The Algorithmic Liquidity Trap: High-Velocity Swiping vs. Deep Engagement in Short-Video Platforms*

Qianzi Zhu†

Shenghao Zhu‡

Preliminary

First version: December 2025
This version: December 26, 2025
[latest version]

Abstract

Short-video platforms balance user retention with high-frequency impression generation. We model this ecosystem using a Mean-Field Stackelberg Game with "traffic recycling," where users constantly re-enter the content pool. Solving the coupled HJB and Non-Local Fokker-Planck equations via a novel Spectral Soft Actor-Critic algorithm, we identify a critical "Algorithmic Liquidity Trap." While unconstrained optimization mathematically maximizes swipe flux, it drives dwell times to a physiological floor ($\tau \approx 0.38s$). We term this "Phantom Liquidity"—high throughput with zero effective economic value (violating IAB standards). Furthermore, we reveal a "Dopamine Treadmill" effect: improved content quality paradoxically increases user impatience, explaining the difficulty of sustaining long-form content. Our results demonstrate that without explicit constraints—such as weights for completion rates and creator interactions—algorithmic maximization inevitably collapses the ecosystem, validating the necessity of current industry governance structures.

Keywords: *short-video platforms, mean-field games, traffic recycling, deep reinforcement learning, algorithmic liquidity trap, platform governance.*
JEL Classification: *L86; C73; M37.*

*We acknowledge the financial support from the National Natural Science Foundation of China (Grant No. 72450003).

†University of International Business and Economics. Email: zhuqianzi836@gmail.com
‡University of International Business and Economics. Email: zhus02755@uibe.edu.cn

1 Introduction

In the digital attention economy, short-video platforms such as TikTok, Instagram Reels, and YouTube Shorts have established a dominant paradigm of content consumption: the infinite scroll. Unlike traditional media (e.g., Netflix) that monetizes user retention via subscriptions, or search engines (e.g., Google) that monetize specific intent, short-video platforms operate on a unique hybrid model. They generate revenue through a delicate balance of two conflicting metrics: *duration* (total time spent on the app) and *flux* (the frequency of exposure to new content, primarily ads). This creates a fundamental operational trade-off: To maximize ad inventory, the platform is incentivized to accelerate the "swiping" velocity of users. However, indiscriminate acceleration risks decoupling algorithmic optimization from physiological reality, potentially leading to a state of hollow engagement. The central managerial question, therefore, is: *How should a platform regulate the granular flow of user attention to avoid the "race to the bottom" and balance immediate impression yield against long-term ecosystem sustainability?*

Existing theoretical frameworks in operations management and economics struggle to fully capture the dynamics of this ecosystem due to a structural disconnect between micro-level behavior and macro-level flows. The classical literature on user behavior, rooted in Optimal Stopping Theory (Peskir & Shiryaev, 2006; Stokey, 2008), typically treats the decision-maker as facing a static environment—once a user exercises the option to "stop" (or switch), they exit the system. However, the short-video ecosystem is characterized by a closed-loop mechanism we term *"Traffic Recycling."* When a user swipes away from a video, they do not leave the platform; rather, they are immediately "recycled" back into the content pool to face a new draw. This recycling mechanism implies that the platform's algorithmic interventions do not merely affect a single viewing session but fundamentally reshape the macroscopic equilibrium of the entire user population.

To bridge this gap, this paper develops a *Mean-Field Stackelberg Game (MFSG)* framework that explicitly models the interplay between the platform's dual-objective optimization and the users' continuous-time stopping decisions. We model the platform as a Stackelberg Leader that regulates the ecosystem through two levers: *Retention Control* (micro-interventions to extend the duration of the current video) and *Matching Control* (investment to improve the quality of the next video). Users, acting as rational Followers, observe these macroscopic controls and solve an optimal stopping problem with an endogenous threshold. Mathematically, this leads to a novel coupled system comprising a Hamilton-Jacobi-Bellman (HJB) equation governing user strategies and a *Non-Local* Fokker-Planck equation describing the population evolution with a source term induced by traffic recycling. We solve this high-dimensional control problem using a physics-informed Deep Reinforcement Learning (DRL) algorithm, calibrated with large-scale randomized experiment data (KuaiRand) to ensure empirical realism.

Our analysis yields three distinct insights that challenge the blind pursuit of engagement metrics. First, we identify a *"Dopamine Treadmill"* effect. One might intuitively expect that improving recommendation quality would satisfy users and prolong their stay on in-

dividual videos. Contrary to this, we find that a higher quality content pool raises the users' rational expectation of the "next best alternative," thereby increasing their optimal switching threshold. This paradox explains the structural difficulty of sustaining *long-form content* or deep engagement within a pure feed architecture: better matching algorithms inadvertently condition users to be more impatient.

