Passivity-based Analysis and Design for Population Dynamics with Conformity Biases

Shunya Yamashita, Student Member, IEEE, Kodai Irifune, Takeshi Hatanaka, Senior Member, IEEE, Yasuaki Wasa, Member, IEEE, Kenji Hirata, Member, IEEE and Kenko Uchida, Member, IEEE

Abstract—This paper addresses mechanisms for boundedly rational decision makers in discrete choice problem. First, we introduce two mathematical models of population dynamics with conformity biases. We next analyze the models in terms of δ-passivity, and show that the conformity biases work to break passivity of decision makers. Based on the passivity perspective, we propose mechanisms so as to induce decision makers to a desired population state. Furthermore, we analyze a convergence property of designed mechanisms, and present parameter conditions to guarantee stable inducements.

Index Terms—Population dynamics, Bounded rationality, Conformity bias, Behavior modification, Passivity

I. INTRODUCTION

Human behavior is a critically important factor in design and analysis of large-scale social systems like transportation network [1] and energy management systems [2]–[4]. As seen in these publications, a typical approach to address such systems involving humans is to model social decisions/dynamics assuming human rationality, against the background of expected utility theory [5]. Meanwhile, behavioral economics has pointed out that human rationality may be bounded due to information and/or cognitive constraints, as typified by prospect theory [6] or dual-process theory [7]. It is highly uncertain that any system designed under the assumption of human rationality will work when their rationality is bounded. Motivated by the issue, boundedly rational human/social models have begun to be investigated in the field of systems and control [8], [9]. Among various types of bounded rationality, in this paper, we focus on so-called conformity bias [10], tendency to follow the majority, that is observed in the scenes of parking location choice [11] and evacuation decision [12].

One of the decision making issues is discrete choice problem which decision makers choose a strategy from finite number of options [13]. Indeed, some significant issues in society are categorized as discrete choice: route selection in transportation networks [14], choice of energies [15], and water distribution [16], for example. Meanwhile, when we deal with systems including large scale of population, it should be required to consider decision models as a population [17]. Discrete choice behavior by large populations is well-addressed in evolutionary game theory, in which various types of population dynamics have been presented, e.g., logit dynamics [18], [19], Smith dynamics [20] and pairwise comparison dynamics [21]. Meanwhile, some recent publications have pointed out relations between population dynamics and passivity [22]–[24]. The researches in [22], [23] analyze passivity for generally formulated population dynamics, and [24] particularly focuses on the decision model in water distribution system. However, the models dealt with in the above literature do not explicitly consider influences of biases. Although several publications [25], [26] deal with bounded rationality in population dynamics, they address the models without focusing on passivity.

In social systems, it is sometimes preferred to induce humans to a desired social state. From the viewpoint of inducement for populations, behavior modification mechanism might be a key concept to achieve a desired behavior. This includes incentive [27] which is an economic approach, or nudge [28] which is an informational one. These kinds of mechanisms have recently attracted the attention of control community [29], [30], and more specifically, the ones for population dynamics have also been studied [31], [32]. Whereas, the publication which explicitly consider bounded rationality is rare. The authors in [25] propose a nudging mechanism for biased population dynamics, and analyze stability by using singular perturbation.

In this paper, we address mechanisms for discrete choice problem under bounded rationality. First, we introduce a well-used population dynamics model, called the logit dynamics, and its δ-passivity [23]. Next, we extend the decision making model to the one with conformity biases. We show two types of the bias models respectively addressed in [33] and [25]. We then analyze the biased logit dynamics in terms of δ-passivity, and reveal the impacts of conformity biases. Inspired by passivity-based control methods, we propose mechanisms so as to achieve desired social behavior for two types of biased models, respectively. Furthermore, we analyze their convergence, and show parameter conditions in order to stabilize the proposed mechanisms.

We summarize the contributions of this paper as follows:

- the impact by conformity biases to population dynamics is shown in terms of passivity, and
- passivity-based mechanisms are presented, which stably
achieve desired social state.

Parts of the contents in this paper are similar to the previous work in [34]. Meanwhile, variation of bias models and exact analysis of mechanisms are the incremental contributions added anew in this paper.

II. MODEL DESCRIPTION

A. Preliminaries: δ-Passivity

This subsection introduces δ-pas- sivity [23] which is a similar concept to passive systems [35]. Suppose a system Σ represented by the state space model

\[ \dot{x} = f(x, u), \]

where \( x \in \mathbb{R}^n \) is the state, \( u \in \mathbb{R}^m \) is the input, and \( f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \) is a function. Then, Σ is called δ-passive from \( \dot{x} \) if there exists a positive semi-definite function \( S : \mathbb{R}^n \times \mathbb{R}^m \to [0, \infty) \) and a scalar \( \rho \geq 0 \) such that

\[ S(u(t), x(t)) - S(u(0), x(0)) \leq \int_0^t \dot{x}^\top(t) \dot{u}(t) - \rho |\dot{u}(t)|^2 dt \]

for all input \( u \), all initial state \( x(0) \) and all \( t \geq 0 \). The positive semi-definite function \( S \) is particularly called a storage function. In addition, Σ is called δ-input-strictly-passive if the above inequality is satisfied with some positive scalar \( \rho > 0 \).

As widely known, if \( S \) is differentiable, we can replace the above inequality with

\[ \dot{S}(u(t), x(t)) \leq \dot{x}^\top(t) \dot{u}(t) - \rho |\dot{u}(t)|^2. \]

In the same way as passivity shortage [36], we define δ-output-passivity-shortage. The system Σ is called δ-output-passivity-short from \( \dot{x} \) to \( \dot{u} \) if there exists a positive semi-definite function \( S : \mathbb{R}^n \times \mathbb{R}^m \to [0, \infty), \) and a scalar \( \gamma \geq 0 \) such that

\[ S(u(t), x(t)) - S(u(0), x(0)) \leq \int_0^t \dot{x}^\top(t) \dot{u}(t) + \gamma |\dot{u}(t)|^2 dt \]

for all input \( u \), all initial state \( x(0) \) and all \( t \geq 0 \). We call the value \( \gamma \) as an impact coefficient. If \( S \) is differentiable, the above inequality is equivalent to

\[ \dot{S}(u(t), x(t)) \leq \dot{x}^\top(t) \dot{u}(t) + \gamma |\dot{u}(t)|^2. \]

