Self-Consuming Generative Models with Co-Evolving Human Preferences

Xiukun Wei[1*], **Tian Xie**[1*], **Ding Zhu**[1], **Xueru Zhang**[1]
[1]The Ohio State University
{wei.1418, xie.1379, zhu.3723, zhang.12807}@osu.edu

Abstract

Generative models are increasingly trained in self-consuming iterative loops, where users curate preferred samples from model-generated candidates and the curated samples are used to train future generations of the model. Prior work has largely assumed fixed user preferences, but in practice exposure to model outputs gradually reshapes what users perceive as desirable, creating a feedback loop in which model distributions and user preferences co-evolve. We take a first step toward understanding the long-term behavior of such coupled dynamics. We show that when training relies entirely on user-curated synthetic data, iterative curation amplifies initial biases and drives the system toward one of multiple *singleton* equilibria in which the instance holding an initial advantage eventually dominates. In contrast, injecting reference data into training at a sufficiently large rate fundamentally changes the dynamics and yields a *unique globally attracting* equilibrium. Building on this insight, we study how reference-data injection can be used to control long-term outcomes, and propose an efficient algorithm that jointly selects a reference distribution and its mixing weight to steer the coupled system toward equilibria that preserve desired attributes while minimizing data collection costs.

1 Introduction

Modern generative models are increasingly trained on data that contains synthetic outputs produced by earlier versions of themselves, creating a *self-consuming* training loop. Recent studies have shown that such loops can lead to model collapse, training instability, and bias amplification [2, 3, 12, 14, 39, 40, 44], and have proposed various mitigation strategies [6, 14, 15]. In many practical settings, the synthetic data reused for retraining is further filtered through human curation [53]: model-generated samples are presented to users, who express preferences by selecting, rating, editing, reposting, or otherwise amplifying certain outputs over others, and the retained samples are subsequently incorporated into future training datasets. Prior work has shown that repeated retraining on such human-curated synthetic data can implicitly steer models toward distributions aligned with user preferences [11, 45].

However, existing works largely treat the preference signal as **fixed** or **exogenous**: users are assumed to consistently favor the same attributes, and human curation simply filters samples accordingly. In practice, this assumption is often violated because users and the model interact in a closed loop: the model itself reshapes the environment from which future preferences are formed, as its outputs shape what users are exposed to, which in turn shapes what they find desirable. Such model-influenced preference shifts are supported by recent empirical evidence: LLM-assisted writing systematically alters the tone, vocabulary, and meaning of human-produced text [1]; exposure to biased AI writing assistants shifts users' subsequent opinions on the underlying topic [19]; and AI-assisted creation more broadly homogenizes the diversity of content users produce [10]. In each case, synthetic outputs

[*]Equal contribution.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

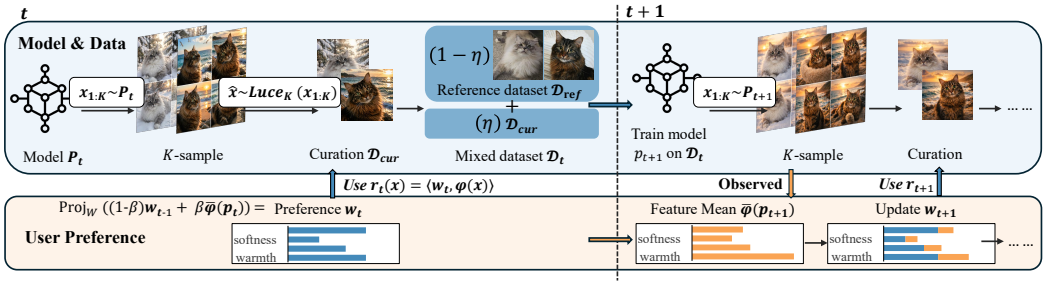

Figure 1: Overview of the coupled model-user dynamics. **Top row (model update):** the current model p_t generates K candidates, of which one is retained via preference-guided K-way Luce choice; the curated set $\mathcal{D}_{\text{cur}}$ is mixed with a reference dataset $\mathcal{D}_{\text{ref}}$ at rate η to form $\mathcal{D}_t$, on which p_{t+1} is trained. **Bottom row (preference update):** the user assesses how each candidate aligns with the current preference w_t via the reward $r_t(x) = \langle w_t, \varphi(x) \rangle$, where φ is the feature map, and the preference vector then drifts toward the mean feature direction $\bar{\varphi}(p_{t+1})$ of the updated model.

are not merely consumed by future models; they reshape the preference landscape that guides future curation, creating a closed feedback loop in which the model and its users co-evolve.

This paper sets out to study the long-term behavior of coupled user-model dynamics. We formulate a dynamical system over the joint state (p_t, w_t) in which the model distribution p_t and the underlying user preference vector w_t co-evolve over discrete time steps, as illustrated in Fig. 1. At each round t, the model generates K candidates and the user curates one according to their current preferences, with samples whose features align more closely with w_t selected with higher probability. The model is then retrained on the curated synthetic samples, optionally mixed with a fixed reference distribution, while the user preferences simultaneously drift toward the feature direction of the updated model.

Our analysis shows that the system exhibits qualitatively different long-run behavior depending on whether reference data is included in training. In the purely synthetic regime, where retraining relies entirely on user-curated synthetic samples, repeated curation amplifies local reward advantages: any item that is even slightly preferred under the current reward eventually dominates, driving the system toward one of multiple singleton equilibria. In contrast, when a sufficient fraction of reference data is mixed in at each round, the coupled update becomes a strict contraction on the joint state space under an explicit condition on the mixing weight, implying a unique globally attracting equilibrium that every trajectory converges to geometrically, regardless of initialization.

The uniqueness of the equilibrium under reference-data mixing opens a natural design question: how should the reference distribution and mixing weight be jointly chosen so that a specified level of diversity is preserved at equilibrium at minimum data-collection cost? We formalize this as a constrained optimization problem, which is bilevel and intractable since the constraint involves the equilibrium of a nonlinear coupled map with no closed form. To tackle this, we identify a worst-case relaxation that yields a linear sufficient condition on the reference distribution, decoupling the bilevel structure and reducing the problem to a family of linear programs indexed by the mixing weight. Experiments verify the theoretical results and demonstrate the effectiveness of the proposed algorithm.

Tbl. 1 summarizes how our paper extends the existing line of self-consuming generative models with human curation. In Appx. A, we discuss more related work, and our contributions can be summarized as follows:

- We propose a dynamic model that captures the co-evolution of user preferences and the model under self-consuming iterative training (Sec. 2).
- We theoretically analyze the convergence of the coupled dynamics and characterize the properties of its equilibria, both with and without reference data mixed into the retraining loop (Sec. 3).
- We formulate equilibrium steering as a constrained optimization problem and show that a worst-case sufficient condition reduces it to a family of linear programs, providing a computationally lightweight algorithm for jointly choosing the reference distribution and mixing weight (Sec. 4).
- We empirically verify the theoretical results and the effectiveness of the proposed algorithm on both image generation (CFM and DDPM on CIFAR-10) and text generation (GPT-2 on AG News). (Sec. 5)

Table 1: Comparison with prior work on self-consuming generative models with curation.

Work	Preference	# Models	Core conclusion	Risk	Role of real data
Ferbach et al. [11]	Fixed	Single	Curation implicitly optimizes a fixed preference signal.	Bias amplification	Stabilizer
Wei and Zhang [45]	Fixed	Single	Adversarial curation can systematically misalign the model.	Malicious feedback	Improves robustness
Zhao et al. [59]	Heterogeneous fixed	Single	Reference mixing yields convergence and stability under noisy heterogeneous curation.	Reward perturbation instability	Stabilizer and regularizer
Zhang et al. [58]	Fixed	Multi	Benign curation may backfire under model interaction.	Cross-model coupling	Convergence aid
This work	**Co-evolving**	**Single**	**Preferences co-evolve with the model, altering equilibrium structure.**	**Preference drift; diversity collapse**	**Stabilizer; Rreference data design**

2 Problem Formulation

Consider a platform that iteratively trains generative models using data curated by human users. Let $\mathcal{X} \subset \mathbb{R}^m$ be the instance space and let $p_t \in \mathcal{P}(\mathcal{X})$ denote the output distribution of the generative model at round t of the retraining loop. We present the model and results in the discrete setting with finite $\mathcal{X} = \{x_1, \dots, x_n\}$, motivated by the fact that in many platforms, outputs fall into a finite number of distinguishable categories (e.g., topic or object classes); the extension to continuous spaces is deferred to Appx. D.

User preference & human curated data. At each round t, the platform presents $K \geq 2$ synthetic samples $\{x_1, \dots, x_K\}$ drawn from the current model p_t to users, who then select a preferred sample $\hat{x}$ that is used to train the next generation of models. Specifically, let $\varphi : \mathcal{X} \to \mathbb{R}^d$ be a fixed feature map satisfying $\|\varphi(x)\| = 1$, which extracts features of an instance $x \in \mathcal{X}$ relevant to user preferences as a unit-norm representation[1]. User preferences over these features at round t are characterized by a vector $w_t \in \mathcal{W} := \{w \in \mathbb{R}^d : \|w\| \leq 1\}$. Given this preference vector, the reward of x at round t is defined as $r_{w_t}(x) := r(x; w_t) := \langle w_t, \varphi(x) \rangle \in [-1, 1]$, which measures the alignment between the instance features and the user's preference direction. Based on these rewards, users select $\hat{x}$ from the candidate set $\{x_1, \dots, x_K\}$ according to the K-way Luce rule:

$$\Pr(\hat{x} = x_k \mid x_1, \dots, x_K) = \frac{e^{\tau \cdot r_{w_t}(x_k)}}{\sum_{j=1}^{K} e^{\tau \cdot r_{w_t}(x_j)}} \tag{1}$$

where $\tau > 0$ controls the sharpness of user choice. When $K = 2$, this reduces to Bradley-Terry model [4, 30]. We denote this sampling process by $\hat{x} \sim \mathrm{Luce}_K(x_1, \dots, x_K)$.

Recursive training loop. Given a fixed reference (real data) distribution $p_{\mathrm{ref}} \in \mathcal{P}(\mathcal{X})$ and user curated synthetic data, the platform updates its model at round $t+1$ using either only user-curated synthetic data ($\eta = 1$) or a mixture of synthetic and reference data $\eta \in (0, 1)$:

$$p_{t+1} \in \arg\max_{p \in \mathcal{P}(\mathcal{X})} (1-\eta)\, \mathbb{E}_{x \sim p_{\mathrm{ref}}}[\log p(x)] + \eta \mathbb{E}_{\substack{x_1, \dots, x_K \sim p_t \\ \hat{x} \sim \mathrm{Luce}_K(x_1, \dots, x_K)}} [\log p(\hat{x})] \tag{2}$$

where mixing parameter $\eta \in [0, 1]$ controls the fraction of synthetic data used in training. Following Ferbach et al. [11], we show that the retraining step (2) admits a multiplicative reweighting of p_t (proof in Appx. C.1):

$$p_{t+1}(x) = (1-\eta) p_{\mathrm{ref}}(x) + \eta p_t(x) H^K_{p_t, w_t}(x), \tag{3}$$

where $H^K_{p_t, w_t}(x)$ is referred to as the *K-way curation operator*, which quantifies the amplification of x induced by K-way preference-based selection from p_t.

[1] If x is text, $\varphi(x)$ could encode tone, topic, or style; if x is an image, it may encode color, sharpness, or semantic content. See Appx. B for complete worked examples.

Definition 2.1 (Curation operator)**.** The K-way curation operator $H^K_{p_t} : \mathcal{X} \to \mathbb{R}_+$ is defined as

$$H^K_{p_t,w_t}(x) := \mathbb{E}_{x_1,\ldots,x_{K-1}\sim p_t}\left[\frac{Ke^{\tau\cdot r_{w_t}(x)}}{e^{\tau\cdot r_{w_t}(x)} + \sum_{i=1}^{K-1} e^{\tau\cdot r_{w_t}(x_i)}}\right]$$

Moreover, for any fixed $x \in \mathcal{X}$, this reweighting converges, as $K \to \infty$, to an exponential tilt of p_t, i.e., $H^K_{p_t,w_t}(x) \xrightarrow{K\to\infty} \frac{e^{\tau\cdot r_{w_t}(x)}}{\mathbb{E}_{z\sim p_t}[e^{\tau\cdot r_{w_t}(z)}]}$.

Preference drift & coupled dynamics. When users are exposed to model-generated content on the platform, their preferences may evolve over time. To capture this effect, we model the preference vector w_t as being gradually steered toward the feature distribution induced by the model. Specifically, define $\bar{\varphi}(p) := \mathbb{E}_{x\sim p}[\varphi(x)]$ as the mean feature vector under distribution p. After retraining at round $t+1$, user preferences are updated according to

$$w_{t+1} = \mathrm{Proj}_{\mathcal{W}}\left((1-\beta)w_t + \beta\bar{\varphi}(p_{t+1})\right) \tag{4}$$

where $\beta \in (0, 1]$ is the preference adaptation rate and $\mathrm{Proj}_{\mathcal{W}}$ denotes Euclidean projection onto $\mathcal{W}$.

Eq. (3) and (4) together define a closed-loop dynamical system over the state (p_t, w_t). We define an **equilibrium** $(p^\star_K, w^\star_K)$ as a fixed point of the joint dynamics, i.e., a state that remains unchanged under data generation and preference adaptation. Formally, $(p^\star_K, w^\star_K)$ satisfies

$$p^\star_K(x) = (1-\eta)p_{\mathrm{ref}}(x) + \eta p^\star_K(x)H^K_{p^\star_K,w^\star_K}(x) \qquad w^\star_K = \mathrm{Proj}_{\mathcal{W}}\left((1-\beta)w^\star_K + \beta\bar{\varphi}(p^\star_K)\right) \tag{5}$$

Objective. We study the evolution of the coupled dynamics (3)-(4) and characterize the long-run behavior of the model p_t and user preferences w_t (Sec. 3). Specifically, we ask: under what conditions does the system converge and admit an equilibrium; how do the choice set size K, the synthetic-data weight η, and the geometry of the feature map φ govern this behavior; and when do the dynamics collapse toward degenerate outcomes versus stabilize under reinjection of reference data.

Beyond stability analysis, we investigate the reference distribution p_{ref} as a steering mechanism. We ask how p_{ref} and the synthetic-data weight η can be jointly chosen so that the resulting equilibrium preserves desired attributes while keeping the cost of intervention small (Sec. 4).

3 Long-Term Behavior of User–Model Dynamics

Next, we study the evolution of (p_t, w_t). We begin with the purely synthetic curation setting ($\eta = 1$), and then extend to the setting in which each round mixes synthetic and reference samples ($\eta \in (0, 1)$).

3.1 Retraining with purely synthetic user-curated data

When $\eta = 1$, update (3) reduces to $p_{t+1}(x) = p_t(x)H^K_{p_t,w_t}(x)$. In this case, the dynamics are driven entirely by repeated preference-guided reweighting of the current distribution via the curation operator $H^K_{p_t,w_t}$. Before analyzing the full dynamics, we first study $H^K_{p_t,w_t}$ and identify its key properties.

Proposition 3.1 (Basic properties of curation operator)**.** *For any $p \in \mathcal{P}(\mathcal{X})$, $w \in \mathcal{W}$, and $K \geq 2$, the curation operator $H^K_{p,w}$ satisfies the following properties:*

1. ***Strict monotonicity.*** *For any $x_i, x_j \in \mathcal{X}$, $r_w(x_i) > r_w(x_j) \iff H^K_{p,w}(x_i) > H^K_{p,w}(x_j)$*
2. ***Normalization.*** $\sum_{i=1}^n p(x_i)H^K_{p,w}(x_i) = 1$.
3. ***Odds contraction.*** *If $r_w(x_i) - r_w(x_j) \geq r_\Delta > 0$, then*

$$\frac{H^K_{p,w}(x_j)}{H^K_{p,w}(x_i)} \leq \lambda_K(r_\Delta) := 1 - \frac{(1-e^{-\tau\cdot r_\Delta})(K-1)}{e^{2\tau}+K-1},$$

with $\lambda_K(r_\Delta) \in (0, 1)$ and $\lambda_K(r_\Delta) \xrightarrow{K\to\infty} e^{-\tau\cdot r_\Delta}$

See proof in Appx. C.2. *Strict monotonicity* implies that $H^K_{p,w}$ preserves the reward ordering induced by user preferences, ensuring higher-reward instances always receive higher selection weight. Moreover, *odds contraction* implies lower-reward items are disproportionately suppressed relative

to higher-reward ones, leading to an increasingly sharp separation in their selection probabilities. As a result, curation does not merely favor better items at each round; instead, it compounds their advantage over time by repeatedly amplifying higher-reward instances and pushing weaker ones further behind.

Equilibrium and convergence analysis. Building on Prop. 3.1, we next analyze the dynamics. Since each retraining step shifts mass toward higher-reward instances at the expense of lower-reward ones, so any distribution that places positive mass on instances with differing rewards is unstable.

Theorem 3.2 (Equilibrium characterization under purely synthetic setting)**.** *When $\eta = 1$, a distribution $p \in \mathcal{P}(\mathcal{X})$ is a fixed point of the retraining step under preference w if and only if all instances in its support share the same reward:*

$$p(x) > 0,\ p(x') > 0 \implies r_w(x) = r_w(x').$$

Consequently, for every instance $x_i \in \mathcal{X}$, the pair $(p^\star = \delta_{x_i}, w^\star = \varphi(x_i))$ is an equilibrium: the model has collapsed to a point mass on x_i while user preferences have locked onto its features.

See proof in Appx. C.3, where δ_{x_i} denotes the Dirac delta centered at x_i. Thm. 3.2 shows that training exclusively on user-curated synthetic data admits n distinct singleton equilibria, one centered at each instance x_i, in which the model concentrates all mass on a single instance and user preferences align entirely with its features. Since every instance supports a singleton equilibrium, the natural question is: *which of these equilibria are locally attracting?* The next theorem shows that any instance with a sufficient initial advantage will eventually dominate, a phenomenon we term "winner lock-in".

Theorem 3.3 (Local convergence to singleton equilibria)**.** *Fix $x_i \in \mathcal{X}$ and suppose $\varphi(x_j) \neq \varphi(x_i)$ for all $j \neq i$. Define the reward gap $\gamma_i = \min_{j \neq i} \langle \varphi(x_i),\ \varphi(x_i) - \varphi(x_j) \rangle > 0$ and let $\lambda_K(\cdot)$ be the odds-contraction factor from Prop. 3.1. If the initial state satisfies*

$$1 - p_0(x_i) \leq \varepsilon_i, \qquad \|w_0 - \varphi(x_i)\| \leq \frac{\gamma_i}{4}$$

for some $\varepsilon_i \in \left(0, \min\left\{\frac{1}{2},\ 1 - \lambda_K(\frac{\gamma_i}{2}),\ \frac{\gamma_i}{8}\right\}\right)$, then for all $t \geq 0$,

$$1 - p_t(x_i) \leq \left(\frac{\lambda_K(\frac{\gamma_i}{2})}{1 - \varepsilon_i}\right)^t \big(1 - p_0(x_i)\big).$$

Consequently, $p_t(x_i) \xrightarrow{t\to\infty} 1$ and $w_t \xrightarrow{t\to\infty} \varphi(x_i)$.

See proof in Appx. C.4. Thm. 3.3 shows that if the initial distribution and preferences are sufficiently biased toward x_i, this bias is self-reinforcing: the mass on all other instances $1 - p_t(x_i)$ decays geometrically and p_t converges to the singleton equilibrium δ_{x_i}. Reward gap γ_i is the minimum reward advantage that x_i holds over any competitor x_j when preferences w are fully aligned with $\varphi(x_i)$: a larger gap expands the set of initial conditions from which convergence to δ_{x_i} is guaranteed.

Remark 3.4 (Role of the choice set size K)**.** Thm. 3.3 holds for every $K \geq 2$, but K governs the speed of convergence. Since λ_K decreases with K, larger choice sets make curation more selective and accelerate the decay of mass on competing instances. In the limit $K \to \infty$, the update approaches exponential tilting, the most aggressive form of preference-guided reweighting.

Note that the condition $\varphi(x_j) \neq \varphi(x_i)$ in Thm. 3.3 is necessary. When two distinct instances share the same feature direction, curation cannot distinguish between them on the basis of rewards alone, and transferring mass from x_i to such a duplicate leaves both the reward geometry and the induced preference direction unchanged. Local stability therefore fails, as we show below.

Corollary 3.5 (Failure of local stability)**.** *Fix $x_i \in \mathcal{X}$. If there exists x_j with $j \neq i$ such that $\varphi(x_j) = \varphi(x_i)$, then the singleton equilibrium δ_{x_i} is not locally asymptotically stable.*

See proof in Appx. C.5. Thm. 3.2, Thm. 3.3 and Cor. 3.5 together characterize the long-term behavior of the dynamics when models are retrained exclusively on human-curated synthetic data: it admits n singleton equilibria, one for each instance; under generic feature separation, any instance with a sufficient initial advantage in model mass and preference alignment locks in as the unique long-run outcome; and when two instances share the same feature direction, the local stability fails.

3.2 Retraining on mixed synthetic and reference data

We next consider $\eta \in (0,1)$, where each retraining round mixes curated synthetic data with a fixed reference distribution p_{ref} according to (3). Unlike purely synthetic setting, p_{ref} prevents the dynamics from collapsing to a single instance and instead drives them toward a globally stable equilibrium.

To analyze convergence and establish contraction of the dynamics, the following lemma quantifies how perturbations in the model distribution p and user preference w propagate through the retraining step $p \to (1-\eta)p + \eta p H^K_{p,w}$ and serves as the key technical tool for the convergence results below.

Lemma 3.6 (Sensitivity of the retraining step)**.** *Let $d_{\mathrm{TV}}(p,q) := \frac{1}{2}\sum_{x\in\mathcal{X}} |p(x)-q(x)|$ denote total variation distance. For any $p, q \in \mathcal{P}(\mathcal{X})$, $w, v \in \mathcal{W}$, and $K \geq 2$, the following hold:*

$$d_{\mathrm{TV}}\left(pH^K_{p,w}, qH^K_{q,w}\right) \leq L_p d_{\mathrm{TV}}(p,q); \quad d_{\mathrm{TV}}\left(pH^K_{p,w}, pH^K_{p,v}\right) \leq L_w \|w-v\|; \quad \|\bar{\varphi}(p) - \bar{\varphi}(q)\| \leq 2 d_{\mathrm{TV}}(p,q)$$

where $L_p := 2e^{2\tau}(2+e^{2\tau})$ and $L_w := 2\tau e^{4\tau}$.

See proof in Appx. C.6. Lemma 3.6 implies that the retraining step is Lipschitz continuous in both p and w: the first bound controls how perturbations in p propagate through the curation step; the second controls the sensitivity to changes in w; and the third shows that distributional perturbations translate to bounded perturbations in the mean feature vector entering the preference update (4).

Equilibrium and convergence analysis. The coupled system (3)-(4) is difficult to analyze directly: the model update depends on the current preference w_t, while the preference update depends on the updated distribution p_{t+1}. We therefore begin by fixing w and studying the model update in isolation. This decoupling breaks the circularity and allows us to characterize the long-term model distribution for any given w. Lem. 3.7 below shows that, unlike the purely synthetic setting which admits n singleton equilibria with long-run outcome determined by initial conditions (Thm. 3.2), mixing with p_{ref} yields a **unique** long-run distribution, provided η is small enough to make the update contractive.

Lemma 3.7 (Unique equilibrium under frozen-w)**.** *Fix $w \in \mathcal{W}$, and consider the the frozen-w model update $p_{t+1}(x) = (1-\eta)p_{\mathrm{ref}}(x) + \eta p(x) H^K_{p,w}(x)$. If $\eta L_p < 1$, the update is a strict contraction on $(\mathcal{P}(\mathcal{X}), d_{\mathrm{TV}})$. Consequently, there exists a unique equilibrium $p^{K,\eta}_w$ and every trajectory converges to it geometrically:*

$$d_{\mathrm{TV}}(p_t, p^{K,\eta}_w) \leq (\eta L_p)^t \, d_{\mathrm{TV}}(p_0, p^{K,\eta}_w).$$

See proof in Appx. C.7. In general, the equilibrium $p^{K,\eta}_w$ must be computed iteratively. In the large-K limit, however, it admits a closed-form expression.

Proposition 3.8 (Closed-form equilibrium)**.** *As $K \to \infty$, the frozen-w equilibrium takes the form*

$$p^{\infty,\eta}_w(x) = (1-\eta)p_{\mathrm{ref}}(x)\,\frac{Z^\eta_w}{Z^\eta_w - \eta e^{\tau\cdot r_w(x)}}.$$

where $Z^\eta_w := \mathbb{E}_{z\sim p^{\infty,\eta}_w}[e^{\tau\cdot r_w(x)}]$ is the unique scalar satisfying

$$\sum_{i=1}^{n}(1-\eta)\,p_{\mathrm{ref}}(x_i)\,\frac{e^{\tau\cdot r_w(x_i)}}{Z^\eta_w - \eta e^{\tau\cdot r_w(x_i)}} = 1, \qquad Z^\eta_w > \eta \max_i e^{\tau\cdot r_w(x_i)}.$$

See proof in Appx. C.7. The closed form reveals how the equilibrium balances three forces: p_{ref} provides a baseline, synthetic-data weight η amplifies the preference-induced exponential tilt $e^{\tau r_w(\cdot)}$ and concentrates mass on high-reward instances, while the constant Z^η_w ensures the result is a valid probability distribution. As $\eta \to 0$, the equilibrium recovers p_{ref}; as $\eta \to 1$, the tilt dominates and mass concentrates on the highest-reward instance, aligned with results of the purely synthetic setting.

