-scissors game," *Scientific reports*, vol. 5, no. 1, p. 8817, 2015.

[16] K. D. Choudhury and T. Aydinyan, "Stochastic replicator dynamics: A theoretical analysis and an experimental assessment," *Games and Economic Behavior*, vol. 142, pp. 851–865, 2023.

[17] H. K. Khalil and J. W. Grizzle, *Nonlinear systems*. Prentice hall Upper Saddle River, NJ, 2002, vol. 3.