Feedback system composed of strictly passive component and passivity-short one is related to the Lyapunov stability [35], [37]. Consider a δ-input-strictly-passive system whose input is \( u_1 \) and state is \( x_1 \), which satisfies \( S_1 \leq \dot{x}_1^\top \dot{u}_1 - \rho |\dot{u}_1|^2 \) for a differentiable storage function \( S_1 \) and \( \rho > 0 \). We also suppose a δ-output-passivity-short system whose input is \( u_2 \) and state is \( x_2 \), which satisfies \( S_2 \leq \dot{x}_2^\top \dot{u}_2 + \gamma |\dot{x}_2|^2 \) for a differentiable storage function \( S_2 \) and \( \gamma > 0 \). Then, the feedback interconnection under \( u_1 = x_2 \) and \( u_2 = -x_1 \) provides

\[ \dot{S}_1 + \dot{S}_2 \leq (\gamma - \rho) |\dot{x}_1|^2. \]

If \( \gamma \leq \rho \) holds and \( S_1 + S_2 \) is radially unbounded, the above inequality suggests stability of the feedback system in the sense of Lyapunov [35]. In other words, δ-input-strictly-passive systems can stabilize δ-output-passivity-short ones by negative feedback.

B. Dynamic Decision Making Model under Rationality

This subsection introduces the logit dynamics [19], [23] which is a well-used dynamical discrete choice model.

Consider the situation that decision makers choose a strategy from \( n \) available strategies. We denote the strategy set as \( A := \{1, 2, \ldots, n\} \). Suppose that the population of the decision makers can be represented as a continuum value. We now define \( S := \{x \in [0, 1]^n \mid 1_n^\top x = 1\} \) as the strategy choice distribution set, where \( 1_n \) describes \( n \) dimensional vector whose elements are all 1. We denote the relative interior of the set \( S \) as int(\( S \)). The population state \( \pi \in S \) implies the distribution of strategy choice. Specifically, the \( k \)-th element of \( \pi \), denoted by \( \pi_k \in [0, 1] \), implies a fraction of the decision makers selecting strategy \( k \in A \).

Let us define a cost vector \( \tau \in \mathbb{R}^n \) of which \( k \)-th element \( \tau_k \in \mathbb{R} \) corresponds to the cost for choosing strategy \( k \in A \).

Then, the logit dynamics [19], [23] is represented as

\[ \dot{\pi} = \eta (Q(\tau) - \pi), \quad \pi(0) \in \text{int}(S), \]

where \( \eta \) is a positive constant which indicates the update rate. The function \( Q : \mathbb{R}^n \to \text{int}(S) \) corresponds to the steady state of (1) and follows

\[ Q_k(\tau) = \frac{\exp(-\beta \tau_k)}{\sum_{l \in A} \exp(-\beta l_k)}, \]

where \( Q_k : \mathbb{R}^n \to (0, 1) \) implies \( k \)-th element of \( Q \), and \( \beta > 0 \) is a constant. The system (1) is known to guarantee \( \pi(t) \in \text{int}(S) \) for any time \( t \geq 0 \). Remarking that \( Q_k(\tau) > Q_1(\tau) \) for any pair \( k \neq i \) satisfying \( \tau_k < \tau_i \), the logit dynamics (1) can be interpreted as an approximated model of best response, i.e., rational decision making [18].

The logit dynamics (1) is also known to satisfy δ-passivity as below.

\textbf{Lemma 1 ([23])} The logit dynamics (1) is δ-passive from \( x \) to \( \dot{x} \) for the storage function

\[ S(\tau, \pi) := \eta \left( \pi^\top \tau + \frac{1}{\beta} \sum_{l \in A} \pi_l \log \pi_l \right) - \eta \min_{\omega \in S} \left( \omega^\top \tau + \frac{1}{\beta} \sum_{l \in A} \omega_l \log \omega_l \right), \]

i.e., \( \dot{S} \leq -\dot{x}^\top \dot{x} \) holds.

Remark that the storage function \( S(\tau, \pi) \) has a kind of radial unboundedness. See Appendix I for the details.

C. Bias Models

The logit model (1) corresponds to an approximation of rational strategy choice. Meanwhile, in behavioral economics field, it has been pointed out that human’s decision process is sometimes biased relying on population state. This subsection introduces two models of conformity bias. In particular, we represent these bias models as the ones to affect the cost \( \tau \). Hereafter, we refer to \( \tau \) as a biased cost. Alternatively, the actual cost is denoted by \( T \in \mathbb{R}^n \), which is basically supposed to be non-negative.
First, we introduce the bias model which is a generalization of interactions model [33]. In this model, the biased cost \( \tau \) is given by

\[
\tau = T + B(\pi),
\]

where \( B : \mathcal{S} \to \mathbb{R}^n \) is the bias function, whose \( k \)-th element obeys \( b_k(\pi_k) \). The function \( b_k : [0, 1] \to \mathbb{R} \) is assumed to satisfy the following:

**Assumption 1:** The function \( b_k (k \in \mathcal{A}) \) obeys the following items:

(i) \( b_k \) is continuous in \([0, 1]\) and continuously differentiable in \((0, 1)\).

(ii) \( b_k \) is decreasing in \( \pi_k \), and uniformly bounded by \([b^L, b^H]\) with some \( b^H \geq b^L > 0 \).

(iii) The first derivative \( \nabla b_k \) is uniformly bounded by \([-c^L, c^H]\) with some \( c^H \geq c^L > 0 \).

The item (ii) in Assumption 1 implies that if \( k \in \mathcal{A} \) is a majority strategy, the decision makers get a lower impression of cost than the actual cost \( T_k \). Hence, \( B(\pi) \) has the tendency that the strategy chosen by many people intensifies its own popularity, which corresponds to conformity bias. The first derivative \( \nabla b_k \) implies the dependency of \( b_k \) on \( \pi_k \), i.e., this corresponds to the strength of the bias. In particular, \( c^H \), which is the maximum of \( |\nabla b_k(\pi_k)| \), can be interpreted as the maximal bias-strength. In the sequel, we call the decision making model composed of (1) and (4) as Model 1. The block diagram of Model 1 is illustrated in Fig. 1.