With the frozen-w equilibrium characterized, we now return to the full coupled system in which both the model distribution and user preferences evolve simultaneously.

Theorem 3.9 (Convergence of coupled dynamics)**.** *Consider the coupled dynamics (3)-(4). If $\eta(L_p + 2L_w) < 1$, then there exists $\xi \in \left(\frac{\eta L_w}{\beta(1-2\eta L_w)}, \frac{1-\eta L_p}{2\beta\eta L_p}\right)$ such that the coupled update is a strict contraction on $(\mathcal{P}(\mathcal{X})\times\mathcal{W}, d_\xi)$, where $d_\xi\big((p,w),(q,v)\big) := d_{\mathrm{TV}}(p,q) + \xi\|w-v\|$.*

Consequently, there exists a unique equilibrium $(p_K^\star, w_K^\star)$ satisfying (5), and every trajectory converges to it geometrically.

See proof in Appx. C.8. The metric d_ξ combines distributional and preference distances into a single metric on $\mathcal{P}(\mathcal{X}) \times \mathcal{W}$; the window condition on ξ ensures both coordinates contract simultaneously, and the contraction condition $\eta(L_p + 2L_w) < 1$ ensures the reference mixing damps both feedback loops sufficiently for the Banach fixed-point theorem to apply. Compared with the frozen-preference condition $\eta L_p < 1$ in Lem. 3.7, the extra L_w term captures the stability cost of preference drift.

Thm. 3.9 highlights a key difference from the purely synthetic setting: whereas pure self-consumption leads to path-dependent collapse driven by initial conditions, mixing with p_{ref} yields a unique equilibrium to which all trajectories converge, regardless of initialization. The resulting equilibrium depends entirely on p_{ref} and η, which raises a natural design question: *how should p_{ref} and η be chosen so that the equilibrium preserves desired attributes?* We address this in Sec. 4.

4 Steering Equilibria via Reference Data Mixing

The uniqueness of the equilibrium under reference data mixing, established in Thm. 3.9, motivates the use of p_{ref} and η as **design variables**: by choosing them appropriately, one can steer the coupled dynamics toward a desired equilibrium. Specifically, denote $(p_K^\star(p_{\text{ref}}, \eta),\ w_K^\star(p_{\text{ref}}, \eta))$ as the unique equilibrium attained under reference distribution p_{ref} and mixing weight η, the corresponding mean feature vector is $\bar{\varphi}(p_K^\star(p_{\text{ref}}, \eta)) := \mathbb{E}_{x \sim p_K^\star(p_{\text{ref}}, \eta)}[\varphi(x)]$.

To specify which attributes should be preserved at equilibrium, suppose the platform identifies L features of interest, each represented by a *unit-norm* direction $v_\ell \in \mathbb{R}^d$ in the feature space. For example, v_ℓ might point toward the feature vectors of instances belonging to a particular genre, topic, or demographic group. The quantity $\langle v_\ell,\ \bar{\varphi}(p_K^\star)\rangle$ then measures how well the equilibrium distribution represents feature ℓ on average. Given thresholds $\theta_1, \ldots, \theta_L > 0$ specifying the minimum acceptable representation for each feature, we say that (p_{ref}, η) satisfies the **preservation constraint** if $\langle v_\ell,\ \bar{\varphi}(p_K^\star(p_{\text{ref}}, \eta))\rangle \geq \theta_\ell, \forall \ell \in \{1, \cdots, L\}$. For example, in an image generation platform with three categories, setting $L = 1$, $v_1 = (\underbrace{1}_{\text{bird}}, \underbrace{0}_{\text{cat}}, \underbrace{0}_{\text{dog}})$ and $\theta_1 = 0.05$ requires that bird content constitutes at least 5% of the equilibrium output.

Let N denote the number of training samples used at each training round. The reference distribution $p_{\text{ref}} = (p_{\text{ref},1}, \cdots, p_{\text{ref},n})$ specifies the allocation of reference samples across instances: at each round, $p_{\text{ref},i} \cdot (1-\eta) \cdot N$ fresh samples of instance x_i are collected and added to training. Collecting these samples is not free. Rare or specialized instances (e.g., those representing underrepresented languages or domain-specific content) require more effort to source, label, and verify than common ones. We capture this heterogeneity by assigning each instance x_i a per-sample cost $c_i > 0$, so that the **total collection cost** per round is $J(p_{\text{ref}}, \eta) := (1-\eta) \cdot N \cdot \sum_{i=1}^n c_i p_{\text{ref},i}$

Design objective. Our goal is to find (p_{ref}, η) that minimizes the total collection cost subject to the preservation constraint at the long-run equilibrium. This is naturally formulated as a bilevel problem: the outer level optimizes over (p_{ref}, η), and the inner level computes the equilibrium that determines whether the preservation constraint is met:

$$\min_{p_{\text{ref}} \in \mathcal{P}(\mathcal{X}),\ \eta \in (0,1)} \quad J(p_{\text{ref}}, \eta) \quad \text{s.t.} \quad \langle v_\ell,\ \bar{\varphi}(p_K^\star(p_{\text{ref}}, \eta))\rangle \geq \theta_\ell, \quad \ell = 1, \ldots, L \tag{6}$$

However, (6) is difficult to solve directly: evaluating the preservation constraint requires computing $p_K^\star(p_{\text{ref}}, \eta)$, which is the fixed point of a nonlinear coupled map and has no closed form. To tackle this, we propose a worst-case relaxation of the preservation constraint that depends directly on p_{ref} and η, bypassing the need to compute the equilibrium explicitly. Enforcing this relaxed condition decouples the bilevel structure and reduces (6) to a collection of linear programs indexed by η.

Proposition 4.1 (Sufficient condition for preservation). *Fix $\eta \in (0, \frac{1}{(L_p+2L_w)})$. If $\forall \ell \in \{1, \ldots, L\}$*

$$\sum_{i=1}^n p_{\text{ref},i} \langle v_\ell, \varphi(x_i)\rangle \ \geq \ \frac{\theta_\ell - \eta \cdot \min_{i=1,\cdots,n} \langle v_\ell, \varphi(x_i)\rangle}{1-\eta},$$

then the equilibrium satisfies the preservation constraint $\langle v_\ell, \bar{\varphi}(p_K^\star(p_{\text{ref}}, \eta))\rangle \geq \theta_\ell$, $\forall \ell$.

See proof in Sec. C.9. Replacing the equilibrium constraint in (6) by the sufficient condition of Prop. 4.1 eliminates the bilevel structure entirely. Since η appears only in the right-hand side and in the cost factor $(1-\eta)$, fixing η decouples the problem into a linear program over p_{ref}.

Algorithm. The complete design procedure combines an outer search over η with an inner linear program (LP) solve for p_{ref}, as summarized in Alg. 1. The search range is $(0, \eta_{\max})$, where $\eta_{\max}$ upper-bounds mixing weight. Thm. 3.9 provides a conservative range $(0, \frac{1}{(L_p+2L_w)})$ for η, but in practice, $\eta_{\max}$ can be set larger since the Lipschitz constants are worst-case. The procedure is performed entirely offline prior to deployment and remains computationally lightweight: the inner problem is a linear program with n variables and $L+1$ constraints, and the outer problem is one-dimensional.

Algorithm 1 Cost-Optimal Reference Data Mixing

Require: Data size N, feature vectors $\{\varphi(x_i)\}_{i=1}^n$, directions $\{v_\ell\}_{\ell=1}^L$, thresholds $\{\theta_\ell\}_{\ell=1}^L$, costs $\{c_i\}_{i=1}^n$
Ensure: Reference distribution $p_{\text{ref}}^\star$, mixing weight $\eta^\star$

```
J_min ← +∞
for η on a grid in (0, η_max) do
    Solve LP (7) → p_ref(η), C*(η)
    if LP feasible and J(η) < J_min then
        J_min ← J(η),  η* ← η,  p*_ref ← p_ref(η)
    end if
end for
return p*_ref, η*
```

Inner problem: optimal reference distribution for fixed η. For any fixed $\eta \in (0, \bar{\eta})$, Prop. 4.1 reduces the design task to the following linear program:

$$C^\star(\eta) \;:=\; \min_{p_{\text{ref}} \in \mathcal{P}(\mathcal{X})} \sum_{i=1}^n p_{\text{ref},i}\, c_i \quad \text{s.t.} \quad \sum_{i=1}^n p_{\text{ref},i}\, \langle v_\ell, \varphi(x_i)\rangle \;\geq\; \frac{\theta_\ell - \eta \min_i \langle v_\ell, \varphi(x_i)\rangle}{1-\eta},\ \forall \ell \quad (7)$$

The solution $p_{\text{ref}}(\eta)$ is the cheapest reference distribution that guarantees the preservation constraint at equilibrium for the given η. By standard linear programming theory, an optimal solution always exists with at most $L+1$ strictly positive entries, meaning the optimal p_{ref} concentrates on at most $L+1$ instances. The platform therefore needs to actively source only a small number of distinct instance types, regardless of how large $\mathcal{X}$ is. Each inner problem can be solved efficiently offline using standard LP solvers such as simplex or interior-point methods.

Outer problem: selecting the mixing weight. Given $C^\star(\eta)$, the per-round collection cost is $J(\eta) := (1-\eta)NC^\star(\eta)$. The outer problem

$$\eta^\star \in \arg\min_{\eta \in (0,\bar{\eta})} (1-\eta)\, N\, C^\star(\eta). \quad (8)$$

balances two competing effects. A larger η means fewer reference samples per round, reducing the direct cost $(1-\eta)\, N\, C^\star(\eta)$; but it simultaneously tightens the preservation constraint, since the right-hand side of (7) grows with η, forcing $C^\star(\eta)$ to increase. The optimal $\eta^\star$ resolves this tension, and since the problem is one-dimensional, it can be found efficiently by a simple grid search.

5 Experiments

This section aims to empirically illustrate our theoretical results in a realistic training condition, where finite samples, imperfect model training, and a concrete model architecture introduce errors absent from the abstract analysis[2].

Datasets. We conduct experiments on two datasets (i) Image dataset: **CIFAR-10** [24]: It contains 60,000 images from 10 classes $\{$airplane $:= 0$, automobile $:= 1$, bird $:= 2$, cat $:= 3$, deer $:= 4$, dog $:= 5$, frog $:= 6$, horse $:= 7$, ship $:= 8$, truck $:= 9\}$. (ii) Text dataset: **AG News** [54]: It contains 127,600 texts from 4 classes $\{$world $:= 0$, sports $:= 1$, business $:= 2$, science/technology $:= 3\}$

Models. We experiment with three generative architectures: OT-CFM [28, 43] and DDPM [17, 35] for image generation, and GPT-2 [37] for text generation. We present OT-CFM results in this section; DDPM and GPT-2 results appear in Appx. E.2 and E.3, respectively. All models are adopted from

[2]Please see our code at https://github.com/osu-srml/Model-Reward-Coevolution

public GitHub repositories and are used with their default hyperparameters. Each model is initialized from its released pretrained checkpoint, keeping architectures and optimization settings unchanged.

Feature map φ and preference w_t. For CIFAR-10, we use a pretrained ResNet-56 [16] classifier (accuracy $\geq 93\%$). The feature map $\varphi(x) \in \mathbb{R}^{10}$ is the normalized softmax output, where $\varphi(x)_c$ represents the predicted probability of class c. The preference vector $w_t \in \mathbb{R}^{10}$ is updated via Eq. (4) with projection onto the unit ball. For AG News, we use a pretrained BERT [8] classifier with 4-dimensional normalized softmax features; see Appx. E.3 for details.

Iterative Retraining. For OT-CFM on CIFAR-10, each round t proceeds as follows. The current model generates 10,000 candidate images, from which 5,000 are selected via K-way Luce curation with reward $r_t(x) = \langle w_t, \varphi(x) \rangle$. The curated samples are mixed with reference data at ratio η to form a training set of size 5,000, on which the model is fine-tuned for 1,000 steps. After training, 10,000 fresh samples are generated for evaluation: we compute per-class proportions and update w_t. We run each experiment for 20 rounds. We set $\tau = 5.0$, $K = \infty$, and $\beta = 0.01$ for the main experiments. The effect of each parameter is discussed in Appx. E.1.1. Due to the computational cost of training and sampling from generative models, we report a single run for each setting, with the random seed varying across rounds. To verify reproducibility, Appx. E.2 reports three independent runs under the same setting, confirming that the trajectories are consistent.

Different initializations, different outcomes ($\eta = 1$). Under pure synthetic training, the system exhibits initialization-dependent lock-in. When $w_0 = [1, 0, 0, 0, \dots]$ (prefer airplane), curation progressively favors airplane images. The model's airplane proportion rises from $\sim$10% to >95% over 20 rounds as shown in Fig. 2(a), while all other classes collapse below 5%. When $w_0 = [0, 0, 0, 1, \dots]$ (prefer cat), the same dynamics drive the model toward cat instead (Fig. 2(c)). In both cases, the curation signal is strong enough ($\tau = 5.0$, $K = \infty$) to steer the entire generated distribution. Appx. E.1.1 varies β and shows that preference adaptation mainly changes the speed of preference drift, while the qualitative lock-in behaviors remain consistent.

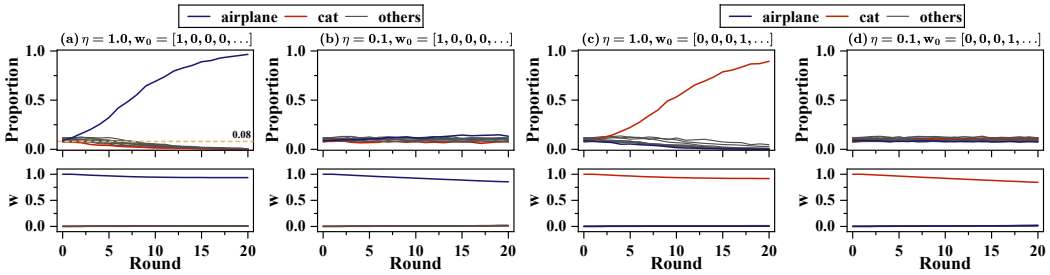

Figure 2: Co-evolution dynamics under different initializations and mixing ratios. **Top**: per-class generation proportions. **Bottom**: preference vector w_t.(a) The model locks into airplane, reaching >95% by round 20. (b) Uniform reference data stabilizes the distribution near equilibrium. (c) A different initialization leads to a different outcome. (d) The system converges to a similar equilibrium, independent of initialization.

Enough reference data yields a unique equilibrium ($\eta < 1$). When 90% of each training batch comes from the uniform CIFAR-10 training set ($\eta = 0.1$), regardless of whether $w_0 = [1, 0, 0, 0, \dots]$ (Fig. 2(b)) or $w_0 = [0, 0, 0, 1, \dots]$ (Fig. 2(d)), the class proportions remain close to uniform across all 20 rounds. The two runs converge to a similar equilibrium, consistent with the uniqueness guarantee of Thm. 3.9: for sufficiently small η, the reference distribution stabilizes the system to a single fixed point independent of initialization.

Reference data design. We evaluate the effectiveness of Alg. 1 for preserving minority classes under co-evolution. We set the goal is to guarantee that the last four classes (frog, horse, ship, truck) each maintain a proportion of at least $\theta = 0.08$ at equilibrium. We assign heterogeneous costs reflecting class rarity: cheap classes (airplane, automobile: $c = 1$) are abundant, while protected classes are expensive (frog: 30, horse: 28, ship: 27, truck: 25). The preference is initialized to $w_0 = [1, 0, 0, 0, \dots]$ (airplane).

Tbl. 2 compares our designed reference with uniform reference and Fig. 3 shows the class proportions under our designed reference. Our designed reference achieves two advantages simultaneously. First, it uses a higher synthetic ratio ($\eta = 0.67$) than either uniform baseline, reducing the total amount of

Table 2: Comparison of reference data strategies. C^*: per-sample cost; J: total cost $(1-\eta)\cdot N\cdot C^*$. Last four columns: proportion of each protected class at round 20. $\theta_\ell = 0.08$ for all protected classes.

p_{ref}	η	C^*	J	frog	horse	ship	truck
Uniform	0.50	16.4	41,000	.067	.077	.059	.067
Uniform	0.20	16.4	65,600	.112	.113	.097	.105
Ours	**0.67**	**26.81**	**44,043**	**.084**	**.113**	**.088**	**.092**

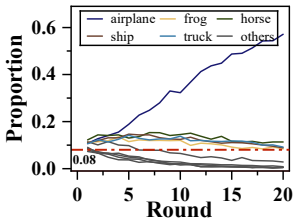

Figure 3: Class proportions under designed reference.

reference data required. Second, at round 20, all protected classes exceed the threshold $\theta_\ell = 0.08$, suggesting that the designed reference distribution steers the trajectory toward a feasible equilibrium. The partial LP search results over a uniform grid of 50 points in $\eta \in [0.2, 0.9]$ and results under a different setting are provided in Appx. E.1.2.

6 Discussion

Our analysis focuses on a basic co-evolutionary setting in which a single aggregate preference state interacts with a self-consuming generative model through preference-guided curation. The same framework naturally extends to richer forms of preference dynamics. In particular, we consider nonlinear gap-response updates that include saturating and state-dependent gated responses; the model-update branch remains unchanged, while the stability analysis becomes response-dependent. Beyond this, stochastic individual-level adaptation provides a natural micro-level interpretation of the deterministic update studied here, while heterogeneous user populations lead to coupled group-specific preferences interacting through a shared generative distribution. We discuss these extensions and the resulting theoretical questions in Appx. F.

7 Conclusion

We studied self-consuming generative retraining in which the model distribution and user preferences co-evolve. The system was formulated as a coupled dynamic over (p_t, w_t), with the model updated by reference-mixed user-curated data and the preference vector drifting toward the model's mean feature. Under purely synthetic curation ($\eta = 1$), the dynamics admit singleton equilibria, and any instance with a sufficient initial advantage locks in. With reference mixing satisfying $\eta(L_p + 2L_w) < 1$, the coupled update becomes a strict contraction with a unique, globally attracting equilibrium. Building on this, we further formulate equilibrium steering as a bilevel problem and reduce it to a family of linear programs indexed by η. Experiments empirically validate our analysis and confirm the effectiveness of our design. Several directions remain open: relaxing the exact observability of φ and w_t to partially observed quantities, extending the contraction analysis to nonlinear rewards, and generalizing the dynamics to multi-platform settings with heterogeneous user populations.

Acknowledgments and Disclosure of Funding

This material is based upon work supported by the U.S. National Science Foundation under Grant Nos. IIS-2202699, IIS-2416895, IIS-2301599, CMMI-2301601, DMS-2529302, and OIA-2616279. The authors declare no competing interests.

References

[1] Marwa Abdulhai, Isadora White, Yanming Wan, Ibrahim Qureshi, Joel Leibo, Max Kleiman-Weiner, and Natasha Jaques. How llms distort our written language, 2026. URL https://arxiv.org/abs/2603.18161.

[2] Sina Alemohammad, Josue Casco-Rodriguez, Lorenzo Luzi, Ahmed Imtiaz Humayun, Hossein Babaei, Daniel LeJeune, Ali Siahkoohi, and Richard Baraniuk. Self-consuming generative models go MAD. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=ShjMHfmPs0.

[3] Quentin Bertrand, Joey Bose, Alexandre Duplessis, Marco Jiralerspong, and Gauthier Gidel. On the stability of iterative retraining of generative models on their own data. In *The Twelfth International Conference on Learning Representations*, 2024.

[4] Ralph Allan Bradley and Milton Edward Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952.

[5] Gavin Brown, Shlomi Hod, and Iden Kalemaj. Performative prediction in a stateful world. In *International conference on artificial intelligence and statistics*, pages 6045–6061. PMLR, 2022.

[6] Zhongteng Cai, Yaxuan Wang, Yang Liu, and Xueru Zhang. Stabilizing self-consuming diffusion models with latent space filtering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 19844–19852, 2026.

[7] Sarah Dean and Jamie Morgenstern. Preference dynamics under personalized recommendations. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, EC '22, page 795–816, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450391504. doi: 10.1145/3490486.3538346. URL https://doi.org/10.1145/3490486.3538346.

[8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1423. URL https://aclanthology.org/N19-1423/.

[9] Elvis Dohmatob, Yunzhen Feng, Pu Yang, Francois Charton, and Julia Kempe. A tale of tails: Model collapse as a change of scaling laws. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=KVvku47shW.

[10] Anil R Doshi and Oliver P Hauser. Generative ai enhances individual creativity but reduces the collective diversity of novel content. *Science advances*, 10(28):eadn5290, 2024.

[11] Damien Ferbach, Quentin Bertrand, Avishek Joey Bose, and Gauthier Gidel. Self-consuming generative models with curated data provably optimize human preferences, 2024. URL https://arxiv.org/abs/2407.09499.

[12] Shi Fu, Yingjie Wang, Yuzhu Chen, Xinmei Tian, and Dacheng Tao. A theoretical perspective: How to prevent model collapse in self-consuming training loops. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=WttfQGwpES.

[13] Etienne Gauthier, Francis Bach, and Michael I. Jordan. Explaining and preventing alignment collapse in iterative rlhf, 2026. URL https://arxiv.org/abs/2605.04266.

[14] Matthias Gerstgrasser, Rylan Schaeffer, Apratim Dey, Rafael Rafailov, Tomasz Korbak, Henry Sleight, Rajashree Agrawal, John Hughes, Dhruv Bhandarkar Pai, Andrey Gromov, Dan Roberts, Diyi Yang, David L. Donoho, and Sanmi Koyejo. Is model collapse inevitable? Breaking the curse of recursion by accumulating real and synthetic data. In *First Conference on Language Modeling*, 2024. URL https://openreview.net/forum?id=5B2K4LRgmz.

[15] Nate Gillman, Michael Freeman, Daksh Aggarwal, Chia-Hong Hsu, Calvin Luo, Yonglong Tian, and Chen Sun. Self-correcting self-consuming loops for generative model training. In *ICML*, 2024. URL https://openreview.net/forum?id=i0nVanexij.

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. doi: 10.1109/CVPR.2016.90.

[17] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.

[18] Zachary Izzo, Lexing Ying, and James Zou. How to learn when data reacts to your model: performative gradient descent. In *International Conference on Machine Learning*, pages 4641–4650. PMLR, 2021.

[19] Maurice Jakesch, Advait Bhat, Daniel Buschek, Lior Zalmanson, and Mor Naaman. Co-writing with opinionated language models affects users' views. In *Proceedings of the 2023 CHI conference on human factors in computing systems*, pages 1–15, 2023.

[20] Zhuangzhuang Jia, Yijie Wang, Roy Dong, and Grani A Hanasusanto. Distributionally robust performative optimization. *arXiv preprint arXiv:2407.01344*, 2024.

[21] Kun Jin, Xueru Zhang, Mohammad Mahdi Khalili, Parinaz Naghizadeh, and Mingyan Liu. Incentive mechanisms for strategic classification and regression problems. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 760–790, 2022.

[22] Kun Jin, Tian Xie, Yang Liu, and Xueru Zhang. Addressing polarization and unfairness in performative prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 22408–22416, 2026.

[23] Michael P Kim and Juan C Perdomo. Making decisions under outcome performativity. *arXiv preprint arXiv:2210.01745*, 2022.

[24] Alex Krizhevsky. Learning multiple layers of features from tiny images. 2009. URL https://api.semanticscholar.org/CorpusID:18268744.

[25] Qiang Li and Hoi-To Wai. State dependent performative prediction with stochastic approximation. In *International Conference on Artificial Intelligence and Statistics*, pages 3164–3186. PMLR, 2022.

[26] Qiang Li and Hoi-To Wai. Stochastic optimization schemes for performative prediction with nonconvex loss. *Advances in Neural Information Processing Systems*, 37:8673–8697, 2024.

[27] Qiang Li, Michal Yemini, and Hoi-To Wai. Clipped sgd algorithms for performative prediction: Tight bounds for clipping bias and remedies. *arXiv preprint arXiv:2404.10995*, 2024.

[28] Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t.

[29] H. Liu, Qiang Li, and Hoi-To Wai. Two-timescale derivative free optimization for performative prediction with markovian data. *arXiv preprint arXiv:2310.05792*, 2023.

[30] R.D. Luce. *Individual Choice Behavior: A Theoretical Analysis*. Wiley, 1959. URL https://books.google.com/books?id=a80DAQAAIAAJ.

[31] Celestine Mendler-Dünner, Juan Perdomo, Tijana Zrnic, and Moritz Hardt. Stochastic optimization for performative prediction. *Advances in Neural Information Processing Systems*, 33:4929–4939, 2020.

[32] Celestine Mendler-Dünner, Frances Ding, and Yixin Wang. Anticipating performativity by predicting from predictions. *Advances in neural information processing systems*, 35:31171–31185, 2022.

[33] John P Miller, Juan C Perdomo, and Tijana Zrnic. Outside the echo chamber: Optimizing the performative risk. In *International Conference on Machine Learning*, pages 7710–7720. PMLR, 2021.

[34] Mehrnaz Mofakhami, Ioannis Mitliagkas, and Gauthier Gidel. Performative prediction with neural networks. In *International Conference on Artificial Intelligence and Statistics*, pages 11079–11093. PMLR, 2023.

[35] Alex Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models, 2021. URL https://arxiv.org/abs/2102.09672.

[36] Juan Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative prediction. In *International Conference on Machine Learning*, pages 7599–7609. PMLR, 2020.

[37] Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2019.

[38] Mitas Ray, Lillian J Ratliff, Dmitriy Drusvyatskiy, and Maryam Fazel. Decision-dependent risk minimization in geometrically decaying dynamic environments. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 8081–8088, 2022.

[39] Mohamed El Amine Seddik, Suei-Wen Chen, Soufiane Hayou, Pierre Youssef, and Merouane Abdelkader DEBBAH. How bad is training on synthetic data? A statistical analysis of language model collapse. In *First Conference on Language Modeling*, 2024. URL https://openreview.net/forum?id=t3z6UlV09o.

[40] Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson, and Yarin Gal. AI models collapse when trained on recursively generated data. *Nature*, 631(8022):755–759, 2024.

[41] Ananda Theertha Suresh, Andrew Thangaraj, and Aditya Nanda Kishore Khandavally. Rate of model collapse in recursive training, 2024. URL https://arxiv.org/abs/2412.17646.

[42] Rohan Taori and Tatsunori B. Hashimoto. Data feedback loops: Model-driven amplification of dataset biases. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023.

[43] Alexander Tong, Kilian FATRAS, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=CD9Snc73AW. Expert Certification.

[44] Yaxuan Wang, Zhongteng Cai, Yujia Bao, Xueru Zhang, and Yang Liu. Observations and remedies for large language model bias in self-consuming performative loop. *arXiv preprint arXiv:2601.05184*, 2026.

[45] Xiukun Wei and Xueru Zhang. Self-consuming generative models with adversarially curated data, 2025. URL https://arxiv.org/abs/2505.09768.

[46] Xiukun Wei, Min Shi, and Xueru Zhang. Market games for generative models: Equilibria, welfare, and strategic entry. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=T7c5LXuebY.

[47] Xiukun Wei, Yang Zhang, and Xueru Zhang. Stability and diversity of networked self-consuming generative ecosystems. In *The Fortieth Annual Conference on Neural Information Processing Systems*, 2026. URL https://openreview.net/forum?id=Mq8Ixa9IkT.

[48] Tian Xie and Xueru Zhang. Automating data annotation under strategic human agents: Risks and potential solutions. *Advances in Neural Information Processing Systems*, 37:127436–127482, 2024.

[49] Tian Xie and Xueru Zhang. Non-linear welfare-aware strategic learning. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, volume 7, pages 1660–1671, 2024.

[50] Tian Xie, Zhiqun Zuo, Mohammad Mahdi Khalili, and Xueru Zhang. Learning under imitative strategic behavior with unforeseeable outcomes. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=82bNZGMNZa.

[51] Tian Xie, Ding Zhu, Jia Liu, Mahdi Khalili, and Xueru Zhang. Sprint: Stochastic performative prediction with variance reduction. *arXiv preprint arXiv:2509.17304*, 2025.

[52] Songkai Xue and Yuekai Sun. Distributionally robust performative prediction. *Advances in Neural Information Processing Systems*, 37:55030–55052, 2024.

[53] Bingji Yi, Qiyuan Liu, Yuwei Cheng, and Haifeng Xu. Escaping model collapse via synthetic data verification: Near-term improvements and long-term convergence. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=yfk6c39omW.

[54] Xiang Zhang, Junbo Zhao, and Yann LeCun. Character-level convolutional networks for text classification, 2016. URL https://arxiv.org/abs/1509.01626.

[55] Xueru Zhang, Mohammadmahdi Khaliligarekani, Cem Tekin, et al. Group retention when using machine learning in sequential decision making: the interplay between user dynamics and fairness. *Advances in neural information processing systems*, 32, 2019.

[56] Xueru Zhang, Ruibo Tu, Yang Liu, Mingyan Liu, Hedvig Kjellstrom, Kun Zhang, and Cheng Zhang. How do fair decisions fare in long-term qualification? *Advances in neural information processing systems*, 33: 18457–18469, 2020.

[57] Xueru Zhang, Mohammad Mahdi Khalili, Kun Jin, Parinaz Naghizadeh, and Mingyan Liu. Fairness interventions as (dis) incentives for strategic manipulation. In *International Conference on Machine Learning*, pages 26239–26264. PMLR, 2022.

[58] Yang Zhang, Xiukun Wei, and Xueru Zhang. When and how human curation backfires: Preference alignment under multi-model self-consuming loop, 2026. URL https://arxiv.org/abs/2605.29267.

[59] Hongru Zhao, Jinwen Fu, and Tuan Pham. Convergence and stability analysis of self-consuming generative models with heterogeneous human curation, 2025. URL https://arxiv.org/abs/2511.09002.

[60] Yulai Zhao. Optimizing the performative risk under weak convexity assumptions. *arXiv preprint arXiv:2209.00771*, 2022.

[61] Xue Zheng, Tian Xie, Xuwei Tan, Aylin Yener, and Xueru Zhang. Profl: Performative robust optimal federated learning. *arXiv preprint arXiv:2410.18075*, 2024.

[62] Zihan Zhu, Ethan Fang, and Zhuoran Yang. Online performative gradient descent for learning nash equilibria in decision-dependent games. *Advances in Neural Information Processing Systems*, 36:47902–47913, 2023.

Appendix Contents

A Related Work **16**

B Complete Instantiation Examples **17**

B.1 Discrete Examples 17

B.2 Continuous Examples 18

C Proofs **19**

C.1 Proof of Explicit Retraining Update 19

C.2 Proof of Propostion 3.1 20

C.3 Proof of Theorem 3.2 23

C.4 Proof of Theorem 3.3 24

C.5 Proof of Corollary 3.5 25

C.6 Proof of Proposition 3.6 26

C.7 Proofs of Lemma 3.7 and Proposition 3.8 29

C.8 Proof of Theorem 3.9 30

C.9 Proof of Proposition 4.1 31

D Extension to Continuous Instance Spaces **33**

D.1 Retraining with Purely Synthetic User-Curated Data 33

D.2 Retraining on Mixed Synthetic and Reference Data 36

E Additional Experimental Results **38**

E.1 Additional CFM Experiments on CIFAR-10 39

E.2 DDPM Experiments on CIFAR-10 41

E.3 GPT-2 Experiments on AG News 43

F Discussion and Extensions **44**

F.1 Nonlinear Preference Responses 44

F.2 Stochastic Preference Adaptation 45

F.3 Heterogeneous User Populations 46

G Broader impacts **46**

A Related Work

Self-consuming When generative models are repeatedly trained on their own outputs, the training distribution can drift from the original data distribution. Early empirical studies demonstrated that this loop leads to progressive loss of distributional coverage and tail information [2, 14, 40], with Taori and Hashimoto [42], Xie and Zhang [48] showing that feedback loops can amplify existing dataset biases or even introduce complex feature distribution shifts. Subsequent theoretical work characterized the conditions under which collapse occurs and the rate at which errors accumulate [3, 9, 12, 39, 41]. Several mitigation strategies have been proposed, including retaining real data in the training loop [2, 3, 14], filtering generated samples via verification or latent-space methods [6, 53], and incorporating self-correction during generation [15]. Recent work has further broadened the study of generative feedback from isolated single-model loops to settings with multiple interacting models, networked generative ecosystems, and strategic interactions among generative platforms [46, 47, 58]. Our work builds on this line by studying a setting where the training loop is further shaped by preference-guided curation and where the preference signal itself co-evolves with the model.

Self-consuming with curated data A growing body of work has studied self-consuming retraining with human curation. Ferbach et al. [11] show that, under a fixed reward signal, human curation acts as an implicit preference-optimization mechanism. Wei and Zhang [45] further show that when the same curation process is noisy or adversarially manipulated, the retraining loop can become systematically misaligned from genuine user preferences. Work on multi-model self-consuming systems shows that even benign human curation need not improve alignment once models interact through recycled data: its effect can be amplified, dampened, or even reversed by cross-model coupling [58]. More recently, Zhao et al. [59] studied self-consuming generative models with heterogeneous human curation, proving convergence and stability results across pure synthetic and reference-mixed retraining regimes. Related iterative alignment Gauthier et al. [13] studies the co-evolution of a policy and a learned reward-model estimate around a fixed underlying human preference. Their Stackelberg decomposition characterizes how gradient-based policy optimization can corrupt the reward-model estimate, and their FPO intervention regularizes the policy update to account for this influence. However, these formulations still treat the underlying human preference as exogenous. In many deployed settings, model outputs can reshape what users read, compare, and upload, thereby altering the preferences that guide future curation. Our work studies this endogenous feedback loop by allowing model retraining and user preferences to co-evolve. Tbl. 1 summarizes the key differences.

Performative prediction. Performative prediction (PP) studies learning problems where the deployed model changes the data distribution it will later face [36]. This feedback effect appears in many human-facing ML systems: applicants may strategically modify their features, users may change their participation or engagement behavior, and adversaries may adapt to deployed classifiers. A central difficulty in PP is that the learner is no longer optimizing over a fixed distribution. Previous works mainly focused on the optimization in discriminative PP settings. Several works develop algorithms and convergence guarantees for finding performatively stable solutions. Perdomo et al. [36] introduced repeated risk minimization and repeated gradient descent under sensitivity and convexity assumptions. Subsequent work studied stochastic and online variants, including greedy and lazy deployment schemes [31, 32]. State-dependent performative prediction further generalizes the framework by allowing the next distribution to depend not only on the deployed model but also on the current population state [5, 25]. More recent work has moved beyond the strongly convex setting. For example, Mofakhami et al. [34] considered neural-network-based PP under structural assumptions, and Li and Wai [26] established convergence to stationary performatively stable points for smooth non-convex losses, though with variance-dependent error terms. Another line of work directly optimizes the performative risk or seeks performatively optimal solutions under additional assumptions on the distribution map [18, 29, 33, 38, 61]. Others study PP under weaker convexity assumptions, distributional robustness, privacy, or welfare-oriented objectives [20, 22, 23, 27, 51, 52, 60, 62]. These works show that the choice of solution concept is not only technical: it determines whether the learner aims for equilibrium-like stability, direct performative optimality, or robustness to uncertainty in the distributional response.

A related perspective on endogenous user responses comes from preference dynamics under personalized recommendations [7, 55] and strategic behavior [21, 49, 50, 56, 57]. Dean and Morgenstern [7] study how users' preferences evolve through interaction with a recommendation platform. Their setting is closely related on the preference side, but the platform selects from a fixed catalog without generative retraining on user-curated outputs. Our framework additionally closes the self-consuming training loop: model outputs shape user preferences, the evolving preferences guide curation, and the retained outputs are used to retrain the generator. The resulting dynamics involve both the generative distribution and the underlying preference, bringing curation-induced generator collapse and reference-data mixing into the analysis of preference evolution.

B Complete Instantiation Examples

This appendix provides concrete examples that instantiate every symbol in our framework across several domains. We give two discrete examples (Appx. B.1) and one continuous examples (Appx. B.2).

B.1 Discrete Examples

Example A: Image Generation with Object Classes Consider an image generation platform with $n = 10$ object categories.

- **Sample space:** $\mathcal{X} = \{x_1, \ldots, x_{10}\}$, representing the types: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, truck.
- **Distribution space:** $\mathcal{P}(\mathcal{X})$ is the probability simplex over 10 types.
- **Feature map:** $\varphi : \mathcal{X} \to \mathbb{R}^{10}$ is the softmax output of a pretrained classifier, normalized to $\|\varphi(x_i)\| = 1$.

Concrete feature vectors. After normalization, the feature vectors for three representative types are:

$$\begin{aligned}
\varphi(x_{\text{airplane}}) &= \frac{(0.95,\ 0.02,\ 0.01,\ 0.00,\ 0.00,\ 0.00,\ 0.00,\ 0.01,\ 0.01,\ 0.00)}{\|\cdot\|} \\
&\approx (0.999,\ 0.021,\ 0.011,\ 0.000,\ 0.000,\ 0.000,\ 0.000,\ 0.011,\ 0.011,\ 0.000), \\
\varphi(x_{\text{cat}}) &= \frac{(0.01,\ 0.01,\ 0.03,\ 0.90,\ 0.01,\ 0.02,\ 0.00,\ 0.01,\ 0.00,\ 0.01)}{\|\cdot\|} \\
&\approx (0.011,\ 0.011,\ 0.033,\ 0.998,\ 0.011,\ 0.022,\ 0.000,\ 0.011,\ 0.000,\ 0.011), \\
\varphi(x_{\text{frog}}) &= \frac{(0.00,\ 0.00,\ 0.02,\ 0.01,\ 0.02,\ 0.01,\ 0.92,\ 0.01,\ 0.00,\ 0.01)}{\|\cdot\|} \\
&\approx (0.000,\ 0.000,\ 0.022,\ 0.011,\ 0.022,\ 0.011,\ 0.999,\ 0.011,\ 0.000,\ 0.011).
\end{aligned}$$

Each vector is dominated by one coordinate, reflecting high classifier confidence.

Initial distribution. $p_0 = (0.10,\ 0.10,\ \ldots,\ 0.10)$ (uniform across all types).

User preference. Suppose the user prefers airplane and ship content. The initial preference vector is

$$w_0 = \frac{(0.6,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0.4,\ 0)}{\|(0.6,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0.4,\ 0)\|} \approx (0.832,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0.555,\ 0),$$

with $\|w_0\| = 1$.

Reward computation. The reward $r_0(x) = \langle w_0, \varphi(x) \rangle$ at round $t = 0$:

$$\begin{aligned}
r_0(x_{\text{airplane}}) &= 0.832 \times 0.999 + 0.555 \times 0.011 + \cdots \approx 0.837, \\
r_0(x_{\text{cat}}) &= 0.832 \times 0.011 + 0.555 \times 0.000 + \cdots \approx 0.009, \\
r_0(x_{\text{frog}}) &= 0.832 \times 0.000 + 0.555 \times 0.000 + \cdots \approx 0.000.
\end{aligned}$$

Airplane receives reward ≈ 0.84, while cat and frog receive near-zero reward. With curation strength $\tau > 0$, the K-way Luce rule strongly favors airplane. Over successive rounds, p_t concentrates on airplane, w_t drifts further toward the airplane direction, and frog and horse vanish (Thm. 3.3).

Protection. To prevent frog and horse from disappearing, the designer sets $L = 2$:

$$\begin{aligned}
v_1 &= (\underbrace{0}_{0}, \underbrace{0}_{1}, \underbrace{0}_{2}, \underbrace{0}_{43}, \underbrace{0}_{4}, \underbrace{0}_{5}, \underbrace{1}_{6}, \underbrace{0}_{7}, \underbrace{0}_{8}, \underbrace{0}_{9}), \quad \theta_1 = 0.05 \quad (\text{frog} \geq 5\%), \\
v_2 &= (\underbrace{0}_{0}, \underbrace{0}_{1}, \underbrace{0}_{2}, \underbrace{0}_{6}, \underbrace{0}_{4}, \underbrace{0}_{5}, \underbrace{0}_{6}, \underbrace{1}_{7}, \underbrace{0}_{8}, \underbrace{0}_{9}), \quad \theta_2 = 0.05 \quad (\text{horse} \geq 5\%).
\end{aligned}$$

The constraint $\langle v_1, \bar{\varphi}(p^\star) \rangle \geq 0.05$ ensures that frog-type images constitute at least 5% of the equilibrium output in the feature-weighted sense.

Cost. If real-world images are equally easy to collect across all types, we set $c_i = 1$ uniformly. In practice, some types are naturally rarer. For instance, high-quality frog images may be scarce compared to airplane or cat images, making them more expensive to collect. Setting $c_{\text{frog}} = 5$ reflects this scarcity.

Example B: Text Generation with Writing Style Consider a text generation platform where outputs are categorized by writing style.

- **Sample space:** $\mathcal{X} = \{x_1, \ldots, x_5\}$, representing five style types: formal, casual, poetic, technical, humorous.
- **Distribution space:** $\mathcal{P}(\mathcal{X}) = \Delta_5$.

- **Feature map:** $\varphi : \mathcal{X} \to \mathbb{R}^5$ is the softmax output of a fine-tuned style classifier, normalized to $\|\varphi(x_i)\| = 1$.

Concrete feature vectors.

$$\varphi(x_{\text{casual}}) \approx (0.033,\ 0.997,\ 0.033,\ 0.022,\ 0.055)/\|\cdot\|,$$
$$\varphi(x_{\text{poetic}}) \approx (0.055,\ 0.033,\ 0.993,\ 0.022,\ 0.055)/\|\cdot\|,$$
$$\varphi(x_{\text{technical}}) \approx (0.044,\ 0.022,\ 0.011,\ 0.995,\ 0.011)/\|\cdot\|.$$

User preference and reward. Most users prefer casual content:

$$w_0 = \frac{(0.1,\ 0.8,\ 0.0,\ 0.0,\ 0.1)}{\|(0.1,\ 0.8,\ 0.0,\ 0.0,\ 0.1)\|} \approx (0.124,\ 0.990,\ 0.000,\ 0.000,\ 0.124).$$

Then:

$$r_0(x_{\text{casual}}) = \langle w_0, \varphi(x_{\text{casual}})\rangle \approx 0.99,$$
$$r_0(x_{\text{poetic}}) = \langle w_0, \varphi(x_{\text{poetic}})\rangle \approx 0.04,$$
$$r_0(x_{\text{technical}}) = \langle w_0, \varphi(x_{\text{technical}})\rangle \approx 0.02.$$

Casual text dominates curation. Over rounds, poetic and technical writing vanish.

Protection. Preserve poetic and technical writing:

$$v_1 = (0,\ 0,\ 1,\ 0,\ 0), \quad \theta_1 = 0.08 \quad (\text{poetic} \geq 8\%),$$
$$v_2 = (0,\ 0,\ 0,\ 1,\ 0), \quad \theta_2 = 0.05 \quad (\text{technical} \geq 5\%).$$

Cost. $c_{\text{poetic}} = 3$ (high-quality poetic text requires expert curation), $c_{\text{casual}} = 1$ (abundant online).

B.2 Continuous Examples

Example C: Image Generation with CLIP Embedding Consider an image generation platform where outputs live in a continuous space.

- **Sample space:** $\mathcal{X} \subset \mathbb{R}^m$ is a compact subset of image space.
- **Distribution space:** $\mathcal{P}(\mathcal{X})$ is the set of all Borel probability measures on $\mathcal{X}$.
- **Feature map:** $\varphi : \mathcal{X} \to \mathbb{R}^{512}$ is a pretrained CLIP image encoder, with $\|\varphi(x)\| = 1$. Feature dimension $d = 512$. Semantically similar images have nearby embeddings.

Concrete feature vectors. Three representative images and their (truncated) CLIP embeddings:

$$\varphi(x_{\text{sunset}}) = (0.12,\ -0.08,\ 0.43,\ \ldots)_{512}, \quad \text{(vibrant, warm tones)}$$
$$\varphi(x_{\text{portrait}}) = (-0.05,\ 0.31,\ 0.15,\ \ldots)_{512}, \quad \text{(human subject, neutral)}$$
$$\varphi(x_{\text{mono}}) = (-0.22,\ -0.14,\ -0.35,\ \ldots)_{512}, \quad \text{(dark, monochrome)}$$

each normalized to $\|\varphi\| = 1$.

User preference and reward. Users consistently select vibrant, colorful images. After 5 rounds, the preference vector has drifted to

$$w_5 = (0.15,\ -0.06,\ 0.38,\ \ldots)_{512}, \quad \|w_5\| = 0.73.$$

The reward for each image:

$$r_5(x_{\text{sunset}}) = \langle w_5, \varphi(x_{\text{sunset}})\rangle \approx 0.71,$$
$$r_5(x_{\text{portrait}}) = \langle w_5, \varphi(x_{\text{portrait}})\rangle \approx 0.18,$$
$$r_5(x_{\text{mono}}) = \langle w_5, \varphi(x_{\text{mono}})\rangle \approx -0.41.$$

Sunset images receive high reward; monochrome images receive negative reward. Curation suppresses monochrome content, p_t concentrates on vibrant images, w_t drifts further from the monochrome direction.

Mean feature vector. The equilibrium mean feature vector is

$$\bar{\varphi}(p^\star) = \int_{\mathcal{X}} \varphi(x)\, p^\star(dx) \in \mathbb{R}^{512}.$$

As p_t concentrates on vibrant images, $\bar{\varphi}(p_t)$ moves toward the vibrant cluster in CLIP space and away from the monochrome direction.

Protection. To preserve monochrome imagery, the designer collects 500 reference monochrome images and computes their mean CLIP embedding:

$$v_1 = \frac{\frac{1}{500}\sum_{j=1}^{500} \varphi(x_j^{\text{mono}})}{\left\|\frac{1}{500}\sum_{j=1}^{500} \varphi(x_j^{\text{mono}})\right\|} \in \mathbb{R}^{512}, \qquad \theta_1 = 0.05.$$

The constraint $\langle v_1, \bar{\varphi}(p^\star)\rangle \geq 0.05$ ensures the equilibrium retains at least 0.05 projection onto the monochrome direction in CLIP space.

Cost. $c(x)$ varies across CLIP space: reference images in the monochrome region are expensive to source (niche artistic photography), while vibrant natural images are cheap (abundant on the web).

C Proofs

C.1 Proof of Explicit Retraining Update

Lemma C.1 (Explicit retraining update)**.** *Let p_{t+1} be defined by (2). Then,*

$$p_{t+1}(x) = (1-\eta)p_{\mathrm{ref}}(x) + \eta p_t(x) H^K_{p_t,w_t}(x), \tag{9}$$

where

$$H^K_{p_t,w_t}(x) := \mathbb{E}_{x_1,\ldots,x_{K-1}\sim p_t}\left[\frac{K e^{\tau\cdot r_{w_t}(x)}}{e^{\tau\cdot r_{w_t}(x)} + \sum_{i=1}^{K-1} e^{\tau\cdot r_{w_t}(x_i)}}\right].$$

Moreover, for every fixed $x \in \mathcal{X}$,

$$H^K_{p_t,w_t}(x) \xrightarrow{K\to\infty} \frac{e^{\tau r_t(x)}}{\mathbb{E}_{z\sim p_t}\left[e^{\tau r_t(z)}\right]}.$$

Proof. First, by minimization of the cross-entropy, we know that for any distribution $q \in \mathcal{P}(\mathcal{X})$, $\arg\max_{p\in\mathcal{P}(\mathcal{X})} \mathbb{E}_{x\sim q}[\log p(x)] = q$

Case 1: $\eta = 1$. When $\eta = 1$, the objective in (2) reduces to

$$p_{t+1} = \arg\max_{p\in\mathcal{P}(\mathcal{X})} \mathbb{E}_{\substack{x_1,\ldots,x_K\sim p_t \\ \hat{x}\sim \mathrm{Luce}_K(x_1,\ldots,x_K)}}[\log p(\hat{x})]$$

Let p_t^{cur} denote the distribution of the retained sample $\hat{x}$. By the population MLE argument above, $p_{t+1} = p_t^{\mathrm{cur}}$

To characterize p_t^{cur}, fix round t and sample$X_1,\ldots,X_K \overset{\text{i.i.d.}}{\sim} p_t$. Let $i_K \in [K]$ be selected according to

$$\Pr(i_K = i \mid x_1,\ldots,x_K) = \frac{e^{\tau r_t(x_i)}}{\sum_{j=1}^K e^{\tau r_t(x_j)}}.$$

Define the retained sample by $\hat{X} := X_{i_K}$. Then p_t^{cur} is the law of $\hat{X}$. By symmetry of the K candidates,

$$\begin{aligned}
p_{t+1}(x) = p_t^{\mathrm{cur}}(x) &= \sum_{i=1}^K \int_{y_j,\ j\neq i} p_t(y_1,\ldots,y_{i-1},x,y_{i+1},\ldots,y_K)\ \Pr(i_K = i \mid x, y_j,\ j\neq i) \prod_{j\neq i} dy_j \\
&= K \int_{y_1,\ldots,y_{K-1}} p_t(y_1,\ldots,y_{K-1},x)\ \Pr(i_K = K \mid y_1,\ldots,y_{K-1},x)\, dy_1\cdots dy_{K-1} \\
&= p_t(x)\, K \int_{y_1,\ldots,y_{K-1}} \frac{e^{\tau r_t(x)}}{e^{\tau r_t(x)} + \sum_{i=1}^{K-1} e^{\tau r_t(y_i)}} \prod_{i=1}^{K-1} p_t(y_i)\, dy_1\cdots dy_{K-1}.
\end{aligned}$$

Therefore,

$$p_{t+1}(x) = p_t(x)\, H^K_{p_t,w_t}(x),$$

where

$$\begin{aligned}
H^K_{p_t,w_t}(x) &:= K \int_{y_1,\ldots,y_{K-1}} \frac{e^{\tau r_t(x)}}{e^{\tau r_t(x)} + \sum_{i=1}^{K-1} e^{\tau r_t(y_i)}} \prod_{i=1}^{K-1} p_t(y_i)\, dy_1\cdots dy_{K-1} \\
&= \mathbb{E}_{X_1,\ldots,X_{K-1}\overset{\text{i.i.d.}}{\sim} p_t}\left[\frac{K e^{\tau r_t(x)}}{e^{\tau r_t(x)} + \sum_{i=1}^{K-1} e^{\tau r_t(X_i)}}\right].
\end{aligned}$$

Now consider the large-K limit. Since $r_t(x) \in [-1,1]$, we have

$$e^{-\tau} \leq e^{\tau r_t(x)} \leq e^{\tau} \qquad \text{for all } x \in \mathcal{X}.$$

Rewrite the denominator as

$$\frac{e^{\tau r_t(x)}}{K} + \frac{K-1}{K}\cdot\frac{1}{K-1}\sum_{i=1}^{K-1} e^{\tau r_t(X_i)}.$$

By the strong law of large numbers,

$$\frac{1}{K-1}\sum_{i=1}^{K-1} e^{\tau r_t(X_i)} \xrightarrow{\text{a.s.}} \mathbb{E}_{z\sim p_t}[e^{\tau r_t(z)}].$$

Hence, for every fixed $x \in \mathcal{X}$,

$$\frac{e^{\tau r_t(x)}}{\frac{e^{\tau r_t(x)}}{K} + \frac{K-1}{K}\cdot\frac{1}{K-1}\sum_{i=1}^{K-1} e^{\tau r_t(X_i)}} \xrightarrow{\text{a.s.}} \frac{e^{\tau r_t(x)}}{\mathbb{E}_{z\sim p_t}[e^{\tau r_t(z)}]}.$$

Therefore,

$$H^K_{p_t,w_t}(x) \xrightarrow{K\to\infty} \frac{e^{\tau r_t(x)}}{\mathbb{E}_{z\sim p_t}[e^{\tau r_t(z)}]}.$$

Consequently, in the large-K limit,

$$p_{t+1}(x) \xrightarrow{K\to\infty} p_t(x)\frac{e^{\tau r_t(x)}}{\mathbb{E}_{z\sim p_t}[e^{\tau r_t(z)}]}$$

Moreover, for every $K \geq 2$,

$$0 \leq H^K_{p_t,w_t}(x) \leq \frac{Ke^{\tau}}{e^{-\tau}+(K-1)e^{-\tau}} \leq 2e^{2\tau}.$$

Hence the family $\{H^K_{p_t,w_t}(x)\}_{K\geq 2}$ is uniformly bounded. By dominated convergence,

$$\int_{\mathcal{X}} \left| H^K_{p_t,w_t}(x) - \frac{e^{\tau r_t(x)}}{\mathbb{E}_{z\sim p_t}[e^{\tau r_t(z)}]} \right| p_t(dx) \to 0.$$

This proves the total-variation convergence of the curated distribution to its large-K exponential-reweighting limit.