In this paper, we address another model of conformity bias, presented in [25], [34]. For given \( T \), in this model, the biased cost \( \tau \) obeys

\[
\tau = W(\pi)T,
\]

where \( W(\pi) = \text{diag}(w_1(\pi_1), w_2(\pi_2), \ldots, w_n(\pi_n)) \) is the bias matrix. The \( k \)-th diagonal element \( w_k : [0, 1] \to \mathbb{R} (k \in \mathcal{A}) \) is the bias function for strategy \( k \). Throughout this paper, we set the following assumption for the bias function \( w_k \).

**Assumption 2:** The function \( w_k (k \in \mathcal{A}) \) obeys the following items:

(i) \( w_k \) is continuous in \([0, 1]\) and continuously differentiable in \((0, 1)\).

(ii) \( w_k \) is decreasing in \( \pi_k \), and uniformly bounded by \([w^L, w^H]\) with some \( w^H \geq w^L > 0 \).

(iii) The first derivative \( \nabla w_k \) is uniformly bounded by \([-v^L, v^H]\) with some \( v^H \geq v^L > 0 \).

In the subsequent discussion, we call the model composed of (1) and (5) as Model 2. Under Assumption 2 the biased cost for strategy \( k \) decreases when \( \pi_k \) gets large. In other words, the majority strategy tends to be well chosen. Since \( |\nabla w_k| \) is the dependency of \( w_k \) on \( \pi_k \), this implies the strength of the bias and its maximum value \( v^H \) represents the maximal bias-strength of Model 2. The block diagram of Model 2 is illustrated in Fig. 2.

### III. PASSIVITY ANALYSIS OF DECISION MAKING MODELS

In this section, we analyze the biased decision making models, introduced in the last section, in terms of \( \delta \)-passivity.

First, we consider Model 1, which is composed of (1) and (4). From (4), the signal \( \dot{\tau} \) is given by

\[
\dot{\tau} = \dot{T} + B'(\pi)\hat{\pi},
\]

where \( B'(\pi) := \text{diag}(\nabla b_1(\pi_1), \nabla b_2(\pi_2), \ldots, \nabla b_n(\pi_n)) \) is negative definite because of Assumption 1. From Lemma 1 and (5), we have the following lemma.

**Lemma 2:** Under Assumption 1 the system (1) with (4) is \( \delta \)-output-passivity-short from \(-\dot{T}\) to \( \hat{\pi} \).

**Proof:** Consider the function \( S(\tau, \pi) \) defined in (3) with substituting \( \tau = T + B(\pi) \). From Lemma 1 we have

\[
\dot{S} \leq -\dot{T}^T \hat{\pi} = -(\dot{T} + B'(\pi)\hat{\pi})^T \hat{\pi} = -\dot{T}^T \hat{\pi} - \hat{\pi}^T B'(\pi)\hat{\pi} \leq -\dot{T}^T \hat{\pi} + c^H \|\hat{\pi}\|^2.
\]

This completes the proof.

Focusing on the \( \delta \)-passivity, the block diagram of Model 1 can be represented as Fig. 3. Lemma 2 suggests that a positive feedback by \(-B'(\pi)\) appears on the outside of the non-biased logit model, as shown in Fig. 3. We can confirm from (7) that the impact coefficient incrementally varies according to the strength of the bias \( c^H \). In other words, the conformity bias modeled in (4) violates passivity of decision making.

We next analyze Model 2. Here, we define the diagonal matrix \( W'(\pi) := \text{diag}(\nabla w_1(\pi_1), \nabla w_2(\pi_2), \ldots, \nabla w_n(\pi_n)) \) which is negative definite due to Assumption 2. Then, the bias model (5) yields

\[
\dot{\tau} = W(\pi)\dot{T} + \text{diag}(W'(\pi)\hat{\pi})T = W(\pi)\dot{T} - \Phi(T, \pi)\hat{\pi},
\]

where \( \Phi(T, \pi) := -W'(\pi)\text{diag}(T) \) is a diagonal matrix. Substituting (8) to the inequality in Lemma 1 we can prove the following lemma which is a particular case of [34, Lemma 5].

**Lemma 3:** Suppose that there exists a constant \( T_{\text{max}} > 0 \) such that \( T_k(t) \leq T_{\text{max}} \) holds for all \( k \in \mathcal{A} \) and for all \( t \geq 0 \). Under Assumption 2 the system (1) with (5) is \( \delta \)-output-passivity-short from \(-\dot{T}\) to \( y := W(\pi)\hat{\pi} \).
In this section, we focus on the output-passivity-shortage of the decision making models, shown in the last section, and design mechanism based on passivity paradigm. In the sequel, we assume that the population state $\pi$ is observable, and the desired population state, denoted as $\pi^*$ is given as a constant vector. In this paper, a mechanism indicates the system to update the actual cost $\pi$ on the original logit model $(\pi)$.

A. Mechanism for Model 1

In this subsection, we consider a mechanism for Model 1. The structure of the mechanism is illustrated in Fig. 5. The gray part in Fig. 5 implies Model 1, and the block $\Sigma_1$ is the mechanism.

Proof: From (10), $\tilde{\pi} = \mu + \kappa \hat{\pi}$ holds. Noticing this result and (10a), we have

$$\dot{H} = \rho (\pi - \pi^* \right) \hat{\pi} = \mu \hat{\pi}$$

$$= \tilde{T}^T \hat{\pi} - \kappa \| \hat{\pi} \|^2.$$  

This result shows the $\delta$-input-strict-passivity of (10).

In the case of incentive design, the cost $T(t)$ is added to decision makers as an economic input. Whereas, in the case of nudge, it is announced to them as an informational input.
As a preparation for convergence analysis, we introduce the following lemma about the storage function $S$ under the mechanism \( \mathbf{10} \).

**Lemma 5:** Consider the system \( \mathbf{10} \). Let a signal $\pi(t) \in \text{int}(S)$ achieves $\|T(t)\| \to \infty$ when $t \to \infty$. For this specific signal $\pi(t)$, $S(T + B(\pi), \pi) \to \infty$ holds for any $\pi^* \in \text{int}(S)$.

**Proof:** See Appendix [II-A]

We are now ready to prove the following theorem. In the proof, we use the notation $\mathcal{L}_{\infty}$ as the set of all signals $u : [0, \infty) \to \mathbb{R}^n$ satisfying $\sup_{t \in [0, \infty)} \|u(t)\| < \infty$.