Case 2: $\eta < 1$. When $\eta < 1$, the objective in (2) can be written as

$$(1-\eta)\mathbb{E}_{x\sim p_{\text{ref}}}[\log p(x)] + \eta\mathbb{E}_{x\sim p_t^{\text{cur}}}[\log p(x)] = \mathbb{E}_{x\sim q_t}[\log p(x)],$$

where

$$q_t := (1-\eta)p_{\text{ref}} + \eta\, p_t^{\text{cur}}.$$

Applying the same population MLE argument yields

$$p_{t+1} = q_t = (1-\eta)p_{\text{ref}} + \eta\, p_t^{\text{cur}}.$$

For finite K, since

$$p_t^{\text{cur}}(x) = p_t(x)\, H^K_{p_t,w_t}(x),$$

we obtain

$$\begin{aligned} p_{t+1}(x) &= (1-\eta)p_{\text{ref}}(x) + \eta\, p_t^{\text{cur}}(x) \\ &= (1-\eta)p_{\text{ref}}(x) + \eta\, p_t(x)\, H^K_{p_t,w_t}(x) \\ &= (1-\eta)p_{\text{ref}}(x) + \eta\, p_t(x)\, \mathbb{E}_{X_1,\ldots,X_{K-1}\overset{\text{i.i.d.}}{\sim} p_t}\left[\frac{Ke^{\tau r_t(x)}}{e^{\tau r_t(x)} + \sum_{i=1}^{K-1} e^{\tau r_t(X_i)}}\right]. \end{aligned}$$

As $K \to \infty$,

$$H^K_{p_t,w_t}(x) \to \frac{e^{\tau r_t(x)}}{\mathbb{E}_{z\sim p_t}[e^{\tau r_t(z)}]},$$

and therefore

$$p_{t+1}(x) \xrightarrow{K\to\infty} (1-\eta)p_{\text{ref}}(x) + \eta\, p_t(x)\frac{e^{\tau r_t(x)}}{\mathbb{E}_{z\sim p_t}[e^{\tau r_t(z)}]}.$$

□

C.2 Proof of Propostion 3.1

Proposition 3.1 (Basic properties of curation operator)**.** *For any $p \in \mathcal{P}(\mathcal{X})$, $w \in \mathcal{W}$, and $K \geq 2$, the curation operator $H^K_{p,w}$ satisfies the following properties:*

1. ***Strict monotonicity.*** *For any $x_i, x_j \in \mathcal{X}$, $r_w(x_i) > r_w(x_j) \iff H^K_{p,w}(x_i) > H^K_{p,w}(x_j)$*

2. ***Normalization.*** $\sum_{i=1}^{n} p(x_i) H_{p,w}^K(x_i) = 1$.
3. ***Odds contraction.*** *If* $r_w(x_i) - r_w(x_j) \geq r_\Delta > 0$, *then*

$$\frac{H_{p,w}^K(x_j)}{H_{p,w}^K(x_i)} \leq \lambda_K(r_\Delta) := 1 - \frac{(1 - e^{-\tau \cdot r_\Delta})(K-1)}{e^{2\tau} + K - 1},$$

with $\lambda_K(r_\Delta) \in (0,1)$ *and* $\lambda_K(r_\Delta) \xrightarrow{K \to \infty} e^{-\tau \cdot r_\Delta}$

Proof. Write

$$a(x) := e^{\tau r_w(x)}$$

For a fixed $X_1, \ldots, X_{K-1} \overset{\text{iid}}{\sim} p$, let

$$S := \sum_{l=1}^{K-1} a(X_l) \geq 0$$

Since $a(X_\ell) > 0$ and $K \geq 2$, we have $S > 0$

So

$$H_{p,w}^K(x) = \mathbb{E}_{X_1,\ldots,X_{K-1} \overset{\text{i.i.d.}}{\sim} p} \left[\frac{K e^{\tau r_w(x)}}{e^{\tau r_w(x)} + \sum_{i=1}^{K-1} e^{\tau r_w(X_i)}} \right] = \mathbb{E}_{X_1,\ldots,X_{K-1} \overset{\text{i.i.d.}}{\sim} p} \frac{K a(x)}{a(x) + S}$$

We prove the three claims in order.

Part 1: strict monotonicity. Define

$$f_S(a) := \frac{Ka}{a + S}$$

Then

$$f_S'(a) = \frac{K(a+S) - Ka}{(a+S)^2} = \frac{KS}{(a+S)^2} > 0$$

Hence $f_S(a)$ is strictly increasing in $a \in (0, \infty)$.

On the other hand,

$$a'(x) = \frac{d}{dr_w(x)} e^{\tau r_w(x)} = \tau e^{\tau r_w(x)} > 0$$

so $a(x)$ is strictly increasing in $r_w(x)$.

Therefore, for every fixed sample path,

$$r_w(x_i) > r_w(x_j) \quad \Longrightarrow \quad a(x_i) > a(x_j) \quad \Longrightarrow \quad f_S(a(x_i)) > f_S(a(x_j))$$

That is,

$$r_w(x_i) > r_w(x_j) \quad \Longrightarrow \quad \frac{K e^{\tau r_w(x_i)}}{e^{\tau r_w(x_i)} + S} > \frac{K e^{\tau r_w(x_j)}}{e^{\tau r_w(x_j)} + S}$$

Taking expectation over $X_1, \ldots, X_{K-1}$ yields

$$r_w(x_i) > r_w(x_j) \quad \Longrightarrow \quad H_{p,w}^K(x_i) > H_{p,w}^K(x_j)$$

Conversely, since $H_{p,w}^K(x)$ is a strictly increasing function of $r_w(x)$, it also reflects strict order, and hence

$$r_w(x_i) > r_w(x_j) \quad \Longleftrightarrow \quad H_{p,w}^K(x_i) > H_{p,w}^K(x_j)$$

For $K = \infty$, we have

$$H_{p,w}^\infty(x) = \frac{e^{\tau r_w(x)}}{\mathbb{E}_{z \sim p}[e^{\tau r_w(z)}]}$$

where $\mathbb{E}_{z \sim p}[e^{\tau r_w(z)}]$ is independent of x. Thus strict monotonicity is immediate.

Part 2: normalization. For sample$X_1, \ldots, X_K \overset{\text{i.i.d.}}{\sim} p_t$, let $i_K \in [K]$ be selected according to

$$\Pr(i_K = i \mid x_1, \ldots, x_K) = \frac{e^{\tau r_t(x_i)}}{\sum_{j=1}^K e^{\tau r_t(x_j)}}$$

Define the retained sample by $\hat{X} := X_{i_K}$. By the derivation in Lemma C.1, the law of $\hat{X}$ is precisely

$$\Pr(\hat{X} = x_i) = p(x_i) H_{p,w}^K(x_i)$$

Therefore,

$$\sum_{i=1}^{n} p(x_i) H_{p,w}^{K}(x_i) = \sum_{i=1}^{n} \Pr(\hat{X} = x_i) = 1$$

For $K = \infty$, we have

$$H_{p,w}^{\infty}(x) = \frac{e^{\tau r_w(x)}}{\mathbb{E}_{z\sim p}[e^{\tau r_w(z)}]}$$

so

$$\sum_{i=1}^{n} p(x_i) H_{p,w}^{\infty}(x_i) = \frac{\sum_{i=1}^{n} p(x_i) e^{\tau r_w(x_i)})}{\mathbb{E}_{z\sim p}[e^{\tau r_w(z)}]} = 1$$

Part 3: explicit odds contraction. Write

$$a(x_i) := e^{\tau r_w(x_i)} \qquad a(x_j) := e^{\tau r_w(x_j)}$$

If $r_w(x_i) - r_w(x_j) \geq r_\Delta > 0$, then

$$a(x_j) \leq e^{-\tau r_\Delta} a(x_i)$$

and therefore

$$a(x_i) - a(x_j) \geq (1 - e^{-\tau r_\Delta}) a(x_i)$$

Let

$$S := \sum_{l=1}^{K-1} a(X_l) \geq 0$$

For any $S \geq 0$,

$$\frac{a(x_j)/(a(x_j)+S)}{a(x_i)/(a(x_i)+S)} = \frac{a(x_j)(a(x_i)+S)}{a(x_i)(a(x_j)+S)} = 1 - \frac{S \cdot (a(x_i) - a(x_j))}{a(x_i)(a(x_j)+S)}$$

So,

$$\begin{aligned}
\frac{H_{p,w}^{K}(x_j)}{H_{p,w}^{K}(x_i)} &= \frac{\mathbb{E}_{X_1,\ldots,X_{K-1} \overset{\text{i.i.d.}}{\sim} p} \left[\frac{K e^{\tau r_w(x_j)}}{e^{\tau r_w(x_j)} + \sum_{i=1}^{K-1} e^{\tau r_w(X_i)}} \right]}{\mathbb{E}_{X_1,\ldots,X_{K-1} \overset{\text{i.i.d.}}{\sim} p} \left[\frac{K e^{\tau r_w(x_i)}}{e^{\tau r_w(x_i)} + \sum_{i=1}^{K-1} e^{\tau r_w(X_i)}} \right]} \\
&= 1 - \mathbb{E}_{X_1,\ldots,X_{K-1} \overset{\text{i.i.d.}}{\sim} p} \frac{S \cdot (a(x_i) - a(x_j))}{a(x_i)(a(x_j)+S)}
\end{aligned} \tag{10}$$

Since $r_w(x) \in [-1, 1]$, each $a(x) \in [e^{-\tau}, e^{\tau}]$.

Using $a(x_i) - a(x_j) \geq (1 - e^{-\tau r_\Delta}) a(x_i)$ and $a(x_j) \leq e^{\tau}$, we obtain

$$\frac{S \cdot (a(x_i) - a(x_j))}{a(x_i)(a(x_j)+S)} \geq (1 - e^{-\tau r_\Delta}) \frac{S}{e^{\tau} + S}$$

Since $a(X_\ell) \geq e^{-\tau}$ so

$$S \geq (K-1) e^{-\tau}$$

Because the map $s \mapsto \frac{s}{(e^{\tau}+s)}$ is increasing,

$$\mathbb{E}_{X_1,\ldots,X_{K-1} \overset{\text{i.i.d.}}{\sim} p} \frac{S}{e^{\tau} + S} \geq \frac{(K-1)e^{-\tau}}{e^{\tau} + (K-1)e^{-\tau}} = \frac{K-1}{e^{2\tau} + K - 1}$$

Substituting this into (10) yields

$$\frac{H_{p,w}^{K}(x_j)}{H_{p,w}^{K}(x_i)} \leq 1 - (1 - e^{-\tau r_\Delta}) \frac{K-1}{e^{2\tau} + K - 1} =: \lambda_K(r_\Delta)$$

For $K = \infty$, the claim is immediate:

$$\frac{H_{p,w}^{\infty}(x_j)}{H_{p,w}^{\infty}(x_i)} = \frac{\frac{e^{\tau r_w(x_j)}}{\mathbb{E}_{z\sim p}[e^{\tau r_w(z)}]}}{\frac{e^{\tau r_w(x_i)}}{\mathbb{E}_{z\sim p}[e^{\tau r_w(z)}]}} = \frac{a(x_j)}{a(x_i)} \leq e^{-\tau r_\Delta}$$

Finally,

$$\lambda_K(r_\Delta) \xrightarrow{K\to\infty} e^{-\tau r_\Delta}$$

And for $K \geq 2$, $r_\Delta > 0$,

$$0 < \lambda_K(r_\Delta) < 1.$$

This completes the proof. □

C.3 Proof of Theorem 3.2

Theorem 3.2 (Equilibrium characterization under purely synthetic setting)**.** *When $\eta = 1$, a distribution $p \in \mathcal{P}(\mathcal{X})$ is a fixed point of the retraining step under preference w if and only if all instances in its support share the same reward:*

$$p(x) > 0,\ p(x') > 0 \implies r_w(x) = r_w(x').$$

Consequently, for every instance $x_i \in \mathcal{X}$, the pair $(p^\star = \delta_{x_i}, w^\star = \varphi(x_i))$ is an equilibrium: the model has collapsed to a point mass on x_i while user preferences have locked onto its features.

Proof. We prove the two claims in turn.

Part 1: Data fixed-point characterization. *(⇒)* Suppose p is a fixed point of the data update at preference w, i.e., for all $x \in \mathcal{X}$,

$$p(x)H^K_{p,w}(x) = p(x)$$

Then for every x with $p(x) > 0$,

$$H^K_{p,w}(x) = 1$$

Now let $x, x' \in \mathcal{X}$ satisfy $p(x) > 0$ and $p(x') > 0$. If

$$r_w(x) \neq r_w(x')$$

then by Proposition 3.1 (strict monotonicity),

$$r_w(x) > r_w(x') \quad \implies \quad H^K_{p,w}(x) > H^K_{p,w}(x')$$

or vice versa.

But both x and x' lie in the support of p, so we already showed

$$H^K_{p,w}(x) = H^K_{p,w}(x') = 1$$

a contradiction. Therefore,

$$p(x) > 0,\ p(x') > 0 \quad \implies \quad r_w(x) = r_w(x')$$

(⇐) Conversely, suppose that for all $x, x' \in \mathcal{X}$.

$$p(x) > 0,\ p(x') > 0 \quad \implies \quad r_w(x) = r_w(x')$$

Let c denote the common reward value on the support of p, i.e., for every x with $p(x) > 0$.

$$r_w(x) = c$$

Then for every such x,

$$e^{\tau r_w(x)} = e^{\tau c}$$

Hence, for any x in the support of p,

$$H^K_{p,w}(x) = \mathbb{E}\left[\frac{Ke^{\tau c}}{e^{\tau c} + \sum_{\ell=1}^{K-1} e^{\tau c}}\right] = 1$$

Therefore,

$$p(x)H^K_{p,w}(x) = p(x)$$

so p is a fixed point of the data update.

For $K = \infty$, the same conclusion holds since

$$H^\infty_{p,w}(x) = \frac{e^{\tau c}}{\mathbb{E}_{z\sim p}[e^{\tau r_w(z)}]} = 1$$

for every x with $p(x) > 0$.

Part 2: Singleton equilibria. Fix $i \in \{1, \ldots, n\}$, and define $p^\star$ by

$$p^\star(x_i) = 1, \qquad p^\star(x_j) = 0 \text{ for all } j \neq i,$$

and let

$$w^\star = \varphi(x_i)$$

Since $p^\star$ is supported on a single point, the condition in Part 1 is trivially satisfied, so $p^\star$ is a fixed point of the data update.

Moreover,

$$\bar{\varphi}(p^\star) = \varphi(x_i) = w^\star$$

Therefore,

$$\mathrm{Proj}_{\mathcal{W}}((1-\beta)w^\star + \beta\bar{\varphi}(p^\star)) = \mathrm{Proj}_{\mathcal{W}}(\varphi(x_i)) = \varphi(x_i) = w^\star$$

where we used that $\|\varphi(x_i)\| = 1$ and $\mathcal{W}$ is the closed unit ball.

Hence $(p^\star, w^\star)$ satisfies both parts of (5), and is therefore an equilibrium of the pure coupled dynamics. □

C.4 Proof of Theorem 3.3

Theorem 3.3 (Local convergence to singleton equilibria). *Fix $x_i \in \mathcal{X}$ and suppose $\varphi(x_j) \neq \varphi(x_i)$ for all $j \neq i$. Define the reward gap $\gamma_i = \min_{j\neq i}\langle \varphi(x_i),\, \varphi(x_i) - \varphi(x_j)\rangle > 0$ and let $\lambda_K(\cdot)$ be the odds-contraction factor from Prop. 3.1. If the initial state satisfies*

$$1 - p_0(x_i) \le \varepsilon_i, \qquad \|w_0 - \varphi(x_i)\| \le \frac{\gamma_i}{4}$$

for some $\varepsilon_i \in \big(0, \min\{\frac{1}{2},\, 1-\lambda_K(\frac{\gamma_i}{2}),\, \frac{\gamma_i}{8}\}\big)$, then for all $t \ge 0$,

$$1 - p_t(x_i) \le \left(\frac{\lambda_K(\frac{\gamma_i}{2})}{1-\varepsilon_i}\right)^t \big(1 - p_0(x_i)\big).$$

Consequently, $p_t(x_i) \xrightarrow{t\to\infty} 1$ and $w_t \xrightarrow{t\to\infty} \varphi(x_i)$.

Proof. Let

$$q_t := 1 - p_t(x_i) = \sum_{j\neq i} p_t(x_j)$$

denote the off-winner mass.

Since $\|\varphi(x)\| = 1$ for all x and $\varphi(x_j) \neq \varphi(x_i)$ for every $j \neq i$, we have

$$\gamma_i = \min_{j\neq i}\langle \varphi(x_i),\, \varphi(x_i) - \varphi(x_j)\rangle = \min_{j\neq i}\big(1 - \langle \varphi(x_i), \varphi(x_j)\rangle\big) > 0$$

Fix K, $\lambda_K(\frac{\gamma_i}{2}) \in (0,1)$, where $\lambda_K(\cdot)$ is the odds-contraction factor from Proposition 3.1.

Choose

$$\varepsilon_i \in \left(0, \min\left\{\frac{1}{2},\, 1-\lambda_K(\frac{\gamma_i}{2}),\, \frac{\gamma_i}{8}\right\}\right),$$

and define

$$U_i := \left\{(p,w) : 1 - p(x_i) \le \varepsilon_i,\ \|w - \varphi(x_i)\| \le \frac{\gamma_i}{4}\right\}.$$

We show that U_i is forward invariant and that q_t decays geometrically.

Uniform reward gap inside U_i. Suppose $(p_t, w_t) \in U_i$. Using $r_{w_t}(x) = \langle w_t, \varphi(x)\rangle$, we have

$$r_{w_t}(x_i) - r_{w_t}(x_j) = \langle w_t, \varphi(x_i) - \varphi(x_j)\rangle$$

Writing

$$w_t = \varphi(x_i) + \big(w_t - \varphi(x_i)\big)$$

we obtain

$$r_{w_t}(x_i) - r_{w_t}(x_j) = \langle \varphi(x_i), \varphi(x_i) - \varphi(x_j)\rangle + \langle w_t - \varphi(x_i), \varphi(x_i) - \varphi(x_j)\rangle$$

for every $j \neq i$.

Hence

$$r_{w_t}(x_i) - r_{w_t}(x_j) \ge \gamma_i - \|w_t - \varphi(x_i)\|\|\varphi(x_i) - \varphi(x_j)\|$$

Since $\|\varphi(x_i) - \varphi(x_j)\| \le 2$ and $\|w_t - \varphi(x_i)\| \le \gamma_i/4$, we obtain

$$r_{w_t}(x_i) - r_{w_t}(x_j) \ge \gamma_i - 2\cdot\frac{\gamma_i}{4} = \frac{\gamma_i}{2}$$

Odds contraction. By Proposition 3.1, for every $j \neq i$,

$$\frac{H^K_{p_t,w_t}(x_j)}{H^K_{p_t,w_t}(x_i)} \le \lambda_K(\frac{\gamma_i}{2})$$

Since the pure data update is

$$p_{t+1}(x) = p_t(x) H^K_{p_t,w_t}(x),$$

we have

$$\frac{p_{t+1}(x_j)}{p_{t+1}(x_i)} = \frac{p_t(x_j)}{p_t(x_i)} \cdot \frac{H^K_{p_t,w_t}(x_j)}{H^K_{p_t,w_t}(x_i)} \le \lambda_K(\frac{\gamma_i}{2})\frac{p_t(x_j)}{p_t(x_i)}$$

Summing over all $j \neq i$ gives

$$\sum_{j\neq i}\frac{p_{t+1}(x_j)}{p_{t+1}(x_i)} \le \sum_{j\neq i}\lambda_K(\frac{\gamma_i}{2})\frac{p_t(x_j)}{p_t(x_i)}$$

that is,

$$\frac{q_{t+1}}{1-q_{t+1}} \le \lambda_K(\frac{\gamma_i}{2})\frac{q_t}{1-q_t}$$

Since $q_t \le \varepsilon_i$, this implies

$$q_{t+1} \le \frac{\lambda_K(\frac{\gamma_i}{2})}{1-q_t} q_t \le \frac{\lambda_K(\frac{\gamma_i}{2})}{1-\varepsilon_i} q_t$$

Define

$$\bar{\lambda}_K := \frac{\lambda_K(\frac{\gamma_i}{2})}{1-\varepsilon_i}$$

Because $\varepsilon_i \le 1 - \lambda_K(\frac{\gamma_i}{2})$, we have $\bar{\lambda}_K \in (0, 1]$. Thus

$$q_{t+1} \le \bar{\lambda}_K q_t$$

Iterating yields

$$q_t \le \bar{\lambda}_K^t q_0$$

that is,

$$1 - p_t(x_i) \le \bar{\lambda}_K^t \big(1 - p_0(x_i)\big)$$

Forward invariance of U_i. From the geometric bound,

$$q_{t+1} \le \bar{\lambda}_K q_t \le \bar{\lambda}_K \varepsilon_i \le \varepsilon_i$$

So the first defining condition of U_i is preserved.

Next, since

$$\bar{\varphi}(p_{t+1}) - \varphi(x_i) = \sum_{j \ne i} p_{t+1}(x_j)\big(\varphi(x_j) - \varphi(x_i)\big),$$

we obtain

$$\|\bar{\varphi}(p_{t+1}) - \varphi(x_i)\| \le \sum_{j \ne i} p_{t+1}(x_j)\|\varphi(x_j) - \varphi(x_i)\| \le 2q_{t+1}$$

Using the preference update,

$$w_{t+1} = \mathrm{Proj}_{\mathcal{W}}((1-\beta)w_t + \beta\bar{\varphi}(p_{t+1}))$$

and

$$\|w_{t+1} - \varphi(x_i)\| \le (1-\beta)\|w_t - \varphi(x_i)\| + \beta\|\bar{\varphi}(p_{t+1}) - \varphi(x_i)\|$$

Hence

$$\|w_{t+1} - \varphi(x_i)\| \le (1-\beta)\frac{\gamma_i}{4} + 2\beta q_{t+1} \le (1-\beta)\frac{\gamma_i}{4} + 2\beta\varepsilon_i$$

Since $\varepsilon_i \le \gamma_i/8$, we get

$$\|w_{t+1} - \varphi(x_i)\| \le (1-\beta)\frac{\gamma_i}{4} + \beta\frac{\gamma_i}{4} = \frac{\gamma_i}{4}$$

So the second defining condition of U_i is also preserved. Therefore U_i is forward invariant.

Convergence of w_t. We already proved $q_t \to 0$. Using again

$$\|\bar{\varphi}(p_{t+1}) - \varphi(x_i)\| \le 2q_{t+1},$$

the preference recursion implies

$$\|w_{t+1} - \varphi(x_i)\| \le (1-\beta)\|w_t - \varphi(x_i)\| + 2\beta q_{t+1}$$

Unrolling this recursion gives

$$\|w_t - \varphi(x_i)\| \le (1-\beta)^t\|w_0 - \varphi(x_i)\| + 2\beta\sum_{s=0}^{t-1}(1-\beta)^{t-1-s}q_{s+1}$$

Since $q_s \to 0$ geometrically, the right-hand side tends to 0. Hence

$$w_t \to \varphi(x_i)$$

Together with $q_t \to 0$, this implies

$$p_t(x_i) \xrightarrow{t\to\infty} 1, \qquad w_t \xrightarrow{t\to\infty} \varphi(x_i).$$

□

C.5 Proof of Corollary 3.5

Corollary 3.5 (Failure of local stability). *Fix $x_i \in \mathcal{X}$. If there exists x_j with $j \neq i$ such that $\varphi(x_j) = \varphi(x_i)$, then the singleton equilibrium δ_{x_i} is not locally asymptotically stable.*

Proof. Assume there exists $j \neq i$ such that

$$\varphi(x_j) = \varphi(x_i)$$

For any $\varepsilon \in (0, 1)$, define

$$p^\varepsilon(x_\ell) = \begin{cases} 1-\varepsilon, & \ell = i, \\ \varepsilon, & \ell = j, \\ 0, & \text{otherwise}, \end{cases} \qquad w^\varepsilon := \varphi(x_i)$$

Since $\varphi(x_j) = \varphi(x_i)$, we have

$$r_{w^\varepsilon}(x_j) = r_{w^\varepsilon}(x_i),$$

so all points in the support of p^ε have the same reward.

By Theorem 3.2, p^ε is a fixed point of the data update.

Moreover,

$$\bar{\varphi}(p^\varepsilon) = (1-\varepsilon)\varphi(x_i) + \varepsilon\varphi(x_j) = \varphi(x_i) = w^\varepsilon,$$

so the preference update also fixes w^ε.

Hence $(p^\varepsilon, w^\varepsilon)$ is an equilibrium for every $\varepsilon \in (0, 1)$.

As $\varepsilon \to 0$, the equilibria $(p^\varepsilon, w^\varepsilon)$ converge to the singleton equilibrium at x_i.

Therefore this singleton equilibrium is not isolated, and hence cannot be locally asymptotically stable. □

C.6 Proof of Proposition 3.6

Lemma 3.6 (Sensitivity of the retraining step). *Let $d_{\mathrm{TV}}(p, q) := \frac{1}{2}\sum_{x\in\mathcal{X}} |p(x) - q(x)|$ denote total variation distance. For any $p, q \in \mathcal{P}(\mathcal{X})$, $w, v \in \mathcal{W}$, and $K \geq 2$, the following hold:*

$$d_{\mathrm{TV}}\left(pH^K_{p,w}, qH^K_{q,w}\right) \leq L_p d_{\mathrm{TV}}(p, q); \quad d_{\mathrm{TV}}\left(pH^K_{p,w}, pH^K_{p,v}\right) \leq L_w\|w - v\|; \quad \|\bar{\varphi}(p) - \bar{\varphi}(q)\| \leq 2d_{\mathrm{TV}}(p, q)$$

where $L_p := 2e^{2\tau}(2 + e^{2\tau})$ and $L_w := 2\tau e^{4\tau}$.

Proof. Recall that

$$a(x) := e^{\tau r_w(x)}$$

Since $r_w(x) \in [-1, 1]$, we have

$$e^{-\tau} \leq a(x) \leq e^{\tau}$$

For finite K,

$$H^K_{p,w}(x) = \mathbb{E}_{X_1,\ldots,X_{K-1}\overset{\text{iid}}{\sim}p}\left[\frac{Ka(x)}{a(x) + \sum_{\ell=1}^{K-1} a(X_\ell)}\right]$$

For $K = \infty$,

$$H^\infty_{p,w}(x) = \frac{a_(x)}{\mathbb{E}_{z\sim p}[a(z)]}$$

We prove the three bounds in turn.

Part 1: Lipschitz continuity in p. Fix $w \in \mathcal{W}$.

$$2d_{\mathrm{TV}}(pH^K_{p,w},\ qH^K_{q,w}) = \sum_{x\in\mathcal{X}} \left|p(x)H^K_{p,w}(x) - q(x)H^K_{q,w}(x)\right|$$

For each $x \in \mathcal{X}$,

$$p(x)H^K_{p,w}(x) - q(x)H^K_{q,w}(x) = (p(x) - q(x))H^K_{p,w}(x) + q(x)\big(H^K_{p,w}(x) - H^K_{q,w}(x)\big)$$

Hence

$$2d_{\mathrm{TV}}(pH^K_{p,w},\ qH^K_{q,w}) \leq I + II,$$

where

$$I := \sum_{x\in\mathcal{X}} |p(x) - q(x)|\, H^K_{p,w}(x), \qquad II := \sum_{x\in\mathcal{X}} q(x)\left|H^K_{p,w}(x) - H^K_{q,w}(x)\right|.$$

We first bound I. Since for all $x,\ p,\ w,\ K$,

$$0 \le H^K_{p,w}(x) \le 2e^{2\tau}$$

we have

$$I \le 2e^{2\tau} \sum_{x\in\mathcal{X}} |p(x) - q(x)| = 4e^{2\tau} d_{\mathrm{TV}}(p, q)$$

Next we bound II. Fix $x \in \mathcal{X}$. For finite K, define

$$g_x(y_1, \dots, y_{K-1}) := \frac{K a_w(x)}{a_w(x) + \sum_{\ell=1}^{K-1} a_w(y_\ell)}$$

Then

$$H^K_{p,w}(x) = \mathbb{E}_{p^{K-1}}[g_x], \qquad H^K_{q,w}(x) = \mathbb{E}_{q^{K-1}}[g_x]$$

We compare these two expectations by replacing the coordinate distributions one at a time. Define, for $\ell = 0, 1, \dots, K-1$,

$$M_\ell := \mathbb{E}\big[g_x(Y_1, \dots, Y_{K-1})\big], \qquad Y_j \sim \begin{cases} q & \text{if } j \le \ell, \\ p & \text{if } j > \ell. \end{cases}$$

Then $M_0 = H^K_{p,w}(x)$, $M_{K-1} = H^K_{q,w}(x)$, and

$$H^K_{p,w}(x) - H^K_{q,w}(x) = \sum_{\ell=1}^{K-1} (M_{\ell-1} - M_\ell)$$

The difference $M_{\ell-1} - M_\ell$ isolates the effect of switching coordinate ℓ from p to q, with all other coordinates unchanged.

In $M_{\ell-1} - M_\ell$, condition on all coordinates except y_ℓ. Write $S_{-\ell} := \sum_{m\neq\ell} a(y_m)$ for their total As a function of y_ℓ alone, the integrand becomes

$$h(y_\ell) := \frac{K\, a(x)}{a(x) + a(y_\ell) + S_{-\ell}}.$$

Since h is decreasing in $a(y_\ell)$ and $a(y_\ell) \in [e^{-\tau},\ e^{\tau}]$:

$$\sup_y h(y) - \inf_y h(y) = K\, a(x) \cdot \frac{e^{\tau} - e^{-\tau}}{(a(x) + e^{-\tau} + S_{-\ell})(a(x) + e^{\tau} + S_{-\ell})}$$

We bound the numerator and denominator separately:

Numerator: Using $a(x) \le e^{\tau}$ and $e^{\tau} - e^{-\tau} \le 2e^{\tau}$.

$$K\, a(x)\, (e^{\tau} - e^{-\tau}) \le K\, e^{\tau} \cdot 2e^{\tau} = 2K\, e^{2\tau}$$

Denominator: each factor satisfies

$$a(x) + e^{\pm\tau} + S_{-\ell} \ge e^{-\tau} + e^{-\tau} + (K-2)\, e^{-\tau} = K\, e^{-\tau}$$

so the product is $\ge K^2\, e^{-2\tau}$.

Therefore

$$\sup_y h(y) - \inf_y h(y) \le \frac{2K\, e^{2\tau}}{K^2\, e^{-2\tau}} = \frac{2e^{4\tau}}{K}.$$

This bound holds for every realization of the conditioned coordinates, since it uses only the worst-case $S_{-\ell} \ge (K-2)\, e^{-\tau}$.

Now, for any bounded function h and any two distributions p, q,

$$\big|H^K_{p,w}(x) - H^K_{q,w}(x)\big| \le (\sup_y h(y) - \inf_y h(y)) \cdot d_{\mathrm{TV}}(p, q).$$

(This follows from the total-variation duality: center h so that $\|h - \mathrm{const}\|_\infty \le (\sup_y h(y) - \inf_y h(y))/2$, then apply $|\mathbb{E}_p[f] - \mathbb{E}_q[f]| \le 2\|f\|_\infty\, d_{\mathrm{TV}}(p, q)$.)

So:

$$\big|H^K_{p,w}(x) - H^K_{q,w}(x)\big| \le \sum_{\ell=1}^{K-1} \frac{2e^{4\tau}}{K}\, d_{\mathrm{TV}}(p, q) = \frac{(K-1)\cdot 2e^{4\tau}}{K}\, d_{\mathrm{TV}}(p, q) < 2e^{4\tau}\, d_{\mathrm{TV}}(p, q)$$

This is a pointwise bound on x, uniform in K.

Hence

$$II \le 2e^{4\tau} d_{\mathrm{TV}}(p, q).$$

Combining the bounds on I and II, we get

$$2d_{\mathrm{TV}}(pH^K_{p,w},\ qH^K_{q,w}) \le \big(4e^{2\tau} + 2e^{4\tau}\big) d_{\mathrm{TV}}(p, q).$$

Equivalently,

$$d_{\mathrm{TV}}(pH^K_{p,w},\ qH^K_{q,w}) \le L_p\, d_{\mathrm{TV}}(p, q),$$

where

$$L_p := 2e^{2\tau}(2 + e^{2\tau})$$

For $K = \infty$, the same inequality remains valid. Indeed,

$$H^\infty_{p,w}(x) = \frac{a_(x)}{\mathbb{E}_{z\sim p}[a(z)]}$$

and the dependence on p is simpler.

Part 2: Lipschitz continuity in w. Fix $p \in \mathcal{P}(\mathcal{X})$. For finite K, use the same samples $X_1, \dots, X_{K-1} \overset{\text{iid}}{\sim} p$, and write

$$S_w := \sum_{\ell=1}^{K-1} a_w(X_\ell), \qquad S_v := \sum_{\ell=1}^{K-1} a_v(X_\ell).$$

Then for any $x \in \mathcal{X}$,

$$\begin{aligned}\left|\frac{Ka_w(x)}{a_w(x)+S_w} - \frac{Ka_v(x)}{a_v(x)+S_v}\right| \le{}& \left|\frac{Ka_w(x)}{a_w(x)+S_w} - \frac{Ka_v(x)}{a_v(x)+S_w}\right| \\ &+ \left|\frac{Ka_v(x)}{a_v(x)+S_w} - \frac{Ka_v(x)}{a_v(x)+S_v}\right|\end{aligned}$$

For the first term, since

$$|a_w(x) - a_v(x)| \le \tau e^{\tau} \|w - v\|$$

we have

$$\left|\frac{Ka_w(x)}{a_w(x)+S_w} - \frac{Ka_v(x)}{a_v(x)+S_w}\right| = \frac{KS_w|a_w(x) - a_v(x)|}{(a_w(x)+S_w)(a_v(x)+S_w)}.$$

Using

$$S_w \le (K-1)e^{\tau}, \qquad a_w(x) + S_w \ge Ke^{-\tau}, \qquad a_v(x) + S_w \ge Ke^{-\tau},$$

we obtain

$$\left|\frac{Ka_w(x)}{a_w(x)+S_w} - \frac{Ka_v(x)}{a_v(x)+S_w}\right| \le \tau e^{4\tau} \|w - v\|.$$

For the second term,

$$|S_v - S_w| \le \sum_{\ell=1}^{K-1} |a_v(X_\ell) - a_w(X_\ell)| \le (K-1)\tau e^{\tau} \|w - v\|.$$

Therefore

$$\left|\frac{Ka_v(x)}{a_v(x)+S_w} - \frac{Ka_v(x)}{a_v(x)+S_v}\right| = \frac{Ka_v(x)|S_v - S_w|}{(a_v(x)+S_w)(a_v(x)+S_v)} \le \tau e^{4\tau} \|w - v\|.$$

Combining the two bounds and taking expectation gives

$$|H^K_{p,w}(x) - H^K_{p_v}(x)| \le 2\tau e^{4\tau} \|w - v\|.$$

Hence

$$d_{\mathrm{TV}}(pH^K_{p,w},\ pH^K_{p_v}) \le L_w\, \|w - v\|,$$

where

$$L_w := 2\tau e^{4\tau}.$$

For $K = \infty$, the same bound follows directly from

$$H^\infty_{p,w}(x) = \frac{a_(x)}{\mathbb{E}_{z\sim p}[a(z)]}$$

Part 3: Lipschitz continuity of $\bar{\varphi}$. Let $u \in \mathbb{R}^d$ satisfy $\|u\| = 1$. Since $\|\varphi(x)\| = 1$,

$$|\langle u, \varphi(x)\rangle| \le 1$$

Therefore

$$|\langle u, \bar{\varphi}(p) - \bar{\varphi}(q)\rangle| = \left| \sum_{x \in \mathcal{X}} \langle u, \varphi(x)\rangle \big(p(x) - q(x)\big) \right| \le 2 d_{\mathrm{TV}}(p, q).$$

Taking the supremum over all unit vectors u yields

$$\|\bar{\varphi}(p) - \bar{\varphi}(q)\| \le 2 d_{\mathrm{TV}}(p, q).$$

This completes the proof. □

C.7 Proofs of Lemma 3.7 and Proposition 3.8

Lemma 3.7 (Unique equilibrium under frozen-w)**.** *Fix $w \in \mathcal{W}$, and consider the the frozen-w model update $p_{t+1}(x) = (1-\eta)p_{\mathrm{ref}}(x) + \eta p(x) H^K_{p,w}(x)$. If $\eta L_p < 1$, the update is a strict contraction on $(\mathcal{P}(\mathcal{X}), d_{\mathrm{TV}})$. Consequently, there exists a unique equilibrium $p^{K,\eta}_w$ and every trajectory converges to it geometrically:*

$$d_{\mathrm{TV}}(p_t, p^{K,\eta}_w) \le (\eta L_p)^t \, d_{\mathrm{TV}}(p_0, p^{K,\eta}_w).$$

Proof. Fix $w \in \mathcal{W}$. For any $p, q \in \mathcal{P}(\mathcal{X})$, define

$$T^K_w(p)(x) := (1-\eta)p_{\mathrm{ref}}(x) + \eta\, p(x) H^K_{p,w}(x)$$

We show that T^K_w is a contraction on $(\mathcal{P}(\mathcal{X}), d_{\mathrm{TV}})$.

Let $p, q \in \mathcal{P}(\mathcal{X})$. Then

$$\begin{aligned} d_{\mathrm{TV}}(T^K_w(p), T^K_w(q)) &= d_{\mathrm{TV}}\Big((1-\eta)p_{\mathrm{ref}} + \eta\, p H^K_{p,w},\ (1-\eta)p_{\mathrm{ref}} + \eta\, q H^K_{q,w}\Big) \\ &= \eta\, d_{\mathrm{TV}}(p H^K_{p,w},\ q H^K_{q,w}). \end{aligned}$$

By Proposition 3.6,

$$d_{\mathrm{TV}}(p H^K_{p,w},\ q H^K_{q,w}) \le L_p\, d_{\mathrm{TV}}(p, q).$$

Therefore,

$$d_{\mathrm{TV}}(T^K_w(p), T^K_w(q)) \le \eta L_p\, d_{\mathrm{TV}}(p, q).$$

If $\eta L_p < 1$, then T^K_w is a strict contraction.

Since $\mathcal{P}(\mathcal{X})$ is complete under d_{TV} in the finite-state setting, the Banach fixed-point theorem implies that T^K_w admits a unique fixed point, denoted $p^{K,\eta}_w$, satisfying

$$p^{K,\eta}_w(x) = (1-\eta)p_{\mathrm{ref}}(x) + \eta\, p^{K,\eta}_w(x) H^K_{(p^{K,\eta}_w)_w}(x)$$

Moreover, for any initialization $p_0 \in \mathcal{P}(\mathcal{X}$, the iterates satisfy

$$d_{\mathrm{TV}}(p_t, p^{K,\eta}_w) \le (\eta L_p)^t \, d_{\mathrm{TV}}(p_0, p^{K,\eta}_w),$$

which proves geometric convergence. □

Proposition 3.8 (Closed-form equilibrium)**.** *As $K \to \infty$, the frozen-w equilibrium takes the form*

$$p^{\infty,\eta}_w(x) = (1-\eta)p_{\mathrm{ref}}(x)\, \frac{Z^\eta_w}{Z^\eta_w - \eta e^{\tau \cdot r_w(x)}}.$$

where $Z^\eta_w := \mathbb{E}_{z \sim p^{\infty,\eta}_w}[e^{\tau \cdot r_w(x)}]$ is the unique scalar satisfying

$$\sum_{i=1}^n (1-\eta)\, p_{\mathrm{ref}}(x_i)\, \frac{e^{\tau \cdot r_w(x_i)}}{Z^\eta_w - \eta e^{\tau \cdot r_w(x_i)}} = 1, \qquad Z^\eta_w > \eta \max_i e^{\tau \cdot r_w(x_i)}.$$

Proof. Recall that

$$a_w(x) := e^{\tau r_w(x)}$$

$$H^\infty_{p,w}(x) = \frac{a_w(x)}{\mathbb{E}_{z \sim p_w}[a_w(z)]}$$

For fixed w, the frozen-w update therefore becomes

$$p_{t+1}(x_i) = (1-\eta)p_{\mathrm{ref}}(x_i) + \eta\, p_t(x_i)\frac{a_w(x_i)}{\mathbb{E}_{z\sim p_t}[a_w(z)]},$$

Let $p_w^{\infty,\eta}$ be a frozen-w fixed point, then the fixed-point equation is

$$p_w^{\infty,\eta} = (1-\eta)p_{\mathrm{ref}} + \eta p_w^{\infty,\eta}\frac{a_w(x)}{\mathbb{E}_{z\sim p_w^{\infty,\eta}}[a_w(z)]}$$

Rearranging,

$$p_w^{\infty,\eta}\left(1-\eta\frac{a_w(x)}{\mathbb{E}_{p_w^{\infty,\eta}}[a_w(z)]}\right) = (1-\eta)p_{\mathrm{ref}},$$

so

$$p_w^{\infty,\eta} = \frac{(1-\eta)p_{\mathrm{ref}}}{1-\eta\frac{a_w(x)}{\mathbb{E}_{z\sim p_w^{\infty,\eta}[a_w(z)]}}} = (1-\eta)p_{\mathrm{ref}}\,\frac{\mathbb{E}_{z\sim p_w^{\infty,\eta}p}[a_w(z)]}{\mathbb{E}_{z\sim p_w^{\infty,\eta}}[a_w(z)]-\eta a_w(x)}.$$

That is

$$p_w^{\infty,\eta} = (1-\eta)p_{\mathrm{ref}}\,\frac{\mathbb{E}_{z\sim p_w^{\infty,\eta}}[e^{\tau r_w(x)}]}{\mathbb{E}_{z\sim p_w^{\infty,\eta}}[e^{\tau r_w(x)})]-\eta e^{\tau r_w(x)}}.$$

It remains to determine $\mathbb{E}_{z\sim p_w^{\infty,\eta}}[a_w(z)]$. Substituting the formula above, we obtain

$$\mathbb{E}_{z\sim p_w^{\infty,\eta}}[a_w(z)] = \sum_{i=1}^{n}(1-\eta)p_{\mathrm{ref}}\,\frac{\mathbb{E}_{z\sim p_w^{\infty,\eta}p}[a_w(z)]}{\mathbb{E}_{z\sim p_w^{\infty,\eta}}[a_w(z)]-\eta a_w(x_i)}a_w(x_i)$$

Since $\mathbb{E}_{z\sim p_w^{\infty,\eta}}[a_w(z)]$, dividing both sides by $\mathbb{E}_{z\sim p_w^{\infty,\eta}}[a_w(z)]$ yields

$$\sum_{i=1}^{n}(1-\eta)p_{\mathrm{ref}}\,\frac{a_w(x_i)}{\mathbb{E}_{z\sim p_w^{\infty,\eta}}[a_w(z)]-\eta a_w(x_i)} = 1$$

We now show that this equation has a unique solution on

$$(\eta\max_i a_w(x_i),\infty).$$

Define

$$F(Z_w^\eta) := \sum_{i=1}^{n}(1-\eta)\,p_{\mathrm{ref}}(x_i)\,\frac{a_w(x_i)}{Z_w^\eta-\eta a_w(x_i)}.$$

For $Z_w^\eta > \eta\max_i a_w(x_i)$, all denominators are positive, and

$$F'(Z_w^\eta) = -\sum_{i=1}^{n}(1-\eta)\,p_{\mathrm{ref}}(x_i)\,\frac{a_w(x_i)}{(Z_w^\eta-\eta a_w(x_i))^2} < 0.$$

Thus F is strictly decreasing on $(\eta\max_i a_w(x_i),\infty)$. Moreover, $F(Z_w^\eta)\to\infty$ as $Z_w^\eta\downarrow\eta\max_i a_w(x_i)$, while

$$F(Z_w^\eta)\to 0 \qquad \text{as } Z_w^\eta\to\infty.$$

Therefore the equation $F(Z_w^\eta)=1$ has a unique solution when $Z_w^\eta > \eta\max_i a_w(x_i)$.

Substituting this unique Z_w^η into the formula yields the unique frozen-w fixed point. □

C.8 Proof of Theorem 3.9

Theorem 3.9 (Convergence of coupled dynamics)**.** *Consider the coupled dynamics (3)-(4). If $\eta(L_p+2L_w)<1$, then there exists $\xi\in\left(\frac{\eta L_w}{\beta(1-2\eta L_w)},\frac{1-\eta L_p}{2\beta\eta L_p}\right)$ such that the coupled update is a strict contraction on $(\mathcal{P}(\mathcal{X})\times\mathcal{W},d_\xi)$, where $d_\xi\big((p,w),(q,v)\big) := d_{\mathrm{TV}}(p,q)+\xi\|w-v\|$. Consequently, there exists a unique equilibrium $(p_K^\star,w_K^\star)$ satisfying (5), and every trajectory converges to it geometrically.*

Proof. Let (p,w) and (q,v) be two states, then:

$$p_{t+1}(x) = (1-\eta)p_{\mathrm{ref}}(x) + \eta\, p_t(x)H^K_{p_t,w}(x), \qquad q_{t+1} = (1-\eta)p_{\mathrm{ref}}(x) + \eta\, q_t(x)H^K_{q_t,v}(x),$$
$$w_{t+1} = \mathrm{Proj}_{\mathcal{W}}((1-\beta)w_t+\beta\bar{\varphi}(p_{t+1}))\,, \qquad v_{t+1} = \mathrm{Proj}_{\mathcal{W}}((1-\beta)v_t+\beta\bar{\varphi}(q_{t+1}))$$

We first bound the data coordinate. By Proposition 3.6,

$$\begin{aligned} d_{\mathrm{TV}}(p_{t+1}, p_{t+1}) &= d_{\mathrm{TV}}\Big((1-\eta)p_{\mathrm{ref}} + \eta\, p_t H^K_{p_t,w},,\ (1-\eta)p_{\mathrm{ref}} + \eta\, q_t H^K_{q_t,v},\Big) \\ &= \eta\, d_{\mathrm{TV}}(p_t H^K_{p_t,w},\ q_t H^K_{q_t,v}) \\ &\le \eta\, d_{\mathrm{TV}}(p_t H^K_{p_t,w},\ q_t H^K_{q_t w}) + \eta\, d_{\mathrm{TV}}(q_t H^K_{q_t,w},\ q_t H^K_{q_t,v},) \\ &\le \eta L_p\, d_{\mathrm{TV}}(p_t, q_t) + \eta L_w\, \|w_t - v_t\|. \end{aligned}$$

Next we bound the preference coordinate. Using the nonexpansiveness of Euclidean projection,

$$\begin{aligned} \|w_{t+1} - v_{t+1}\| &\le \big\|(1-\beta)(w_t - v_t) + \beta\big(\bar\varphi(p_{t+1}) - \bar\varphi(q_{t+1})\big)\big\| \\ &\le (1-\beta)\|w_t - v_t\| + \beta\|\bar\varphi(p_{t+1}) - \bar\varphi(q_{t+1})\|. \end{aligned}$$

Applying Proposition 3.6 again,

$$\|\bar\varphi(p_{t+1}) - \bar\varphi(q_{t+1})\| \le 2\, d_{\mathrm{TV}}(p_{t+1}, q_{t+1}),$$

and hence

$$\|w_{t+1} - v_{t+1}\| \le (1-\beta)\|w_t - v_t\| + 2\beta\, d_{\mathrm{TV}}(p_{t+1}, q_{t+1}).$$

Substituting the bound on $d_{\mathrm{TV}}(p_{t+1}, q_{t+1})$, we obtain

$$\|w_{t+1} - v_{t+1}\| \le 2\beta\eta L_p\, d_{\mathrm{TV}}(p_t, q_t) + \big(1 - \beta + 2\beta\eta L_w\big)\|w_t - v_t\|.$$

Now multiply this inequality by ξ and add it to the data-coordinate bound. This gives

$$\begin{aligned} d_\xi\big((p_{t+1}, w_{t+1}), (q_{t+1}, v_{t+1})\big) &= d_{\mathrm{TV}}(p_{t+1}, q_{t+1}) + \xi\|w_{t+1} - v_{t+1}\| \\ &\le \eta L_p\, d_{\mathrm{TV}}(p_t, q_t) + \eta L_w\, \|w_t - v_t\| + \xi\Big((1-\beta)\|w_t - v_t\| + 2\beta\, d_{\mathrm{TV}}(p_{t+1}, q_{t+1})\Big) \\ &\le \Big(\eta L_p + 2\beta\eta L_p\, \xi\Big) d_{\mathrm{TV}}(p, q) + \xi\Big(\frac{\eta L_w}{\xi} + (1 - \beta + 2\beta\eta L_w)\Big)\|w - v\|. \end{aligned}$$

Therefore the map is a contraction under d_ξ as soon as both coefficients are strictly smaller than the corresponding weights in d_ξ:

$$\begin{cases} \eta L_p + 2\beta\eta L_p\, \xi < 1, \\ \dfrac{\eta L_w}{\xi} + (1 - \beta + 2\beta\eta L_w) < 1. \end{cases}$$

so,

$$\begin{cases} \xi < \dfrac{1 - \eta L_p}{2\beta\eta L_p} \\ \xi > \dfrac{\eta L_w}{\beta(1 - 2\eta L_w)} \end{cases}$$

Thus any ξ in the window

$$\frac{\eta L_w}{\beta(1 - 2\eta L_w)} < \xi < \frac{1 - \eta L_p}{2\beta\eta L_p}$$

makes the coupled update a strict contraction.