**Theorem 1:** Consider the feedback system composed of (1), (4) and (10). When $\kappa > c^H$ holds, the population state $\pi$ achieves $\lim_{t \to \infty} \pi(t) = \pi^*$.

**Proof:** We define the function $V_1 := S(T + B(\pi), \pi) + H(\pi)$. From Lemma 2 and Lemma 4 we obtain

$$\dot{V}_1 \leq -\bar{T}^T \dot{\pi} + c^H \|\dot{\pi}\|^2 + \bar{T}^T \dot{\pi} - \kappa \|\dot{\pi}\|^2$$

$$= -(\kappa - c^H) \|\dot{\pi}\|^2 \leq 0$$  \hspace{1cm} (12)

under $\kappa > c^H$. Denote the initial values of $S$ and $H$ by $S_0$ and $H_0$, respectively. Then, $V_1 \leq S_0 + H_0 < \infty$ holds for any time $t \geq 0$. Remark that $\pi \in \text{int}(S)$, this implies $T \in \mathcal{L}_{\infty}$ due to Lemma 5. Thus, we can apply the LaSalle’s invariance principle [38], and hence solution of (1), (4) and (10) for any initial conditions $T(0) \in \mathbb{R}^n$ and $\pi(0) \in \text{int}(S)$ converges to the largest invariant set satisfying $V_1 = 0$.

Consider the state trajectories such that $\dot{V}_1 = 0$ holds. From (12), $\dot{V}_1 = 0$ implies $\dot{\pi} = 0$. Hence, $\pi$ should be constant. Focusing on (10), $T \equiv \mu \geq 0$ holds and $\dot{\mu} = \rho(\pi - \pi^*)$ should be constant. If $\pi \neq \pi^*$, $\dot{T} \neq 0$ identically holds and thus $T$ should diverge. However, the divergence contradicts $T \in \mathcal{L}_{\infty}$. Accordingly, $T$ is constant and $\pi \equiv \pi^*$ is satisfied.

As a result, by invoking the LaSalle’s invariance principle, we can prove that $\pi$ asymptotically converges to $\pi^*$.

**B. Mechanism for Model 2**

Let us consider to design a mechanism for Model 2. Its structure is illustrated in Fig. 6. The gray part is the decision makers, and $\Sigma_2$ is the mechanism. Similar to [IV-A] we design $\Sigma_2$ to guarantee the passivity from $y$ to $\bar{T}$.

Applying the mechanism (10) to $\Sigma_2$, the following $\delta$-passivity is satisfied, that is proved in the previous work [34].

**Fact 1:** Under Assumption 1, and $\delta$-input-strictly-passive by using (10) for $\Sigma_2$, certainly, the whole system in Fig. 6 is composed of the feedback of a $\delta$-output-passivity-short system and a $\delta$-input-strictly-passive system. However, the analysis in [34] leaves an important issue about signal boundedness.

Therefore, we should redesign a mechanism for Model 2 to guarantee the existence of $T_{\text{max}}$, with keeping passivity.

To solve the above issue, we propose the following mechanism.

$$\dot{\mu} = \min \{\rho(\pi - \pi^*), -\alpha \mu\}, \mu(0) \leq 0,$$  \hspace{1cm} (13a)

$$T = \bar{T}1_n + \mu + \kappa(\pi - \pi^*),$$  \hspace{1cm} (13b)

where $\rho > 0$ and $\kappa > 0$ are design parameters, $\alpha > 0$ is a parameter satisfying $\alpha > \frac{1}{2\bar{\kappa}}$, and $\bar{T} > 0$ is a constant. Remark that (13a) ensures $\mu \leq 0$ for all time. Thus, the system (13b) guarantees the existence of $T_{\text{max}} < T + \kappa$ such that $T_k \leq T_{\text{max}} \forall k \in A$ holds for all time. About the time derivative of $T$, $\dot{T} = \dot{\mu} + \kappa \dot{\pi}$ holds from (13b).

For the passivity analysis of (13), we consider the primitive function of $w_k$. Under Assumption 2, $w_k$ is continuous. Hence, we can define the following function:

$$F_k(\pi_k) := \begin{cases} \int_0^{\pi_k} \left(\int_0^\phi w_k(\theta) d\theta\right) d\phi & \text{if } \pi_k \in [0, 1], \\ \infty & \text{otherwise.} \end{cases}$$  \hspace{1cm} (14)

For $\pi_k \in (0, 1)$, $\nabla^2 F_k(\pi_k) = w_k(\pi_k) \geq w^L$ clearly holds and this reveals that $F_k$ is convex. We denote $F_k^*: \mathbb{R} \to \mathbb{R}$ as the
Block diagram of a mechanism for Model 2. The gray part corresponds to the decision makers. The system enclosed by the red line is \(\delta\)-output-passivity-short (Lemma 3). The system \(\Sigma_2\) is a mechanism for Model 2 to make the system in blue line \(\delta\)-passive.

Due to the convexity of \(F_k\),
\[
\pi_k = \nabla F_k^\ast(\nabla F_k(\pi_k)), \quad \pi_k \in (0, 1)
\]  

is satisfied [39]. By using these results, we can prove the following lemma.

**Lemma 6:** Define \(\zeta_k := \nabla F_k(\min\{\pi_k, \pi_k - \frac{\alpha}{\rho}\mu_k\})\) and \(\zeta_k^\ast := \nabla F_k(\pi_k^\ast)\) for each \(k \in \mathcal{A}\). Under Assumption 2, the system \(\Sigma_k\) is \(\delta\)-input-strictly-passive from \(y\) to \(\dot{T}\) for the storage function
\[
U(\mu, \pi) := \sum_{k \in \mathcal{A}} U_k(\mu_k, \pi_k),
\]
\[
U_k(\mu_k, \pi_k) := \rho(F_k^\ast(\zeta_k) - F_k^\ast(\zeta_k^\ast)) - (\zeta_k - \zeta_k^\ast)\nabla F_k^\ast(\zeta_k^\ast).
\]

**Proof:** Due to the convexity of \(F_k^\ast\),
\[
F_k^\ast(\zeta_k) - F_k^\ast(\zeta_k^\ast) \geq (\zeta_k - \zeta_k^\ast)\nabla F_k^\ast(\zeta_k^\ast)
\]
holds and hence the function \(U_k\) becomes positive semi-definite. We first consider the time when mode switch does not happen in (13) for \(k\)-th element. Then, the time derivative of \(U_k\) along (13) is given by
\[
\dot{U}_k = \rho(\nabla F_k^\ast(\zeta_k) - \nabla F_k^\ast(\zeta_k^\ast))\dot{\zeta}_k.
\]

If \(\pi_k - \pi_k^\ast \leq -\frac{\alpha}{\rho}\mu_k\), then \(\zeta_k = \nabla F_k(\pi_k)\) holds and hence
\[
\dot{U}_k = \rho(\pi_k - \pi_k^\ast)\nabla^2 F_k(\pi_k)\pi_k = \rho(\pi_k - \pi_k^\ast)\pi_k y_k.
\]

Under \(\pi_k - \pi_k^\ast \leq -\frac{\alpha}{\rho}\mu_k\), \(\dot{T}_k = \rho(\pi_k - \pi_k^\ast) + \kappa\pi_k\) holds. Thus, we have
\[
\dot{U}_k = \dot{T}_k y_k - \kappa\pi_k y_k
\]
\[
= \dot{T}_k y_k - \kappa w_k(\pi_k)\pi_k^2
\]
\[
\leq \dot{T}_k y_k - \frac{\kappa}{2} w_k(\pi_k)\pi_k^2.
\]

Next, we consider the time when a mode switch occurs in (13) for \(k\)-th element. Since \(U_k(\mu_k, \pi_k)\) is not differentiable at \((\mu_k, \pi_k)\) satisfying \(\pi_k - \pi_k^\ast = -\alpha\mu_k\), we now take in the upper Dini derivative, denoted by \(D^+\dot{U}_k\). From the results in (18) and (20), we can confirm that
\[
D^+\dot{U}_k \leq \dot{T}_k y_k - \frac{\kappa}{2} w_k(\pi_k)\pi_k^2
\]
holds for all time \(t \geq 0\). Therefore, we obtain
\[
D^+U \leq \dot{T}^\top y - \frac{\kappa}{2} \pi^\top W(\pi)\pi
\]
\[
= \dot{T}^\top y - \frac{\kappa}{2} y^\top (W(\pi))^{-1} y
\]
\[
\leq \dot{T}^\top y - \frac{\kappa}{2\mu_H} \|y\|^2.
\]

Integrating (21) in time completes the proof.

Thanks to the result in Lemma 6, it is revealed that the proposed mechanism (13) guarantees both \(\delta\)-input-strict-passivity and the existence of \(T_{\text{max}}\). In other words, the remained issue in the previous method [34] can be cleared by (13). Thus, we
can expected to achieve the passivity-based mechanism for Model 2.

Here we introduce the following lemma, which will be used in convergence analysis.

**Lemma 7:** Consider the system (1). Let a signal \( \pi(t) \in \text{int}(S) \) achieves \( \|T(t)\| \to \infty \) when \( t \to \infty \). For this specific signal \( \pi(t) \), \( S(W(\pi)T, \pi) \to \infty \) holds for any \( \pi^* \in \text{int}(S) \).

**Proof:** See Appendix [13].

From Lemma 8 Lemma 6 and Lemma 7 we are now ready to analyze convergence of the proposed mechanism (13) for Model 2. We should remark that the LaSalle’s invariance principle [38] cannot be applied due to non-smoothness of the storage function \( U(\mu, \pi) \). Alternatively, we discuss the convergence by using the invariance principle for non-smooth Lyapunov functions [40]. Then, the following theorem can be proven.

**Theorem 2:** Consider the feedback system composed of (1), (3) and (13) under Assumption 2. When \( \kappa > \frac{2\omega(T_{\infty})}{\omega(T_{\infty})} \) holds, the population state \( \pi \) achieves \( \lim_{t \to \infty} \pi(t) = \pi^* \).

**Proof:** Define the function \( V_2(\pi) := S(W(\pi)T, \pi) + U(\mu, \pi) \). From Lemma 3 and Lemma 6 the upper Dini derivative of \( V_2 \) satisfies

\[
D^+ V_2 \leq -\dot{\pi}^T \pi + \pi^T \Phi(T, \pi) \dot{\pi} + \pi^T y - \frac{\kappa}{2} \pi^T W(\pi) \pi
\]

Under \( \kappa > \frac{2\omega(T_{\infty})}{\omega(T_{\infty})} \), \( \Phi(T, \pi) - \frac{\kappa}{2} W(\pi) < 0 \) is satisfied and hence \( D^+ V_2 \leq 0 \) holds. Denote the initial values of \( S \) and \( U \) by \( S_0 \) and \( U_0 \), respectively. Then, \( V_2 \leq S_0 + U_0 < \infty \) holds for all time. This implies \( T \in \mathcal{L}_\infty \) and \( \mu \in \mathcal{L}_\infty \) from Lemma 7 and \( \pi \in \text{int}(S) \). Accordingly, the invariance principle for non-smooth function [40] is applicable, and trajectories generated by (1), (3) and (13) for any initial conditions \( T(0) \in \mathbb{R}^n \) and \( \pi(0) \in \text{int}(S) \) converge to the largest invariant set satisfying \( D^+ V_2 = 0 \).

Let us now suppose the situation under \( D^+ V_2 = 0 \). From (22), \( D^+ V_2 = 0 \) yields \( \dot{\pi} = 0 \) and hence \( \pi \) must be constant. If \( \pi_k < \pi_k^* \), then \( k \)-th element of (13a) follows \( \dot{\mu}_k = \rho(\pi_k - \pi_k^*) < 0 \), which contradicts the boundedness of \( \mu \). Thus, \( \pi \geq \pi^* \) is identically satisfied. If \( \pi_k > \pi_k^* \) holds for some \( k \in \mathcal{A} \), \( \sum_{\tau \in \mathcal{A}} (\pi_k - \pi_k^*) \geq \pi_k - \pi_k^* > 0 \) must hold. However, this contradicts the fact \( 1_{\mathcal{A}}^\top (\pi - \pi^*) = 0 \) given by \( \pi \in \text{int}(S) \) and \( \pi^* \in \text{int}(S) \). Therefore, \( \pi \equiv \pi^* \) is satisfied when \( D^+ V_2 = 0 \) holds.