It remains to check that this window is nonempty. Since all quantities are positive,

$$\begin{aligned} \frac{\eta L_w}{\beta(1 - 2\eta L_w)} < \frac{1 - \eta L_p}{2\beta\eta L_p} \quad &\Longleftrightarrow \quad 2\eta^2 L_p L_w < (1 - 2\eta L_w)(1 - \eta L_p) \\ &\Longleftrightarrow \quad 0 < 1 - \eta L_p - 2\eta L_w \\ &\Longleftrightarrow \quad \eta(L_p + 2L_w) < 1. \end{aligned}$$

So under the stated condition the interval is nonempty, and we may choose such a ξ.

Finally, $\mathcal{P}(\mathcal{X}) \times \mathcal{W}$ is complete under d_ξ in the finite-state setting, since $\mathcal{P}(\mathcal{X})$ is complete under d_{TV} and $\mathcal{W}$ is closed in finite-dimensional Euclidean space.

By the Banach fixed-point theorem, the coupled reinjected dynamics admit a unique equilibrium $(p^\star_K, w^\star_K)$, and every trajectory converges to it geometrically. □

C.9 Proof of Proposition 4.1

Proposition 4.1 (Sufficient condition for preservation)**.** *Fix* $\eta \in (0, \frac{1}{(L_p+2L_w)})$*. If* $\forall \ell \in \{1, \dots, L\}$

$$\sum_{i=1}^{n} p_{\mathrm{ref},i} \, \langle v_\ell, \varphi(x_i) \rangle \;\geq\; \frac{\theta_\ell - \eta \cdot \min_{i=1,\cdots,n} \langle v_\ell, \varphi(x_i) \rangle}{1-\eta},$$

then the equilibrium satisfies the preservation constraint $\big\langle v_\ell, \bar{\varphi}(p_K^\star(p_{\mathrm{ref}}, \eta)) \big\rangle \geq \theta_\ell$*,* $\forall \ell$*.*

Proof. Fix $\eta \in (0, \bar{\eta})$, and let

$$(p_K^\star, w_K^\star) := (p_K^\star(\mu, \eta),\; w_K^\star(\mu, \eta))$$

denote the unique equilibrium from Theorem 3.9.

For notational simplicity, write

$$p^\star := p_K^\star(\mu, \eta).$$

At equilibrium, the data-update equation gives

$$p^\star(x) = (1-\eta)\mu(x) + \eta \, p^\star(x) H^K_{(p^\star)_{w_K^\star}}(x).$$

Define the curated equilibrium distribution

$$q^\star(x) := p^\star(x) H^K_{(p^\star)_{w_K^\star}}(x)$$

By Proposition 3.1(ii), $q^\star$ is a probability distribution. Hence the equilibrium identity can be written as

$$p^\star = (1-\eta)\mu + \eta q^\star.$$

Applying $\bar{\varphi}(\cdot)$ to both sides and using linearity,

$$\bar{\varphi}(p^\star) = (1-\eta)\bar{\varphi}(\mu) + \eta \bar{\varphi}(q^\star).$$

Fix any protected direction v_ℓ. Taking inner products with v_ℓ, we obtain

$$\big\langle v_\ell, \bar{\varphi}(p^\star) \big\rangle = (1-\eta)\big\langle v_\ell, \bar{\varphi}(\mu) \big\rangle + \eta \big\langle v_\ell, \bar{\varphi}(q^\star) \big\rangle.$$

Now

$$\big\langle v_\ell, \bar{\varphi}(\mu) \big\rangle = \sum_{i=1}^{n} \mu_i \langle v_\ell, \varphi_i \rangle.$$

Define

$$m_\ell := \min_{i=1,\dots,n} \langle v_\ell, \varphi_i \rangle.$$

Since $q^\star \in \mathcal{P}(\mathcal{X})$, we have

$$\big\langle v_\ell, \bar{\varphi}(q^\star) \big\rangle = \sum_{i=1}^{n} q^\star(x_i) \langle v_\ell, \varphi_i \rangle \geq \sum_{i=1}^{n} q^\star(x_i) \, m_\ell = m_\ell.$$

Therefore,

$$\big\langle v_\ell, \bar{\varphi}(p^\star) \big\rangle \geq (1-\eta) \sum_{i=1}^{n} \mu_i \langle v_\ell, \varphi_i \rangle + \eta m_\ell.$$

Now assume

$$\sum_{i=1}^{n} \mu_i \langle v_\ell, \varphi_i \rangle \geq \frac{\theta_\ell - \eta m_\ell}{1-\eta}.$$

Then

$$\big\langle v_\ell, \bar{\varphi}(p^\star) \big\rangle \geq (1-\eta) \cdot \frac{\theta_\ell - \eta m_\ell}{1-\eta} + \eta m_\ell = \theta_\ell.$$

Since this holds for every $\ell = 1, \dots, L$, the equilibrium satisfies

$$\big\langle v_\ell, \bar{\varphi}(p_K^\star(\mu, \eta)) \big\rangle \geq \theta_\ell, \qquad \ell = 1, \dots, L.$$

This proves the claim. □

D Extension to Continuous Instance Spaces

In the main text, we present the model and results in the discrete setting with finite $\mathcal{X} = \{x_1, \dots, x_n\}$. This appendix extends every result to the case where $\mathcal{X} \subset \mathbb{R}^d$ is *compact* instance space.

We denote by $\mathcal{P}(\mathcal{X})$ the set of Borel probability measures on $\mathcal{X}$. The reward $r(x; w) = \langle w, \varphi(x) \rangle$ with $\|\varphi(x)\| = 1$ for every $x \in \mathcal{X}$, feature map, preference state $w \in \mathcal{W} := \{w \in \mathbb{R}^d : \|w\| \le 1\}$, and mean feature operator $\bar{\varphi}(p) := \int_{\mathcal{X}} \varphi(x)\, p(dx)$ carry over unchanged. The coupled dynamics are

$$p_{t+1} = (1-\eta)p_{\mathrm{ref}} + \eta p_t\, H^K_{p_t, w_t} \qquad w_{t+1} = \mathrm{Proj}_{\mathcal{W}}\big((1-\beta)\, w_t + \beta\, \bar{\varphi}(p_{t+1})\big) \tag{11}$$

where $p_{\mathrm{ref}} \in \mathcal{P}(\mathcal{X})$ is a fixed reference measure, $\eta \in [0,1]$, and the curation operator $H^K_{p,w}$ is defined as in Def. 2.1:

$$H^K_{p,w}(x) = \mathbb{E}_{X_1,\dots,X_{K-1} \overset{\text{iid}}{\sim} p}\Big[\frac{K\, e^{\tau r_w(x)}}{e^{\tau r_w(x)} + \sum_{i=1}^{K-1} e^{\tau r_w(X_i)}}\Big].$$

All results below hold for every $K \ge 2$ as well as in the limit $K \to \infty$.

D.1 Retraining with Purely Synthetic User-Curated Data

This appendix extends the results to a compact instance space $X \subset \mathbb{R}^m$. Throughout Appendix C, $\mathcal{P}(X)$ denotes the set of Borel probability measures on X. We assume that the feature map $\phi : X \to \mathbb{R}^d$ is continuous and satisfies $\|\phi(x)\| = 1$ for all $x \in X$. Hence $r_w(x) = \langle w, \phi(x) \rangle$ and $a_w(x) := \exp(\tau r_w(x))$ are continuous and satisfy $e^{-\tau} \le a_w(x) \le e^{\tau}$.

Proposition D.1 (Basic properties of the curation operator on compact spaces)**.** *For any $p \in \mathcal{P}(\mathcal{X})$, $w \in \mathcal{W}$, and $K \ge 2$, the curation operator satisfies:*

1. ***Strict monotonicity.*** *For any $x, y \in \mathcal{X}$, $r_w(x) > r_w(y) \iff H^K_{p,w}(x) > H^K_{p,w}(y)$.*
2. ***Normalization.*** *$\int_{\mathcal{X}} H^K_{p,w}(x)\, p(dx) = 1$.*
3. ***Odds contraction.*** *If $r_w(x) - r_w(y) \ge r_\Delta > 0$, then $\frac{H^K_{p,w}(y)}{H^K_{p,w}(x)} \le \lambda_K(r_\Delta) := 1 - \frac{(1-e^{-\tau r_\Delta})(K-1)}{e^{2\tau}+K-1} \in (0,1)$, and $\lambda_K(r_\Delta) \xrightarrow{K\to\infty} e^{-\tau r_\Delta}$.*

Proof. Each property is proved by arguments that are entirely pointwise and do not rely on finiteness of $\mathcal{X}$; they use only the bound $e^{-\tau} \le e^{\tau r_w(x)} \le e^{\tau}$.

Strict monotonicity. For every fixed realization $S := \sum_{\ell=1}^{K-1} e^{\tau r_w(X_\ell)} > 0$, the map $a \mapsto Ka/(a+S)$ is strictly increasing on $(0,\infty)$, and $a(x) := e^{\tau r_w(x)}$ is strictly increasing in $r_w(x)$. Hence, for every sample path,

$$r_w(x) > r_w(y) \implies \frac{Ka(x)}{a(x)+S} > \frac{Ka(y)}{a(y)+S}.$$

Taking expectations preserves the strict inequality (both sides are continuous functions of the sample path, and the inequality is strict on a set of full measure).

Normalization. Let $X_1, \dots, X_K \overset{\text{iid}}{\sim} p$. By Fubini's theorem,

$$\int_{\mathcal{X}} H^K_{p,w}(x)\, p(dx) = \mathbb{E}\left[\sum_{j=1}^{K} \frac{e^{\tau r_w(X_j)}}{\sum_{i=1}^{K} e^{\tau r_w(X_i)}}\right] = \mathbb{E}[1] = 1.$$

The first equality uses the symmetry of the K i.i.d. draws: each of the K positions contributes identically, so integrating $H^K_{p,w}(x)$ against $p(dx)$ (which accounts for one position) and multiplying by the factor K inside $H^K_{p,w}$ reconstructs the full sum over all K positions.

The second equality holds because the Luce probabilities sum to 1 for every realization $(X_1, \dots, X_K)$.

Odds contraction. Write $a(x) = e^{\tau r_w(x)}$, $a(y) = e^{\tau r_w(y)}$, $S = \sum_{\ell=1}^{K-1} a(X_\ell)$.

Since $r_w(x) - r_w(y) \ge r_\Delta$, we have $a(x) - a(y) \ge (1 - e^{-\tau r_\Delta})\, a(x)$. For any fixed $S \ge 0$,

$$\frac{a(y)/(a(y)+S)}{a(x)/(a(x)+S)} = 1 - \frac{S\,(a(x)-a(y))}{a(x)\,(a(y)+S)}.$$

Using $a(x) - a(y) \ge (1 - e^{-\tau r_\Delta})\, a(x)$, $a(y) \le e^{\tau}$, and $S \ge (K-1)e^{-\tau}$, the ratio $S/(e^{\tau}+S)$ is bounded below by $(K-1)/(e^{2\tau}+K-1)$. Taking expectations yields the stated bound. □

Theorem D.2 (Equilibrium characterization on compact spaces)**.** *When $\eta = 1$, a distribution $p \in \mathcal{P}(\mathcal{X})$ is a fixed point of the retraining step under preference w if and only if the reward r_w is constant on* $\mathrm{supp}(p)$*:*

$$x, x' \in \mathrm{supp}(p) \quad \Longrightarrow \quad r_w(x) = r_w(x').$$

Consequently, for every $x^\star \in \mathcal{X}$, the pair $(p^\star = \delta_{x^\star},\ w^\star = \varphi(x^\star))$ is an equilibrium.

Proof. **Fixed-point characterization.** Let $\mu_p^K(dx) := H_{p,w}^K(x)p(dx)$. The fixed-point equation is $p = \mu_p^K$, equivalently

$$\int_A (H_{p,w}^K(x) - 1)p(dx) = 0, \qquad \forall A \in \mathcal{B}(X).$$

Thus $H_{p,w}^K(x) = 1$ for p-a.e. x. Since ϕ is continuous, $H_{p,w}^K$ is continuous in x by dominated convergence. Therefore $H_{p,w}^K(x) = 1$ on $\mathrm{supp}(p)$.

If $x, y \in \mathrm{supp}(p)$ and $r_w(x) \neq r_w(y)$, strict monotonicity from Proposition C.1 gives $H_{p,w}^K(x) \neq H_{p,w}^K(y)$, contradicting the fact that both equal 1. Hence r_w is constant on $\mathrm{supp}(p)$. Conversely, if r_w is constant on $\mathrm{supp}(p)$, then p-a.s. all sampled points have the same reward as every $x \in \mathrm{supp}(p)$, so directly $H_{p,w}^K(x) = 1$ on $\mathrm{supp}(p)$ and hence $pH_{p,w}^K = p$.

Conversely, if all support points share a common reward c, then $e^{\tau r_w(x)} = e^{\tau c}$ for every $x \in \mathrm{supp}(p)$, and by direct computation $H_{p,w}^K(x) = 1$ for every such x.

Singleton equilibria. For any $x^\star \in \mathcal{X}$, $\delta_{x^\star}$ is trivially supported on a single point, so the equal-reward condition holds. Moreover, $\bar{\varphi}(\delta_{x^\star}) = \varphi(x^\star)$, and $\mathrm{Proj}_{\mathcal{W}}\big((1-\beta)\,\varphi(x^\star) + \beta\,\varphi(x^\star)\big) = \varphi(x^\star)$ since $\|\varphi(x^\star)\| = 1 \leq 1$. Hence $(\delta_{x^\star}, \varphi(x^\star))$ is an equilibrium. □

Assumption D.3 (Multiscale reward gap)**.** Fix $x^\star \in \mathcal{X}$. There exist a radius $r_0 > 0$ and sequences $(a_k)_{k\geq 0}$, $(b_k)_{k\geq 0}$, $(\gamma_k)_{k\geq 0}$ with $a_k \downarrow 0$, $b_k \downarrow 0$, $\gamma_k > 0$, such that the nested balls $U_k := B_{2^{-k}r_0}(x^\star)$, $k = 0, 1, 2, \ldots$ (a sequence of shrinking neighborhoods of $x^\star$ with radii halving at each scale) satisfy, for every $k \geq 0$:

1. **Feature continuity:** $\sup_{x\in U_k} \|\varphi(x) - \varphi(x^\star)\| \leq a_k$.
2. **Reward gap:** $\|w - \varphi(x^\star)\| \leq b_k \implies \inf_{y\in U_{k+1}} \langle w, \varphi(y)\rangle - \sup_{x\notin U_k} \langle w, \varphi(x)\rangle \geq \gamma_k$.
3. **Scale compatibility:** $a_k < b_{k+1}$ for all $k \geq 0$, and $a_0 < b_0$.

Remark D.4 (Interpretation)**.** Assumption D.3 is the continuous analogue of the feature-separation condition $\varphi(x_j) \neq \varphi(x_i)$ for all $j \neq i$ in Theorem 3.3. In the discrete case, the finite set of competitors admits a single uniform reward gap $\gamma_i = \min_{j\neq i}\langle \varphi(x_i), \varphi(x_i) - \varphi(x_j)\rangle > 0$, and the multiscale conditions are automatically satisfied with $a_k = 0$, $b_k = \gamma_i/4$, $\gamma_k = \gamma_i/2$. In the continuous case, competitors can be arbitrarily close to $x^\star$ with arbitrarily small reward disadvantage, so the conditions must hold at every scale separately. Specifically: (i) controls how well the feature mean approximates $\varphi(x^\star)$ after mass concentrates at scale k; (ii) provides the reward advantage that drives mass from U_k^c into U_{k+1}; (iii) ensures that the feature accuracy gained at scale k is sufficient to activate the reward gap at scale $k+1$. These conditions are not implied by ordinary feature separation alone. They hold, for example, whenever the chosen neighborhoods (U_k) satisfy a uniform annular reward separation condition

$$\inf_{y\in U_{k+1}} \langle \phi(x^\star), \phi(y)\rangle > \sup_{x\notin U_k} \langle \phi(x^\star), \phi(x)\rangle$$

with margins large enough to dominate perturbations $\|w - \phi(x^\star)\| \leq b_k$.

Theorem D.5 (Local convergence to singleton equilibria on compact spaces)**.** *Let $x^\star \in \mathcal{X}$ satisfy Assumption D.3. Then there exists $\delta^\star > 0$ such that if*

$$x^\star \in \mathrm{supp}(p_0), \qquad \|w_0 - \varphi(x^\star)\| \leq \delta^\star, \qquad p_0(U_1) \geq 1 - \delta^\star,$$

then $p_t(B_\varepsilon(x^\star)) \to 1$ for every $\varepsilon > 0$ and $w_t \to \varphi(x^\star)$.

Proof. Since $a_0 < b_0$, choose $\delta^\star > 0$ with

$$0 < \delta^\star \leq \min\{\tfrac{1}{2},\ b_0\}, \qquad a_0 + \frac{2\delta^\star}{1-\delta^\star} \leq b_0 \tag{12}$$

Define $q_{k,t} := p_t(U_k^c)$ (the mass outside scale k) and $s_{k,t} := p_t(U_{k+1})$ (the mass inside scale $k+1$).

Assume $x^\star \in \mathrm{supp}(p_0)$, $\|w_0 - \varphi(x^\star)\| \leq \delta^\star$, and $p_0(U_1) \geq 1 - \delta^\star$.

Step 1: The trajectory stays inside the initial preference neighborhood.

We prove $\|w_t - \varphi(x^\star)\| \leq b_0$ for all t by induction.

The base case is $\|w_0 - \varphi(x^\star)\| \leq \delta^\star \leq b_0$.

Suppose $\|w_t - \varphi(x^\star)\| \le b_0$. Then reward gap with $k = 0$ gives

$$\inf_{y \in U_1} r(y; w_t) - \sup_{x \notin U_0} r(x; w_t) \ge \gamma_0.$$

When $\eta = 1$, the retraining step gives $p_{t+1}(A) = \int_A H^K_{p_t, w_t}(x)\, p_t(dx)$.

For every $y \in U_0^c$ and $x \in U_1$,it yields $r_{w_t}(x) - r_{w_t}(y) \ge \gamma_0$, so Proposition D.1 (odds contraction) gives

$$H^K_{p_t, w_t}(y) \le \lambda_K(\gamma_0)\, H^K_{p_t, w_t}(x).$$

Since this holds for *every* $x \in U_1$, $H^K_{p_t, w_t}(y) \le \lambda_K(\gamma_0)\, \inf_{x \in U_1} H^K_{p_t, w_t}(x)$ for all $y \in U_0^c$.

Therefore

$$q_{0,t+1} = \int_{U_0^c} H^K_{p_t, w_t}(y)\, p_t(dy) \le \lambda_K(\gamma_0) \inf_{x \in U_1} H^K_{p_t, w_t}(x) \cdot q_{0,t}$$

and

$$s_{0,t+1} = \int_{U_1} H^K_{p_t, w_t}(x)\, p_t(dx) \ge \inf_{x \in U_1} H^K_{p_t, w_t}(x) \cdot s_{0,t},$$

whence

$$\frac{q_{0,t+1}}{s_{0,t+1}} \le \lambda_K(\gamma_0)\, \frac{q_{0,t}}{s_{0,t}}. \tag{13}$$

Iterating,

$$\frac{q_{0,t}}{s_{0,t}} \le \lambda_K(\gamma_0)^t \cdot \frac{\delta^\star}{(1 - \delta^\star)}$$

and in particular $q_{0,t} \le \frac{\delta^\star}{(1-\delta^\star)}$ for all t.

By feature continuity and $\|\varphi(x)\| = 1$,

$$\|\bar\varphi(p_{t+1}) - \varphi^\star\| \le a_0 + 2\, q_{0,t+1}.$$

Since projection is nonexpansive and $\varphi^\star \in \mathcal{W}$,

$$\|w_{t+1} - \varphi(x^\star)\| \le (1 - \beta)\, \|w_t - \varphi(x^\star)\| + \beta\, (a_0 + 2q_{0,t+1}) \le (1 - \beta)\, b_0 + \beta\Big(a_0 + \frac{2\delta^\star}{1 - \delta^\star}\Big) \le b_0.$$

Step 2: First-scale outer leakage vanishes.

Since $\|w_t - \varphi(x^\star)\| \le b_0$ for all t, inequality (13) holds for all t, so $\frac{q_{0,t}}{s_{0,t}} \to 0$ geometrically.

Since $q_{0,t} \le \frac{q_{0,t}}{s_{0,t}}$, we get $q_{0,t} \to 0$.

Step 3: Scale-by-scale bootstrap.

We prove by induction: for every $k \ge 0$ there exists $T_k < \infty$ with $\|w_t - \varphi(x^\star)\| \le b_k$ for all $t \ge T_k$.

The case $k = 0$ is Step 1. Suppose the claim holds for k.

Support preservation. Since $H^K_{p,w}(x) > 0$ for all x (as $e^{\tau r_w(x)} > 0$ and the denominator in Definition 2.1 is finite and positive), the pure update preserves the support: $\mathrm{supp}(p_t) = \mathrm{supp}(p_0)$ for all t. Because $x^\star \in \mathrm{supp}(p_0)$ and U_{k+1} is open, $s_{k,t} = p_t(U_{k+1}) > 0$ for all t.

Mass concentration. For $t \ge T_k$, reward gap gives at scale k.

Exactly as in Step 1, Proposition D.1 (odds contraction) yields

$$\frac{q_{k,t+1}}{s_{k,t+1}} \le \lambda_K(\gamma_k)\, \frac{q_{k,t}}{s_{k,t}},$$

and iterating gives $q_{k,t} \to 0$.

Preference bootstrap. $\|\bar\varphi(p_t) - \varphi^\star\| \le a_k + 2q_{k,t}$, so $\|w_{t+1} - \varphi(x^\star)\| \le (1 - \beta)\, \|w_t - \varphi(x^\star)\| + \beta\, (a_k + 2q_{k,t+1})$.

Since $q_{k,t} \to 0$, $\limsup_t \|w_t - \varphi(x^\star)\| \le a_k < b_{k+1}$, giving T_{k+1} with $\|w_t - \varphi(x^\star)\| \le b_{k+1}$ for $t \ge T_{k+1}$.

Step 4: Convergence. For every k, $p_t(U_k^c) \to 0$.

Given $\varepsilon > 0$, choose k with $r_k < \varepsilon$; then$p_t(B_\varepsilon(x^\star)^c) \le p_t(U_k^c) \to 0$.

The induction gives $\limsup_t d_t \le a_k \to 0$, so $w_t \to \varphi^\star$. □

Corollary D.6 (Failure of local stability on compact spaces)**.** *Fix $x^\star \in \mathcal{X}$. If there exists $x' \ne x^\star$ with $\varphi(x') = \varphi(x^\star)$, then the singleton equilibrium $(\delta_{x^\star}, \varphi(x^\star))$ is not locally asymptotically stable.*

Proof. For any $\varepsilon \in (0,1)$, define $p_\varepsilon := (1-\varepsilon)\,\delta_{x^\star} + \varepsilon\,\delta_{x'}$ and $w_\varepsilon := \varphi(x^\star)$.

Since $\varphi(x') = \varphi(x^\star)$, the two support points share the same reward under w_ε, so $H^K_{p_\varepsilon,w_\varepsilon}(x^\star) = H^K_{p_\varepsilon,w_\varepsilon}(x') = 1$ by the equal-reward characterization in Theorem D.2.

Moreover, $\bar\varphi(p_\varepsilon) = \varphi(x^\star) = w_\varepsilon$.

Hence $(p_\varepsilon, w_\varepsilon)$ is an equilibrium for every $\varepsilon > 0$.

Since $(p_\varepsilon, w_\varepsilon) \to (\delta_{x^\star}, \varphi(x^\star))$as $\varepsilon \to 0$, the singleton equilibrium is not isolated and therefore not locally asymptotically stable. □

D.2 Retraining on Mixed Synthetic and Reference Data

We now consider $\eta \in (0,1)$, paralleling Section 3.2.

Lemma D.7 (Sensitivity of the retraining step on compact spaces)**.** *For $p \in \mathcal{P}(\mathcal{X})$, define the probability measure $pH^K_{p,w}$ by*

$$(pH^K_{p,w})(A) := \int_A H^K_{p,w}(x)\,p(dx), \qquad A \in \mathcal{B}(\mathcal{X}).$$

Then for any $p, q \in \mathcal{P}(\mathcal{X})$, $w, v \in \mathcal{W}$, and $K \geq 2$,

$$d_{\mathrm{TV}}\big(pH^K_{p,w}, qH^K_{q,w}\big) \leq L_p d_{\mathrm{TV}}(p,q),$$

$$d_{\mathrm{TV}}\big(pH^K_{p,w}, pH^K_{p,v}\big) \leq L_w \|w - v\|,$$

and

$$\|\bar\phi(p) - \bar\phi(q)\| \leq 2 d_{\mathrm{TV}}(p,q),$$

where $L_p := 2e^{2\tau}(2 + e^{2\tau})$ and $L_w := 2\tau e^{4\tau}$.