In summary, the invariance principle for non-smooth function [40] proves that \( \pi \) asymptotically converges to \( \pi^* \).

Thanks to the upper bound condition \( T_k(t) \leq T_{\max} \) \( \forall k \in \mathcal{A}, \forall t \geq 0 \) ensured by (13), the stability and convergence of the feedback system in Fig. 6 can be exactly guaranteed under the gain condition \( \kappa > \frac{2\omega(T_{\max})}{\omega(T_{\max})} \). Similar to Theorem 1 the result in Theorem 2 suggests the tendency that large gain \( \kappa \) will be required for strongly biased decision makers. The quantitative inequality is calculated as a result of passivity-based analysis.

**Remark 1:** The authors in [32] and [25] proposed a similar nudging mechanism to (10), with assuming that the signal \( Q(\tau) \) is observable. This assumption is different from the one dealt with in this paper. Whereas, the information \( Q(\tau) \) is not easy to get from the decision makers since it is implicit variable. Although \( Q(\tau) \) might be estimated by using \( \pi \) and \( \dot{\pi} \), there is another difficulty of identifying \( \eta \). Thus, we suppose the observation of population state \( \pi \), which is commonly used in the field of evolutionary game [23].

**Remark 2:** In [25], the authors addressed the biased population dynamics, and analyzed the convergence of nudging mechanism based on singular perturbation. The convergence condition in [25] implicitly relies on the update rate \( \eta \) of the decision makers (1). Meanwhile, the result in Theorem 2 shows two advantages against [25]. The first one is to explicitly clarify a quantitative condition for stability, which is a benefit of passivity-based analysis. In addition, the proposed mechanism in this paper can design the parameter \( \kappa \) independently of the update rate \( \eta \), which is the second contribution against the nudging method in [25].

**V. Conclusion**

In this paper, we addressed design of mechanisms for decision makers with conformity biases. We first introduced two types of bias models addressed in [33] and [25]. Next, we analyzed the population dynamics with the bias models in terms of \( \delta \)-passivity. Then, we clarified that conformity biases appear as positive feedback terms, and they break passivity of dynamic decision making. We furthermore presented passivity-based mechanisms for biased population dynamics, and showed convergence conditions for the proposed mechanisms. Accordingly, we confirmed that high gain feedback should have been required for the decision makers with strong biases.

**APPENDIX I**

**Radial Unboundedness of Storage Function**

The storage function \( S(\tau, \pi) \), defined in (3), satisfies the following lemma.

**Lemma 8:** Define \( \tau^H := \max\{\tau_1, \tau_2, \ldots, \tau_n\} \) and \( \tau^L := \min\{\tau_1, \tau_2, \ldots, \tau_n\} \). Under \( \pi \in \text{int}(S) \), then

\[
\tau^H - \tau^L \to \infty \Rightarrow S(\tau, \pi) \to \infty
\]

holds.

**Proof:** Let us first focus on the second term in the right hand side of (3), which is a convex optimization about \( \omega \). Remarking that \( \sum_{\tau \in \mathcal{A}} \omega_l \log \omega_l \) works as a barrier function to the inequality constraints, \( \omega^* \in \mathbb{R}^n \) is an optimal solution if and only if there exists \( \lambda^* \in \mathbb{R} \) satisfying the following conditions:

\[
\tau_k + \frac{1}{\beta} (1 + \log \omega_k^*) + \lambda^* = 0 \quad \forall k \in \mathcal{A},
\]

\[
1_{\mathcal{A}}^\top \omega^* = 1.
\]

From (23a), we obtain \( \log \omega_k^* = -\beta \tau_k - \beta \lambda^* - 1 \) and hence

\[
\omega_k^* = \exp(-\beta \tau_k - \beta \lambda^* - 1).
\]
Applying (23b) to (24), we have
\[
\exp(-\beta \lambda^* - 1) \sum_{l \in A} \exp(-\beta \tau_l) = 1,
\]
\[
\therefore \lambda^* = \frac{1}{\beta} \left( -1 + \log \left( \sum_{l \in A} \exp(-\beta \tau_l) \right) \right). \tag{25}
\]
Hence, we can calculate the second term in the right hand side of (3) as
\[
\min_{\omega \in S} \left( \omega^T \tau + \frac{1}{\beta} \sum_{l \in A} \omega_l \log \omega_l \right) = (\omega^*)^T \tau + \frac{1}{\beta} \sum_{l \in A} \omega_l^* \log \omega_l^* \]
\[
= \sum_{l \in A} \omega_l^* \left( \tau_l + \frac{1}{\beta} (\beta \lambda^* - \beta \tau_l - 1) \right) \]
\[
= \sum_{l \in A} \omega_l^* \left( -\lambda^* - \frac{1}{\beta} \right) = -\frac{1}{\beta} \log \left( \sum_{l \in A} \exp(-\beta \tau_l) \right).
\]
Thus, the storage function \( S(\tau, \pi) \) is given as
\[
S(\tau, \pi) = \eta \pi^T \tau + \frac{\eta}{\beta} \sum_{l \in A} \pi_l \log \pi_l + \frac{\eta}{\beta} \log \left( \sum_{l \in A} \exp(-\beta \tau_l) \right).
\]
Let us next focus on the terms depending on \( \tau \). Due to the property of the Log-Sum-Exp function [39],
\[
\frac{1}{\beta} \log \left( \sum_{l \in A} \exp(-\beta \tau_l) \right) \geq -\tau^L
\]
is satisfied. Noticing that \( 1^n \tau = 1 \) holds and there exists \( \varepsilon > 0 \) such that \( \pi_k > \varepsilon \) \( \forall k \in A \), we obtain
\[
\pi^T \tau + \frac{1}{\beta} \log \left( \sum_{l \in A} \exp(-\beta \tau_l) \right) \geq \pi^T \tau - \tau^L
\]
\[
= \sum_{l \in A} \pi_l (\tau_l - \tau^L) \geq \varepsilon (\tau^H - \tau^L). \tag{26}
\]
We now consider the case that \( (\tau^H - \tau^L) \rightarrow \infty \) happens. Then, \( \eta \pi^T \tau + \frac{3}{2} \log \left( \sum_{l \in A} \exp(-\beta \tau_l) \right) \) goes to infinity due to (26). Therefore, since \( \sum_{l \in A} \pi_l \log \pi_l \) is finite, \( S(\tau, \pi) \rightarrow \infty \) holds. This completes the proof.

\section*{APPENDIX II

PROOF OF LEMMAS}

\section*{A. Proof of Lemma 3}
Due to \( 1^n_1 \tau = 1 \) and \( 1^n_1 \pi^* = 1 \), the signal \( \mu^* \) of (10a) satisfies \( 1^n_1 \mu^* \equiv 0 \). Hence, the system (10) satisfies \( 1^n_1 T(t) \equiv 1^n_1 \mu(0) \). Define
\[
T^H(t) := \max\{T_1(t), T_2(t), \ldots, T_n(t)\},
\]
\[
T^L(t) := \min\{T_1(t), T_2(t), \ldots, T_n(t)\}.
\]
When \( \|T(t)\| \rightarrow \infty \) happens, \( T^H(t) \rightarrow \infty \) and \( T^L(t) \rightarrow -\infty \) hold due to the constraint \( 1^n_1 T(t) \equiv 1^n_1 \mu(0) \). Then, \( (T^H(t) - T^L(t) + b^L - b^H) \rightarrow \infty \) is also satisfied. Let us consider \( \tau^H \) and \( \tau^L \) defined in Lemma 3 under (4). Due to \( \tau^H \geq T^H + b^L \) and \( \tau^L \leq T^L + b^H \), we have \( \tau^H - \tau^L \geq T^H - T^L + b^L - b^H \). Thus, \( (\tau^H - \tau^L) \rightarrow \infty \) holds when \( \|T(t)\| \rightarrow \infty \) happens. From the above discussion and Lemma 3 when the system (10) achieves \( \|T(t)\| \rightarrow \infty \) under \( \pi(t) \in \text{int}(S) \), \( S(T + B(\pi), \pi) \rightarrow \infty \) holds. This completes the proof of Lemma 3.

\section*{B. Proof of Lemma 7}
Before we prove Lemma 7, we show the following lemma.
\begin{lemma}
Consider the signal \( T(t) \) generated by (13). For any constant \( \pi^* \in \text{int}(S) \) and for any signal \( \pi(t) \in \text{int}(S) \), the signal \( T^H(t) := \max\{T_1(t), T_2(t), \ldots, T_n(t)\} \) is bounded.
\end{lemma}

\begin{proof}
Let us suppose the case that the system (13) satisfies \( \|\mu_k\| \rightarrow \infty \) for all \( k \in A \). Since (13a) guarantees \( \mu_k \leq 0, \mu_k \rightarrow -\infty \) must hold for all \( k \in A \). Then, there exists a time \( \bar{t} \) > 0 such that \( -\frac{2}{\beta} \mu_k > \pi_k - \pi^*_k \) \( \forall k \in A \) holds for any time \( t \geq \bar{t} \). In other words, (15a) follows \( \mu = \rho(\pi - \pi^*) \) for any \( t \geq \bar{t} \). This yields \( 1^n_1 \mu = 0 \) and hence \( \sum_{l \in A} \mu_l(t) \) must be constant for any time \( t \geq \bar{t} \). This contradicts the assumption satisfying \( \mu_k \rightarrow -\infty \) \( \forall k \in A \).

Therefore, the system (13a) guarantees at least \( -\infty < \max\{\mu_1(t), \mu_2(t), \ldots, \mu_n(t)\} \). As a result, \( T^H(t) \) is bounded since \( \max\{\mu_1(t), \mu_2(t), \ldots, \mu_n(t)\} \rightarrow \infty \) holds under (13b).
\end{proof}

We now prove Lemma 7. Let us introduce the signals \( T^H(t) \) and \( T^L(t) \), defined in Appendix II-A. If (13) achieves \( \|T(t)\| \rightarrow \infty \), then \( T^L(t) \rightarrow -\infty \) must hold. This yields \( (T^H(t) - T^L(t)) \rightarrow \infty \) from Lemma 7. Then, \( (w^H T^H(t) - w^L T^L(t)) \rightarrow \infty \) is also satisfied. Let us consider \( \tau^H \) and \( \tau^L \) defined in Lemma 3 under (5). Due to \( \tau^H - \tau^L \geq w^H T^H(t) - w^L T^L(t), (\tau^H - \tau^L) \rightarrow \infty \) holds when \( \|T(t)\| \rightarrow \infty \) happens. Accordingly, Lemma 3 proves Lemma 7.

\section*{REFERENCES}
[1] G. Como, K. Savla, D. Acemoglu, M.A. Dahleh and E. Frazzoli, “Stability analysis of transportation networks with multiscale driver decisions,” SIAM Journal on Control and Optimization, vol. 51, no. 1, pp. 230–252, 2013.
[2] C.J. Day, B.F. Hobbs and J. Pang, “Oligopolistic competition in power networks: A conjectured supply function approach,” IEEE Transactions on Power Systems, vol. 17, no. 3, pp. 597–607, 2002.
[3] E. Bompadre, Y.C. Ma, R. Napoli, G. Gross and T. Guler, “Comparative analysis of game theory models for assessing the performances of network constrained electricity markets,” IET Generation, Transmission & Distribution, vol. 4, no. 3, pp. 386–399, 2010.
[4] N. Li, L. Chen and M. Dahleh, “Demand response using linear supply function bidding,” IEEE Transactions on Smart Grid, vol. 6, no. 4, pp. 1827–1838, 2015.
[5] R. Sugden, “Rational choice: A survey of contributions from economics and philosophy,” The Economic Journal, vol. 101, no. 407, pp. 751–85, 1991.
[6] D. Kahneman and A. Tversky, “Prospect theory: An analysis of decision under risk,” Econometrica, vol. 47, no. 2, pp. 263–291, 1979.
[7] I. Brocas and J.D. Carrillo, “Dual-process theories of decision-making: A selective survey,” Journal of Economic Psychology, vol. 41, pp. 45–54, 2014.
[8] Y. Guan, A.M. Annaswamy and H.E. Tseng, “Cumulative prospect theory based dynamic pricing for shared mobility on demand services,” Proc. IEEE 58th Conference on Decision and Control, pp. 2239–2244, 2019.
[9] D.M. Mason, L. Stella and D. Bauso, “Evolutionary game dynamics for crowd behavior in emergency evacuations,” Proc. IEEE 59th Conference on Decision and Control, pp. 1672–1677, 2020.
[10] T. Kameda and D. Nakanishi, “Cost-benefit analysis of social/cultural learning in a nonstationary uncertain environment: An evolutionary simulation and an experiment with human subjects,” *Evolution and Human Behavior*, vol. 23, no. 5, pp. 373–393, 2002.