Proof. The proof of Lemma 3.6 (Sec. B.6) relies on three ingredients:

1. the uniform bounds $e^{-\tau} \leq e^{\tau r_w(x)} \leq e^{\tau}$;
2. the variational formula $\left|\int f d(p-q)\right| \leq 2\|f\|_\infty\, d_{\mathrm{TV}}(p,q)$, valid on any measurable space;
3. the coordinate-swapping telescoping argument bounding $|H^K_{p,w}(x) - H^K_{q,w}(x)|$ (Part 1) and $|H^K_{p,w}(x) - H^K_{p,v}(x)|$ (Part 2).

None of these steps use finiteness of $\mathcal{X}$: the sums in the discrete proof become integrals, and all bounds hold for arbitrary Borel probability measures.

Part 1: p-Lipschitz bound.

For any measurable A,

$$(pH^K_{p,w} - qH^K_{q,w})(A) = \int_A H^K_{p,w}\, d(p-q) + \int_A (H^K_{p,w} - H^K_{q,w})\, dq.$$

Taking the supremum over measurable A gives

$$2d_{\mathrm{TV}}(pH^K_{p,w}, qH^K_{q,w}) \leq 2\|H^K_{p,w}\|_\infty d_{\mathrm{TV}}(p,q) + \|H^K_{p,w} - H^K_{q,w}\|_\infty.$$

The first term contributes at most $4e^{2\tau}\, d_{\mathrm{TV}}(p,q)$ since $0 \leq H^K_{p,w}(x) \leq 2e^{2\tau}$. The second term is bounded via the coordinate-swapping argument: the effect of replacing each of the $K-1$ sample coordinates from p to q contributes at most $2e^{4\tau}/K$ in the sup-norm of the integrand, yielding a total of at most $2e^{4\tau}\, d_{\mathrm{TV}}(p,q)$.

Combining: $d_{\mathrm{TV}}(pH^K_{p,w}, qH^K_{q,w}) \leq (2e^{2\tau} + e^{4\tau})\, d_{\mathrm{TV}}(p,q) = L_p\, d_{\mathrm{TV}}(p,q)$.

Part 2: w-Lipschitz bound.

Using $|e^{\tau r_w(x)} - e^{\tau r_v(x)}| \leq \tau e^{\tau}\, \|w - v\|$ and the two-term decomposition of $H^K_{p,w}(x) - H^K_{p,v}(x)$ (see Sec. B.6, Part 2), each term is bounded by $\tau e^{4\tau}\, \|w - v\|$, giving $d_{\mathrm{TV}}(pH^K_{p,w}, pH^K_{p,v}) \leq 2\tau e^{4\tau}\, \|w - v\| = L_w\, \|w - v\|$.

Part 3: Mean-feature bound. For any unit vector u, $|\langle u, \varphi(x)\rangle| \leq 1$, so$|\langle u, \bar\varphi(p) - \bar\varphi(q)\rangle| \leq 2\, d_{\mathrm{TV}}(p,q)$.

Taking the supremum over u gives $\|\bar\varphi(p) - \bar\varphi(q)\| \leq 2\, d_{\mathrm{TV}}(p,q)$. □

Theorem D.8 (Unique equilibrium under frozen-w on compact spaces)**.** *Fix $w \in \mathcal{W}$ and $K \geq 2$ or $K = \infty$. Define $T^K_w : \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{X})$ by*

$$T^K_w(p)(A) := (1-\eta) p_{\mathrm{ref}}(A) + \eta \int_A H^K_{p,w}(x)\, p(dx), \qquad A \in \mathcal{B}(\mathcal{X}).$$

If $\eta L_p < 1$, then T_w^K is a strict contraction on $(\mathcal{P}(\mathcal{X}), d_{\mathrm{TV}})$ with rate ηL_p. Consequently, there exists a unique equilibrium $p_w^{K,\eta}$ and every trajectory converges geometrically:

$$d_{\mathrm{TV}}(p_t, p_w^{K,\eta}) \le (\eta L_p)^t d_{\mathrm{TV}}(p_0, p_w^{K,\eta}).$$

Proof. For any $p, q \in \mathcal{P}(\mathcal{X})$ and measurable A,

$$T_w^K(p)(A) - T_w^K(q)(A) = \eta\big[(pH_{p,w}^K)(A) - (qH_{q,w}^K)(A)\big].$$

Therefore, by Lemma D.7,

$$d_{\mathrm{TV}}(T_w^K(p), T_w^K(q)) = \eta d_{\mathrm{TV}}(pH_{p,w}^K, qH_{q,w}^K) \le \eta L_p d_{\mathrm{TV}}(p, q).$$

If $\eta L_p < 1$, this is a strict contraction. Since $(\mathcal{P}(\mathcal{X}), d_{\mathrm{TV}})$ is complete, Banach's fixed-point theorem gives a unique fixed point $p_w^{K,\eta}$ and geometric convergence. □

Proposition D.9 (Closed-form equilibrium on compact spaces)**.** *Assume the condition of Theorem D.8, in particular $\eta L_p < 1$. For the limiting curation operator $K = \infty$, define*

$$a_w(x) := e^{\tau r_w(x)}, \qquad M_w := \sup_{x \in \mathcal{X}} a_w(x).$$

Then the frozen-w equilibrium is given, for every measurable $A \subseteq \mathcal{X}$, by

$$p_w^{\infty,\eta}(A) = \int_A \frac{(1-\eta)\, Z_w^\eta}{Z_w^\eta - \eta a_w(x)}\, p_{\mathrm{ref}}(dx),$$

where $Z_w^\eta > \eta M_w$ is the unique scalar solving

$$\int_{\mathcal{X}} \frac{(1-\eta)\, a_w(x)}{Z_w^\eta - \eta a_w(x)}\, p_{\mathrm{ref}}(dx) = 1.$$

Proof. For $K = \infty$, the curation operator is

$$H_{p,w}^\infty(x) = \frac{a_w(x)}{\int_{\mathcal{X}} a_w(z)\, p(dz)}.$$

Let p be a frozen-w fixed point and set

$$Z := \int_{\mathcal{X}} a_w(z)\, p(dz).$$

Since $e^{-\tau} \le a_w(x) \le e^{\tau}$, we have $Z \ge e^{-\tau}$. Moreover, $\eta L_p < 1$ implies $\eta < e^{-2\tau}$, and hence

$$\eta M_w \le \eta e^{\tau} < e^{-\tau} \le Z.$$

Therefore $Z - \eta a_w(x) > 0$ for every $x \in \mathcal{X}$.

The fixed-point equation

$$p = (1-\eta)p_{\mathrm{ref}} + \eta p H_{p,w}^\infty$$

is equivalently

$$p(dx)\left(1 - \frac{\eta a_w(x)}{Z}\right) = (1-\eta)p_{\mathrm{ref}}(dx).$$

Thus

$$p(dx) = \frac{(1-\eta)Z}{Z - \eta a_w(x)}\, p_{\mathrm{ref}}(dx).$$

Substituting this expression into the definition of Z and dividing by $Z > 0$ gives the scalar equation

$$\int_{\mathcal{X}} \frac{(1-\eta) a_w(x)}{Z - \eta a_w(x)}\, p_{\mathrm{ref}}(dx) = 1.$$

It remains to show that this equation has a unique solution on $(\eta M_w, \infty)$. Define

$$F(Z) := \int_{\mathcal{X}} \frac{(1-\eta) a_w(x)}{Z - \eta a_w(x)}\, p_{\mathrm{ref}}(dx), \qquad Z > \eta M_w.$$

The function F is continuous and strictly decreasing on $(\eta M_w, \infty)$, and $F(Z) \to 0$ as $Z \to \infty$.

We now check that F is larger than 1 near the left endpoint. Since $a_w(x) \in [e^{-\tau}, M_w]$, the map $a \mapsto a/(Z - \eta a)$ is increasing for $Z > \eta M_w$. Hence

$$F(Z) \ge \frac{(1-\eta)e^{-\tau}}{Z - \eta e^{-\tau}}.$$

Taking $Z \downarrow \eta M_w$ gives

$$\liminf_{Z \downarrow \eta M_w} F(Z) \geq \frac{(1-\eta)e^{-\tau}}{\eta(M_w - e^{-\tau})},$$

with the right-hand side interpreted as $+\infty$ if $M_w = e^{-\tau}$. Since $M_w \leq e^{\tau}$ and $\eta < e^{-2\tau}$, we have $\eta M_w < e^{-\tau}$, which implies

$$\frac{(1-\eta)e^{-\tau}}{\eta(M_w - e^{-\tau})} > 1$$

whenever $M_w > e^{-\tau}$. Therefore $F(Z) > 1$ for Z sufficiently close to ηM_w, while $F(Z) \to 0$ as $Z \to \infty$. By the intermediate value theorem, there exists a solution $Z_w^\eta > \eta M_w$, and strict monotonicity gives uniqueness.

Finally, for this unique Z_w^η, define $p_w^{\infty,\eta}$ by the displayed formula. The scalar equation implies

$$\int_{\mathcal{X}} a_w(x)\, p_w^{\infty,\eta}(dx) = Z_w^\eta,$$

and also

$$p_w^{\infty,\eta}(\mathcal{X}) = (1-\eta) + \eta \int_{\mathcal{X}} \frac{(1-\eta)a_w(x)}{Z_w^\eta - \eta a_w(x)}\, p_{\mathrm{ref}}(dx) = 1.$$

Substituting the formula back into the frozen-w update verifies that $p_w^{\infty,\eta}$ is a fixed point. Uniqueness follows from the uniqueness of Z_w^η, equivalently from the contraction in Theorem D.8. □

Theorem D.10 (Convergence of coupled dynamics on compact spaces)**.** *If $\eta(L_p + 2L_w) < 1$, then there exists $\xi \in \left(\frac{\eta L_w}{\beta(1-2\eta L_w)},\ \frac{1-\eta L_p}{2\beta\eta L_p}\right)$ such that the coupled update is a strict contraction on $(\mathcal{P}(\mathcal{X}) \times \mathcal{W},\ d_\xi)$ with $d_\xi\big((p,w),(q,v)\big) := d_{\mathrm{TV}}(p,q) + \xi\,\|w - v\|$. Consequently, there exists a unique equilibrium $(p_K^\star, w_K^\star)$ and every trajectory converges geometrically.*

Proof. Let $(p^+, w^+) := F(p,w)$ and $(q^+, v^+) := F(q,v)$, where F denotes the joint update.

Data coordinate. By Lemma D.7,

$$d_{\mathrm{TV}}(p^+, q^+) \leq \eta L_p\, d_{\mathrm{TV}}(p,q) + \eta L_w\, \|w - v\|.$$

Preference coordinate. Using nonexpansiveness of projection and $\|\bar\varphi(p^+) - \bar\varphi(q^+)\| \leq 2\, d_{\mathrm{TV}}(p^+, q^+)$,

$$\begin{aligned}\|w^+ - v^+\| &\leq (1-\beta)\,\|w - v\| + 2\beta\, d_{\mathrm{TV}}(p^+, q^+) \\ &\leq 2\beta\eta L_p\, d_{\mathrm{TV}}(p,q) + (1 - \beta + 2\beta\eta L_w)\,\|w - v\|.\end{aligned}$$

Combined metric. Multiplying the preference bound by ξ and adding,

$$d_\xi(F(p,w), F(q,v)) \leq A(\xi)\, d_{\mathrm{TV}}(p,q) + \xi\, B(\xi)\,\|w - v\|,$$

where $A(\xi) := \eta L_p(1 + 2\xi\beta)$ and $B(\xi) := 1 - \beta + 2\beta\eta L_w + \eta L_w/\xi$. The map is a contraction when $A(\xi) < 1$ and $B(\xi) < 1$:

$$\begin{aligned}A(\xi) < 1 &\iff \xi < \frac{1 - \eta L_p}{2\beta\eta L_p}, \\ B(\xi) < 1 &\iff \xi > \frac{\eta L_w}{\beta(1 - 2\eta L_w)}.\end{aligned}$$

Both requirements are implied by $\eta(L_p + 2L_w) < 1$, and under this condition the window is nonempty:

$$\frac{\eta L_w}{\beta(1 - 2\eta L_w)} < \frac{1 - \eta L_p}{2\beta\eta L_p} \iff \eta(L_p + 2L_w) < 1.$$

Completeness. $\mathcal{P}(\mathcal{X})$ is complete under d_{TV} when $\mathcal{X}$ is compact, and $\mathcal{W}$ is closed in $\mathbb{R}^d$, so $(\mathcal{P}(\mathcal{X}) \times \mathcal{W},\ d_\xi)$ is complete. The Banach fixed-point theorem gives the unique equilibrium and geometric convergence. □

E Additional Experimental Results

We conduct all the experiments on a server which has two Intel Xeon 6326 CPU and 6 Nvidia A6000 GPU.

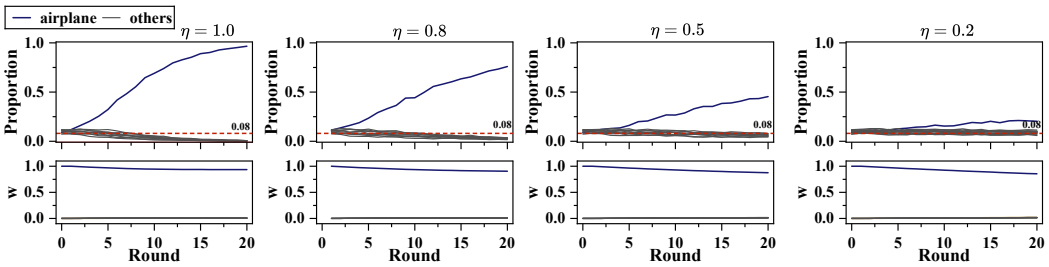

Figure 4: Effect of mixing ratio η ($\tau = 5.0$, $K = \infty$, $\beta = 0.01$).*Top*: per-class proportions. *Bottom*: preference vector w_t. Dashed line marks $\theta = 0.08$.

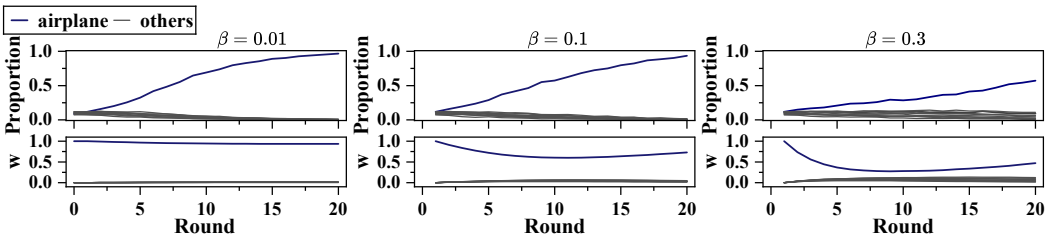

Figure 5: Effect of preference change rate β ($\eta = 1.0$, $\tau = 5.0$, $K = \infty$).

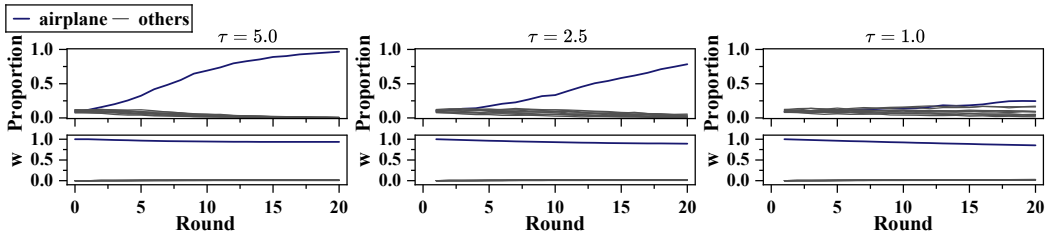

Figure 6: Effect of curation temperature τ ($\eta = 1.0$, $K = \infty$, $\beta = 0.01$).

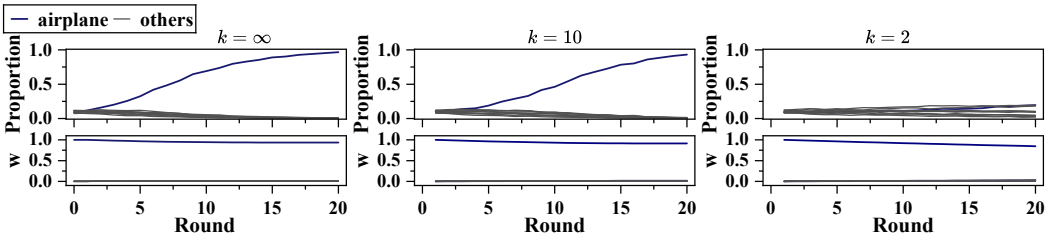

Figure 7: Effect of selection group size K ($\eta = 1.0$, $\tau = 5.0$, $\beta = 0.01$).

E.1 Additional CFM Experiments on CIFAR-10

E.1.1 Effect of Key Parameters

We present the effect of key parameters for OT-CFM on CIFAR-10. The default configuration is $\eta = 1.0$, $\tau = 5.0$, $K = \infty$, $\beta = 0.01$, $w_0 = [1, 0, 0, 0, \dots]$. We vary one parameter at a time.

Mixing Ratio η. Fig. 4 shows the effect of η. At $\eta = 1.0$ (pure synthetic), airplane reaches >95% by round 20. As η decreases, reference data increasingly stabilizes the distribution: at $\eta = 0.8$, the concentration is slightly mitigated; at $\eta = 0.5$, airplane stabilizes around 50%; at $\eta = 0.2$, it stays below 50% with other classes remaining above $\theta = 0.08$.The preference vector w_t evolves similarly across all settings because $\beta = 0.01$ is small, confirming that the mode shift is driven by η rather than by changes in w_t.

Preference change rate β**.** Fig. 5 shows the effect of β. At $\beta = 0.01$, w_t changes slowly and the curation direction remains coherent, producing the strongest concentration (>95% airplane). At $\beta = 0.3$, w_t tracks the current distribution too closely, diluting the preference toward a uniform direction and weakening the curation signal. However, as mode concentration progresses, $\bar{\varphi}(p_t)$ itself becomes dominated by the airplane component,

which in turn pulls w_t back toward $[1, 0, \dots]$. This restores the curation signal and drives further concentration, although it takes effect more slowly compared to small β.

Curation Temperature τ. Fig. 6 shows the effect of τ. At $\tau = 5.0$, curation concentrates sharply: even small reward differences are amplified exponentially, driving airplane above 95% by round 20. At $\tau = 2.5$, the concentration is slower. At $\tau = 1.0$, the softmax weights are nearly uniform and the class distribution remains close to balanced. The speed of mode concentration increases monotonically with τ, consistent with the Lipschitz constants L_p growing rapidly in τ.

Selection Group Size K. Fig. 7 shows the effect of K. At $K = \infty$ and $K = 10$, the concentration trajectories are nearly identical, with airplane exceeding 90% by round 20. At $K = 2$, each Luce group contains only two candidates, limiting the selector's ability to distinguish high-reward samples. This is consistent with the discussion of the role of K.

E.1.2 Reference data design

Setting A: Uniform thresholds. As described in Section 5, all four protected classes share the same threshold $\theta = 0.08$ with $c = (1, 1, 11, 12, 13, 16, 30, 28, 27, 25)$. The feasibility boundary is at $\eta^* = 0.67$. Since the full table is too long, we present Table 3 here to show the partial LP search results.

Table 3: Partial results of searching over 50 evenly spaced points in $\eta \in [0.2, 0.9]$. (Setting A: protect frog/horse/ship/truck, $\theta_\ell = 0.08$, $c = (1, 1, 11, 12, 13, 16, 30, 28, 27, 25)$)

η	air	auto	bird	cat	deer	dog	frog	horse	ship	truck	C^*	J	status
0.20	.000	.600	.000	.000	.000	.000	.100	.100	.100	.100	11.59	46,378	✓
0.30	.000	.543	.000	.000	.000	.000	.114	.114	.114	.114	13.11	45,883	✓
0.40	.000	.467	.000	.000	.000	.000	.133	.133	.133	.133	15.13	45,388	✓
0.50	.000	.360	.000	.000	.000	.000	.160	.160	.160	.160	17.96	44,892	✓
0.60	.000	.200	.000	.000	.000	.000	.200	.200	.200	.200	22.20	44,397	✓
0.67	**.000**	**.026**	**.000**	**.000**	**.000**	**.000**	**.243**	**.243**	**.243**	**.243**	**26.81**	**44,043**	η^*
0.69	infeasible										–	–	✗
0.90	infeasible										–	–	✗

Setting B: Heterogeneous thresholds. The four protected classes have different thresholds reflecting varying preservation priorities: $\theta_{\text{frog}} = 0.05$, $\theta_{\text{horse}} = 0.08$, $\theta_{\text{ship}} = 0.10$, $\theta_{\text{truck}} = 0.15$. And the $c = (1, 1, 11, 12, 13, 16, 30, 28, 27, 25)$ keep same.

Table 4: Partial results of searching over 50 evenly spaced points in $\eta \in [0.2, 0.9]$ (Setting B: protect frog/horse/ship/truck,$\theta = (0.05, 0.08, 0.10, 0.15)$, $c = (1, 1, 11, 12, 13, 16, 30, 28, 27, 25)$).

η	air	auto	bird	cat	deer	dog	frog	horse	ship	truck	C^*	J	status
0.20	.000	.525	.000	.000	.000	.000	.062	.100	.125	.187	13.26	53,030	✓
0.30	.000	.457	.000	.000	.000	.000	.071	.114	.143	.214	15.01	52,534	✓
0.40	.000	.367	.000	.000	.000	.000	.083	.133	.167	.250	17.35	52,039	✓
0.50	.000	.240	.000	.000	.000	.000	.100	.160	.200	.300	20.62	51,544	✓
0.60	.000	.050	.000	.000	.000	.000	.125	.200	.250	.375	25.52	51,049	✓
0.61	**.000**	**.015**	**.000**	**.000**	**.000**	**.000**	**.130**	**.207**	**.259**	**.389**	**26.43**	**50,978**	η^*
0.63	infeasible										–	–	✗
0.90	infeasible										–	–	✗

Table 4 shows the partial LP search results. The feasibility boundary is at $\eta^* = 0.61$. The heterogeneous thresholds shift the feasibility boundary and change the optimal allocation: the LP assigns more reference mass to classes with higher thresholds (ship, truck) and less to those with lower thresholds (frog, horse). Table 5 compares our designed reference with uniform reference and the actual performance of the reference data is shown in the Fig. 8.

E.1.3 Generated Samples

To complement the quantitative results, we provide generated samples at selected rounds in Fig. 9 and Fig. 10, offering a visual illustration of how the model's output evolves under different reference data strategies.

Table 5: Comparison of reference data strategies. C^*: per-sample cost; J: total cost $(1-\eta)\cdot N\cdot C^*$. Last four columns: proportion of each protected class at round 20.

p_{ref}	η	C^*	J	frog	horse	ship	truck
Uniform	0.50	16.4	41,000	.067	.077	.059	.067
Uniform	0.20	16.4	65,600	.112	.113	.097	.105
Ours	**0.61**	**26.43**	**50,978**	**.059**	**.101**	**.116**	**.168**

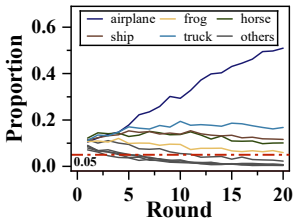

Figure 8: Class proportions under designed reference.

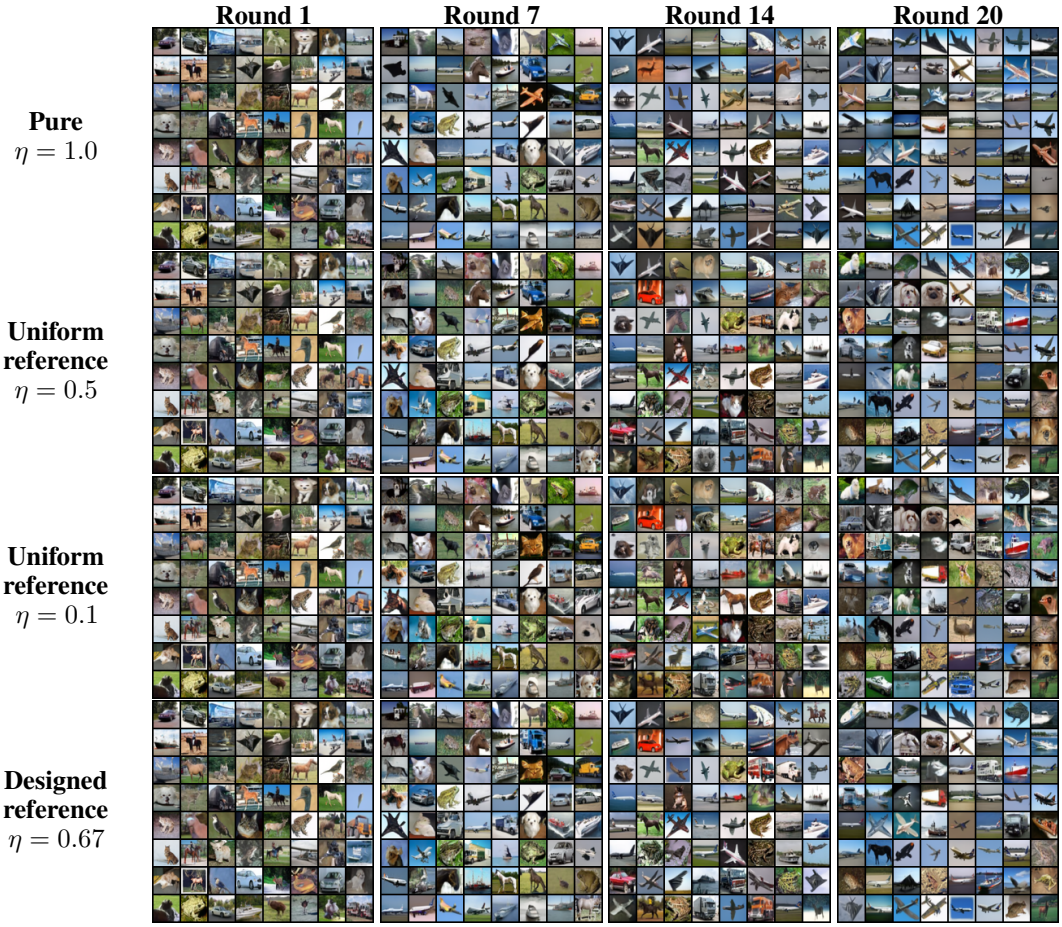

Figure 9: Generated samples from OT-CFM under $w_0 = [1, 0, 0, 0, \dots]$ (airplane preference), $\tau = 5.0$, $\beta = 0.01$. Each row corresponds to a different reference data strategy; each column shows samples at the indicated round. Under pure synthetic training ($\eta = 1.0$), the model collapses to airplane images. Mixing with uniform reference data increasingly preserves diversity. The LP-designed reference with $\theta_\ell = 0.08$ for last four classes (bottom row) maintains diversity at $\eta = 0.67$.