[11] D. Fukuda and S. Morichi, “Incorporating aggregate behavior in an individual’s discrete choice: An application to analyzing illegal bicycle parking behavior,” *Transportation Research Part A*, vol. 41, no. 4, pp. 513–525, 2007.

[12] J. Urata and E. Hato, “Modeling the cooperation network formation process for evacuation systems design in disaster areas with a focus on Japanese megadisasters,” *Leadership and Management in Engineering*, vol. 12, no. 4, pp. 231–246, 2012.

[13] K. Train, *Discrete Choice Methods with Simulation*, SUNY-Oswego, Department of Economics, 2003.

[14] K. Srinivasan and H. Mahmassani, “Modeling inertia and compliance mechanisms in route choice behavior under real-time information,” *Transportation Research Record: Journal of the Transportation Research Board*, no. 1725, pp. 45–53, 2000.

[15] J. Sagebiel, “Preference heterogeneity in energy discrete choice experiments: A review on methods for model selection,” *Renewable and Sustainable Energy Reviews*, vol. 69, pp. 804–811, 2017.

[16] E. Ramírez-Llanos and N. Quijano, “A population dynamics approach for the water distribution problem,” *International Journal of Control*, vol. 83, pp. 1947–1964, 2010.

[17] W. Mei, N.E. Friedkin, K. Lewis and F. Bullo, “Dynamic models of appraisal networks explaining collective learning,” *IEEE Transactions on Automatic Control*, vol. 63, no. 9, pp. 2898–2912, 2018.

[18] J. Hofbauer and E. Hopkins, “Learning in perturbed asymmetric games,” *Games and Economic Behavior*, vol. 52, no. 1, pp. 133–152, 2005.

[19] J. Hofbauer and W.H. Sandholm, “Evolution in games with randomly disturbed payoffs,” *Journal of Economic Theory*, vol. 132, no. 1, pp. 47–69, 2007.

[20] M.J. Smith, “The stability of a dynamic model of traffic assignment: An application of a method of Lyapunov,” *Transportation Science*, vol. 18, no. 3, pp. 245–252, 1984.

[21] W.H. Sandholm, “Pairwise comparison dynamics and evolutionary foundations for Nash equilibrium,” *Games*, vol. 1, no. 1, pp. 3–17, 2010.

[22] M.J. Fox and J.S. Shamma, “Population games, stable games, and passivity,” *Games*, vol. 4, no. 4, pp. 561–583, 2013.

[23] S. Park, J.S. Shamma and N.C. Martins, “Passivity and evolutionary game dynamics,” *Proc. IEEE 57th Conference on Decision and Control*, pp. 3553–3560, 2018.

[24] A. Pashaie, L. Pavel and C.J. Damaren, “A population game approach for dynamic resource allocation problems,” *International Journal of Control*, vol. 90, no. 9, pp. 1957–1972, 2017.

[25] Y. Cheng and C. Langbort, “On informational nudging for boundedly rational decision makers,” *Proc. IEEE 57th Conference on Decision and Control*, pp. 4591–4796, 2018.

[26] W. Zhao, H. Yang, X. Deng and C. Zhong, “Stability of equilibrium for population games with uncertain parameters under bound rationality,” *Journal of Inequalities and Applications*, vol. 2021, no. 15, 2021.

[27] J.J. Laffont and D. Martimort, *The Theory of Incentives: The Principal-Agent Model*, Princeton University Press, 2002.

[28] R.H. Thaler and C.R. Sunstein, *Nudge: Improving Decisions about Health, Wealth, and Happiness*, Yale University Press, 2008.

[29] I. Ratcliffe, R. Dong, H. Ohlsson and S.S. Sastry, “Incentive design and utility learning via energy disaggregation,” *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 3158–3163, 2014.

[30] M. Shakarami, A. Cherukuri and N. Monshizadeh, “Nudging the aggregative behavior of noncooperative agents,” *Proc. IEEE 59th Conference on Decision and Control*, pp. 2579–2584, 2020.

[31] J.S. Weitz, C. Eksin, K. Paaporn, S.P. Brown and W.C. Ratcliffe, “An oscillating tragedy of the commons in replicator dynamics with game-environment feedback,” *Proc. the National Academy of Sciences*, vol. 113, no. 47, E7518–E7525, 2016.

[32] Y. Cheng and C. Langbort, “A model of informational nudging in transportation networks,” *Proc. IEEE 55th Conference on Decision and Control*, pp. 7598–7604, 2016.

[33] W.A. Brock and S.N. Durlauf, “Interactions-based models,” *Handbook of Econometrica*, vol. 5, pp. 3297–3380, 2001.

[34] S. Yamashita, T. Hatanaka, Y. Wasa, N. Hayashi, K. Hirata and K. Uchida, “Passivity-based analysis and nudging design for dynamics social model with bounded rationality,” *IFAC-PapersOnLine*, vol. 53, no. 5, pp. 338–343, 2020.

[35] T. Hatanaka, N. Chopra, M. Fujita and M.W. Spong, *Passivity-Based Control and Estimation in Networked Robotics*, Springer-Verlag, 2015.

[36] Z. Qu and M.A. Smaan, “Modularized design for cooperative control and plug-and-play operation of networked heterogeneous systems,” *Automatica*, vol. 50, no. 9, pp. 2405–2414, 2014.

[37] R. Sepulchre, M. Jankovic and P.V. Kokotovic, *Constructive Nonlinear Control*, Springer, London, 2012.

[38] H.K. Khalil, *Nonlinear Systems*, 3rd edition, Prentice Hall, 2002.

[39] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[40] D. Shevitz and B. Paden, “Lyapunov stability theory of nonsmooth systems,” *IEEE Transactions on Automatic Control*, vol. 39, no. 9, pp. 1910–1914, 1994.