E.2 DDPM Experiments on CIFAR-10

Setup. The model is DDPM [17, 35] with the VLB objective (50M parameter UNet), initialized from the released pretrained checkpoint on CIFAR-10. The feature map $\varphi(x) \in \mathbb{R}^{10}$ is the normalized softmax output of a pretrained ResNet-56 classifier (accuracy $\geq 93\%$), where $\varphi(x)_c$ represents the predicted probability of class c. Each round generates 20,000 candidate images, curates 10,000 via K-way Luce selection, mixes with real CIFAR-10 at ratio η to form a training set of size 10,000, and fine-tunes for 1,000 steps. After training, 10,000 fresh samples are generated to compute per-class proportions and update w_t. We run each experiment for 10 rounds with $\tau = 5.0$, $K = \infty$, and $\beta = 0.01$.

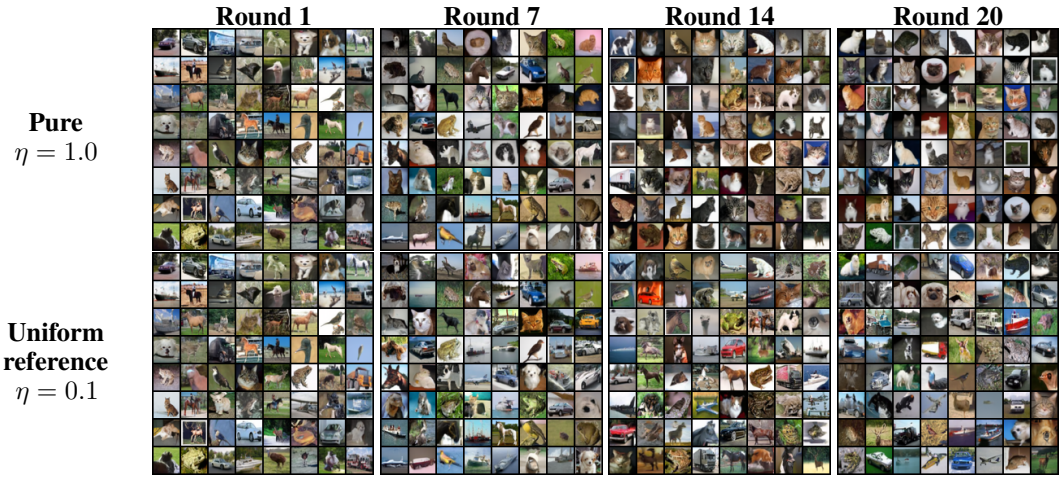

Figure 10: Generated samples from OT-CFM under $w_0 = [0, 0, 0, 1, \dots]$ (cat preference), $\tau = 5.0$, $\beta = 0.01$. Each row corresponds to a different reference data strategy; each column shows samples at the indicated round. Under pure synthetic training ($\eta = 1.0$), the model collapses to airplane images. Mixing with uniform reference data increasingly preserves diversity.

Co-evolution dynamics. Fig. 11 shows results under two initializations and two mixing ratios. Under $\eta = 1.0$, DDPM exhibits the same initialization-dependent lock-in as CFM: $w_0 = [1, 0, 0, 0, \dots]$ drives concentration toward airplane and $w_0 = [0, 0, 0, 1, 0]$ toward cats. Under $\eta = 0.1$, both initializations converge to a similar near-uniform equilibrium, consistent with the uniqueness result. Compared to CFM, DDPM shows slower concentration under $\eta = 1.0$, reaching 60% by round 10. This is expected: DDPM suffers from greater image quality degradation over successive retraining rounds, causing the classifier to assign less confident predictions. More broadly, the theoretical model assumes the retrained distribution exactly matches the training mixture, while in practice finite capacity and training error introduce deviations. Nevertheless, the qualitative predictions of the theory hold: lock-in occurs under pure synthetic training and stabilization occurs under mixture with reference data.

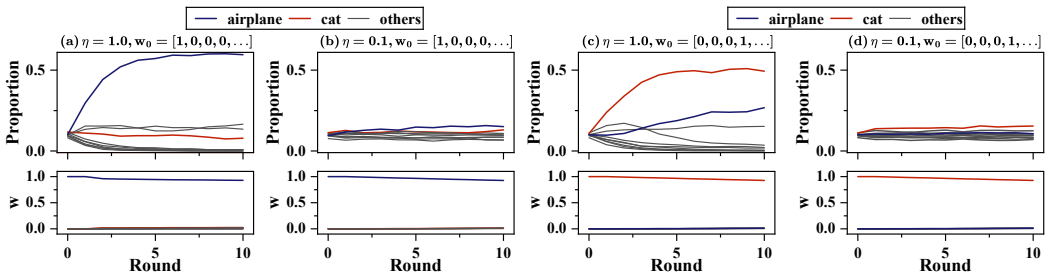

Figure 11: Co-evolution dynamics under different initializations and mixing ratios for DDPM model. **Top**: per-class generation proportions. **Bottom**: preference vector w_t.

Reproducibility across seeds. To verify that the observed dynamics are not artifacts of a particular random seed, we repeat the experiment ($\eta = 1.0$, $w_0 = [1, 0, 0, 0, \dots]$) with three independent seeds. Fig. 12 shows that all three runs produce consistent trajectories: airplane proportion reaches 58.29%-59.02% by round 10 and the preference vector follows a similar path. The small variation across seeds confirms that the co-evolution dynamics are driven by the systematic interaction between curation and retraining, rather than by stochastic fluctuations.

Generated Samples. Fig. 13 shows the corresponding generated samples for DDPM. Compared to OT-CFM, DDPM exhibits noticeable image quality degradation over successive rounds: textures become blurry and fine details are lost, reflecting the accumulation of training error across generations. This degradation is most pronounced under $\eta = 1.0$, where the model retrains entirely on its own imperfect outputs. Nevertheless, the same structural patterns persist: concentration toward a single class under pure synthetic training, and preserved diversity under reference data mixing.

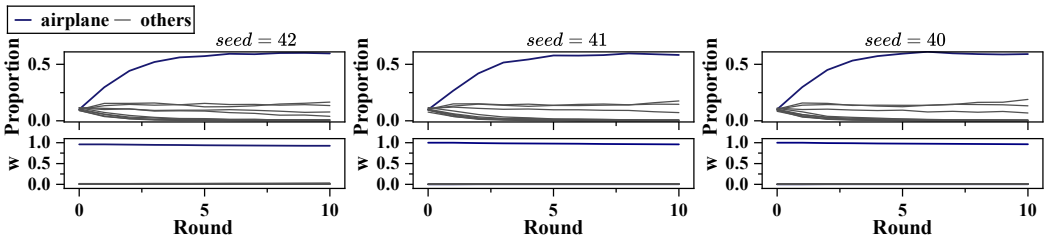

Figure 12: Reproducibility of DDPM co-evolution across three independent seeds ($\eta = 1.0$, $w_0 = [1, 0, 0, 0, \dots]$, $\tau = 5.0$, $K = \infty$, $\beta = 0.01$).

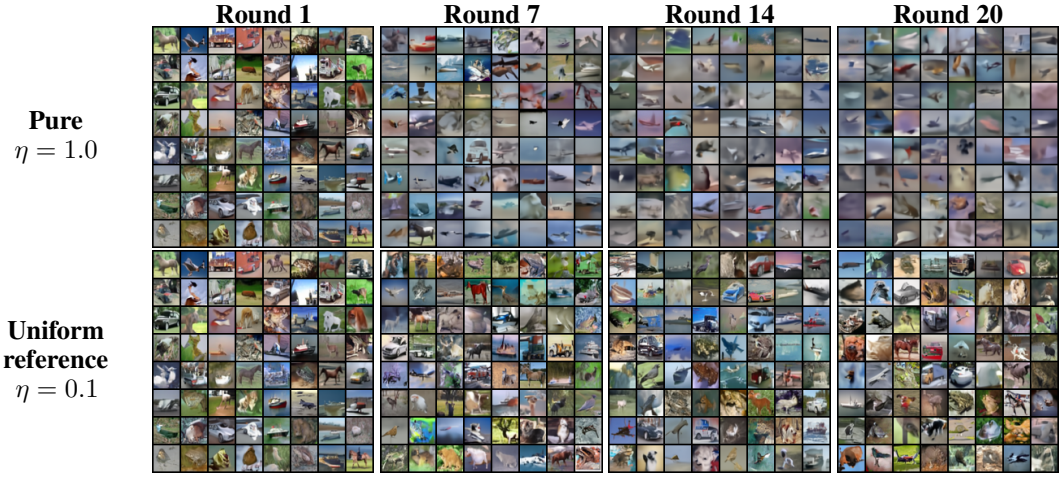

Figure 13: Generated samples from DDPM under $w_0 = [1, 0, 0, 0, \dots]$ (airplane preference), $\tau = 5.0$, $\beta = 0.01$. Each row corresponds to a different reference data strategy; each column shows samples at the indicated round. Under pure synthetic training ($\eta = 1.0$), the model collapses to airplane images. Mixing with uniform reference data increasingly preserves diversity.

E.3 GPT-2 Experiments on AG News

We replicate the co-evolution experiments on text generation.

Setup. The model used is GPT-2-small (124M parameters) [37], initialized from the HuggingFace pretrained checkpoint. The feature map $\varphi(x) \in \mathbb{R}^4$ is the normalized softmax output of a pretrained BERT [8] classifier (accuracy $\geq$ 94%), where $\varphi(x)_c$ represents the predicted probability of class $c \in \{\text{world}, \text{sports}, \text{business}, \text{sciTech}\}$.

Iterative retraining. Each round generates 5,000 candidate texts, curates 2,500 via K-way Luce selection, mixes with real AG News at ratio η to form a training set of size 2,500, and fine-tunes for 1,000 gradient steps. After training, 2,000 fresh samples are generated to compute per-class proportions and update w_t. We run each experiment for 10 rounds with $\tau = 5.0$, $K = \infty$, and $\beta = 0.01$.

Co-evolution dynamics. Fig. 14 shows results with $w_0 = [0, 1, 0, 0]$ (Sports) under four mixing ratios. At $\eta = 1.0$, the Sports proportion exceeds 90% by round 10, mirroring the lock-in observed in the image experiments. As η decreases, reference data stabilizes the distribution. Fig.15 shows the same sweep with $w_0 = [0, 0, 0, 1]$ (SciTech). The dynamics are qualitatively identical: the system locks into SciTech under $\eta = 1.0$ and stabilizes under small η. This confirms that the initialization-dependent lock-in and the stabilizing effect of reference data hold across both modalities and class structures.

Unlike the image models, GPT-2 deviates more noticeably from the idealized assumptions of our theory. Even when trained entirely on uniformly distributed reference data ($\eta = 0.0$), the generated class proportions are not uniform, indicating that GPT-2 is far from a perfect learner: the retrained distribution does not exactly match the training mixture. In addition, the class proportions fluctuate across rounds rather than converging smoothly, likely due to the instability of fine-tuning a pretrained language model on small datasets. Despite these deviations, the qualitative theoretical predictions remain: lock-in under pure synthetic training and stabilization under reference data mixing.

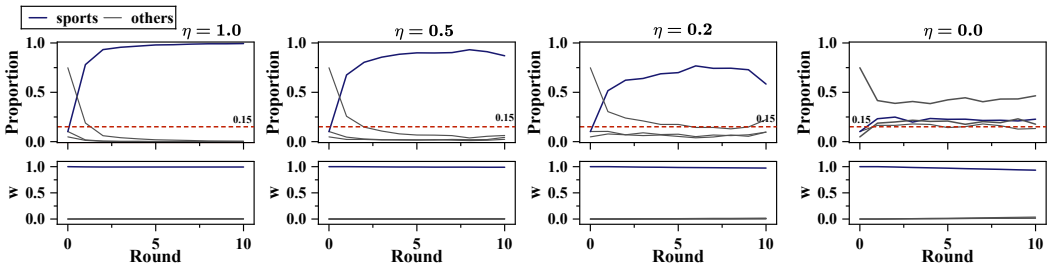

Figure 14: GPT-2 on AG News with $w_0 = [0, 1, 0, 0]$ (Sports). *Top*: per-class proportions. *Bottom*: preference vector w_t.

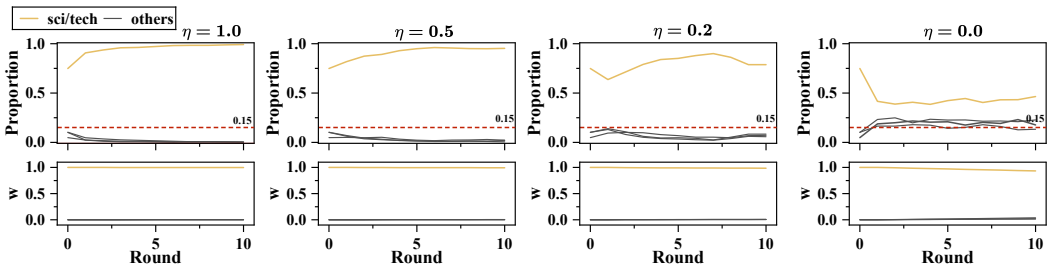

Figure 15: GPT-2 on AG News with $w_0 = [0, 0, 0, 1]$ (SciTech). *Top*: per-class proportions. *Bottom*: preference vector w_t.

Reference data design. We apply Alg. 1 to the AG News setting. We protect the class world with threshold $\theta = 0.15$. To account for the gap between the theoretical model and GPT-2's imperfect learning, we calibrate the threshold used in the LP: since GPT-2 achieves roughly 37.5% of the theoretically predicted level under uniform reference data, we scale the LP input threshold to $\theta_{LP} \approx \frac{0.15}{0.375} = 0.4$, so that the actual achieved proportion approximates the desired $\theta = 0.15$. The costs are $c = (20, 5, 5, 5)$, reflecting that World news requires more effort to source than the other categories.

Table. 6 compares our designed reference with uniform reference and the actual performance of the reference data is shown in the Fig. 16.

Table 6: Comparison of reference data strategies. C^*: per-sample cost; J: total cost $(1 - \eta) \cdot N \cdot C^*$. Last columns: proportion of the protected class at round 20.

p_{ref}	η	C^*	J	world
Uniform	0.50	8.75	10,937.50	.1211
Uniform	0.10	8.75	19,687.50	0.12
Ours	**0.10**	**11.67**	**26,250.00**	**.1404**

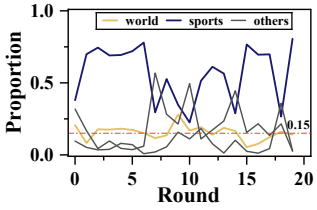

Figure 16: Class proportions under designed reference.

F Discussion and Extensions

Our framework focuses on a basic co-evolutionary setting in which a single aggregate preference state interacts with a self-consuming generative model through preference-guided curation. Building on this framework, several natural extensions arise by enriching the preference dynamics and population structure while retaining the same closed-loop interaction between user feedback and generative retraining. We discuss three such directions below.

F.1 Nonlinear Preference Responses

The preference update in Eq. 4 moves the current preference toward the mean feature vector of the updated model through linear averaging. This form is related to first-order models of preference and opinion formation. More generally, users may respond nonlinearly to the discrepancy between their current preference and the content to which they are exposed. We consider such responses while retaining the reward $r_w(x) = \langle w, \varphi(x) \rangle$, the curation operator, and the model retraining step in Eq. 3.

A gap-response formulation. Let $\Delta_t := \bar{\varphi}(p_{t+1}) - w_t$ denote the mean feature vector after retraining and its discrepancy from the current preference. For a response function $F : \mathbb{R}^d \to \mathbb{R}^d$, define

$$w_{t+1} = G_F(w_t, \bar{\varphi}(p_{t+1})) := \mathrm{Proj}_W[w_t + \beta F(\bar{\varphi}(p_{t+1}) - w_t)]\,. \tag{14}$$

The linear update is recovered by $F(\Delta) = \Delta$. We additionally consider a saturating response and a state-dependent scalar gate.

Saturating response. For $F(\Delta) = \tanh(g\Delta)$ with $g > 0$ and $\tanh$ applied componentwise, the update is

$$w_{t+1} = \mathrm{Proj}_W\left[w_t + \beta \tanh\big(g\big(\bar{\varphi}(p_{t+1}) - w_t\big)\big)\right] \tag{15}$$

This response is smooth and saturates as the feature–preference gap grows, limiting the magnitude of the response in each coordinate.

Scalar-gated response. Let $\gamma_t := \gamma\big(w_t, \bar{\varphi}(p_{t+1})\big) \in (0, 1]$ be a state-dependent response weight. Using $F_t(\Delta_t) = \gamma_t \Delta_t$ in the same update template gives

$$w_{t+1} = \mathrm{Proj}_W[(1 - \beta\gamma_t)w_t + \beta\gamma_t \bar{\varphi}(p_{t+1})]\,. \tag{16}$$

The gate acts as a similarity or confidence weight. It changes the adaptation rate according to the current interaction between preferences and model outputs. In the discussion below, G_F denotes the resulting preference-update map, including this state-dependent gated form.

Components of the analysis that remain unchanged. Changing the preference response leaves the model-update branch unchanged. Consequently, the curation properties in Prop. 3.1, the retraining sensitivity constants L_p and L_w in Lem D.7, and the equal-reward characterization of fixed points of the retraining step at a given preference in Thm. 3.2 continue to apply. These components depend on the reward and feature representation, rather than on the rule used to update preferences.

The singleton equilibria also remain fixed whenever the response vanishes at a zero gap. Indeed, at $p^\star = \delta_{x_i}$, $w^\star = \varphi(x_i)$, we have $\bar{\varphi}(p^\star) - w^\star = 0$. Thus $F(0) = 0$ implies that the preference update leaves $w^\star$ unchanged, while the model update is already fixed in the purely synthetic regime. All three response families above vanish at a zero gap.

Response-dependent sensitivity and coupled dynamics. The response function enters the joint analysis through the preference coordinate. Suppose that the resulting map satisfies

$$\big\|G_F(w, m) - G_F(v, \widetilde{m})\big\| \le a_F\|w - v\| + b_F\|m - \widetilde{m}\|$$

for admissible preference vectors and mean feature vectors, with $a_F, b_F \ge 0$. For two trajectories (p_t, w_t) and (q_t, v_t), the preference-coordinate estimate becomes

$$\|w_{t+1} - v_{t+1}\| \le a_F\|w_t - v_t\| + b_F\big\|\bar{\varphi}(p_{t+1}) - \bar{\varphi}(q_{t+1})\big\|$$

For the linear response, nonexpansiveness of projection gives $(a_F, b_F) = (1 - \beta, \beta)$, recovering the estimate used in the proof of Thm. 3.9. For another response, the same proof strategy replaces $(1 - \beta, \beta)$ with (a_F, b_F) and combines the corresponding preference-coordinate estimate with the unchanged retraining bounds. The admissible weight in the product metric, the contraction condition, and the convergence rate consequently depend on the response. Quantitative lock-in neighborhoods likewise require response-specific control of the preference update. This formulation identifies the part of the analysis that must be adapted, while preserving the curation and retraining components. Deriving explicit bounds for particular nonlinear responses is a direction for further analysis.

F.2 Stochastic Preference Adaptation

Another natural extension is to model preference changes at the individual level as stochastic rather than deterministic. For example, suppose a user updates toward the feature vector of a randomly observed model output, $X_{t+1} \sim p_{t+1}$, $w_{t+1} = \mathrm{Proj}_W[(1 - \beta)w_t + \beta\varphi(X_{t+1})]$. Under the assumptions $\|w_t\| \le 1$, $\|\varphi(x)\| = 1$, and $\beta \in [0, 1]$, the point before projection already lies in W. Therefore,

$$\mathbb{E}\left[w_{t+1} \mid p_{t+1}, w_t\right] = (1 - \beta)w_t + \beta\bar{\varphi}(p_{t+1}),$$

which recovers Eq. 4 as the conditional-mean dynamics. This provides a mean-field interpretation of the deterministic preference update studied in the paper.

The stochastic formulation raises additional questions that do not arise at the mean level. In particular, one may ask whether stochastic trajectories concentrate around the deterministic dynamics, how random exposure affects the probability of entering different lock-in basins under pure self-consumption, and whether reference mixing induces stochastic stability or a concentrated stationary distribution around the deterministic equilibrium. More general noise models, including exogenous perturbations to the preference state, would further require quantifying how random fluctuations propagate through the coupled model–preference feedback loop.

F.3 Heterogeneous User Populations

The current framework represents the population by a single aggregate preference vector. A natural extension is to consider G user groups with group-specific preferences $\{w_{g,t}\}_{g=1}^{G}$ and population weights $\{\alpha_g\}_{g=1}^{G}$, where $\alpha_g \geq 0$ and $\sum_{g=1}^{G} \alpha_g = 1$. One possible model replaces the single curated distribution by a mixture of group-specific curation responses,

$$p_{t+1}(x) = (1-\eta)p_{\mathrm{ref}}(x) + \eta p_t(x) \sum_{g=1}^{G} \alpha_g H^K_{p_t, w_{g,t}}(x),$$

while each group updates according to

$$w_{g,t+1} = \mathrm{Proj}_W \left[(1-\beta_g)w_{g,t} + \beta_g \bar{\varphi}(p_{t+1})\right].$$

This extension introduces qualitatively new interactions because all groups affect, and are simultaneously affected by, the same evolving generative distribution. It therefore raises questions about whether group preferences converge or remain separated, how the feedback of one group indirectly changes the preferences of others through the shared model, and how reference-data design should be adapted when different groups impose distinct preservation objectives. Characterizing the equilibria and stability of such multi-group dynamics, as well as designing reference distributions that balance group-specific objectives, is a promising direction for future work.

G Broader impacts

This work studies feedback loops in self-consuming generative systems, where model outputs are curated by users and reused for future training, while user preferences also evolve over time. This work can help diagnose and mitigate long-term instability, preference drift, and diversity collapse by our theoretical and empirical tools. However, the same steering mechanism is value-neutral with respect to the choice of preserved attributes, and could in principle be used to filter out specific content. Thus, transparent selection of preserved attributes is important in any deployment.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction state the main contribution: a coupled model of self-consuming generative retraining with co-evolving preferences, theoretical convergence and equilibrium results, and a reference-data design algorithm. The claims are scoped to the stated dynamical model and are supported by theoretical analysis and empirical illustrations.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The conclusion discusses limitations and future directions, including partial observability of the feature map and preferences, nonlinear rewards, and extensions to multi-platform settings with heterogeneous user populations.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: The paper states the assumptions in Sec.2. The complete proofs of the main theoretical statements provided in Appx. C and extensions discussed in Appx. D.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Sec. 5 describes key experimental settings (e.g., datasets, models, retraining procedure, and hyperparameters). Additional experimental details, parameter sweeps, and repeated runs are provided in Appx. E.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The paper uses public datasets and public model implementations. We provide experimental details sufficient to reproduce the reported results and provide the link to the code.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.

- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Sec. 5 and Appx. E specifies details in the experiments. The Additional parameter studies are reported in Appx. E.2.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: The paper reports representative closed-loop trajectories rather than averaged curves, because averaging different dynamical paths can obscure behavior. To assess robustness, Appx. E.2 reports three independent repeated runs under the same setting, and the main experiments use different random seeds across rounds.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
   - If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: The appendix reports the compute environment used for the experiments.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

Answer: [Yes]

Justification: The work uses public datasets and pretrained models for theoretical and empirical analysis of generative-model feedback loops. There is no violation of the code of ethics.

Guidelines:

- The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: The paper discusses both positive impacts and potential misuse, motivating transparency and oversight.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The paper does not release new high-risk pretrained models, image generators, or scraped datasets. The experiments use existing public datasets and publicly available model implementations.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: The paper cites the datasets, model architectures, pretrained classifiers, and public repositories used in the experiments.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [N/A]

Justification: The paper does not introduce a new dataset, benchmark, or pretrained model as a primary asset. The proposed algorithm and experimental procedure are described in the paper and appendix.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: The paper does not involve crowdsourcing experiments or human-subject studies. User preferences are modeled mathematically through a preference vector and simulated through the proposed dynamical system.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.