Second, and most critically, we uncover a regime we term the *"Algorithmic Liquidity Trap."* Traditional models often assume that maximizing swipe flux correlates with maximizing value. However, our general equilibrium analysis reveals that under an unconstrained ad-centric strategy, the system converges to a *"High-Velocity Saturation"* point. While mathematically maximizing the traffic recycling loop, this equilibrium drives the average dwell time to a physiological floor ($\tau \approx 0.38s$). We define this state as *"Phantom Liquidity"*—a condition of high system throughput but zero effective economic value (violating standard viewability thresholds, e.g., IAB standards). This finding serves as a counterfactual proof, validating the necessity of current industry practices that strictly enforce constraints on viewability and dwell time.

Third, we characterize the structure of the optimal algorithmic intervention, which exhibits a non-linear *"Retention Trap."* We find that the optimal policy applies peak intervention intensity only when the user's attention approaches the critical churn threshold. This "Trap and Release" strategy exploits the users' decision boundary to maximize ad extraction. While effective in a static sense, this aggressive extraction creates a "Tragedy of the Commons" risk for the creator ecosystem, further underscoring the need for governance mechanisms—such as algorithmic weighting for *completion rates* and *social interactions*—to align platform incentives with sustainable user value.

The remainder of this paper is organized as follows. Section 2 reviews the related literature. Section 3 formulates the Mean-Field Stackelberg Game model. Section 4 details the structural calibration using the KuaiRand dataset. Section 5 presents the main equilibrium analysis and the "Algorithmic Liquidity Trap." Section 6 concludes with managerial implications for platform governance.

2 Literature Review

Our work sits at the intersection of three streams of literature: the economics of rational attention (optimal stopping), algorithmic operations in recommender systems, and mean-field game theory.

2.1 Rational Attention and Optimal Stopping

The micro-foundation of our model builds upon the classical theory of optimal stopping under uncertainty. The seminal works of Stokey (2008) and Peskir and Shiryaev (2006) established the mathematical rigor for single-agent stopping problems driven by geometric Brownian motions. In the context of the digital economy, this framework has been adapted to model online search behavior. For instance, Gershkov and Moldovanu (2009) analyze

consumer search with learning, while Ke, Shen, and Villas-Boas (2016) propose a Bayesian framework for rapid information consumption.

However, these classical models typically assume an "absorbent" exit state—once a user stops (or purchases), they leave the system. Our work extends this by introducing the concept of *"Traffic Recycling,"* where the "stopping" action (swiping) is not an exit but a transition back into the content pool. This requires a non-local boundary condition that couples the exit flux of one state to the entry flux of another, a dynamic absent in standard search theory.

2.2 Algorithmic Management of Attention in Recommender Systems

The second stream focuses on the operational design of content platforms. In the computer science domain, early works prioritized binary metrics such as Click-Through Rate (CTR). More recently, industrial frameworks like those at YouTube (M. Chen et al., 2019; Covington, Adams, & Sargin, 2016) have shifted towards continuous engagement metrics (e.g., watch time) via Deep Reinforcement Learning.

However, a theoretical gap persists in the Operations Management (OM) and Economics literature regarding the *objective function* of these algorithms. Traditional models in online advertising and revenue management typically posit a *linear relationship* between impression flux and revenue (e.g., Goldfarb & Tucker, 2011; Lobel, Patel, Vulcano, & Zhang, 2016). The implicit assumption is that maximizing the throughput of users (i.e., swipe frequency) monotonically increases ad inventory value, effectively treating attention as a standard divisible good.

Our work challenges this "Linear Flux Hypothesis." We argue that in short-video ecosystems, the value function is a *step function* governed by viewability thresholds (e.g., the IAB 2-second standard). By incorporating this discontinuity, we distinguish between *nominal flux* (server load) and *effective economic flow*, a nuance often overlooked in standard inventory maximization models. This allows us to theoretically characterize the phenomenon of "Phantom Liquidity"—a regime where standard OM policies fail by optimizing for empty calories.

2.3 Mean-Field Games in Platform Economies

Finally, we contribute to the emerging application of Mean-Field Games (MFG) in economics. While traditional heterogeneous-agent frameworks in macroeconomics (Aiyagari, 1995) or cooperative control approaches like potential games (Marden, Arslan, & Shamma, 2009) offer valuable insights for static or discrete-time equilibrium, they often face the "curse of dimensionality" when handling continuous-time attention flows with complex recycling dynamics. The adoption of the MFG framework is particularly natural in the context of short-video platforms due to the sheer scale of the user base. Platforms like TikTok and Instagram host billions of active users, rendering traditional N-player game theoretical models computationally intractable. As the number of agents $N \to \infty$, individual idiosyncratic noises average out, allowing us to approximate the complex high-

dimensional system with a tractable interaction between a representative agent and the aggregate population density (Lasry & Lions, 2007).

Since its foundation, MFG has been applied to systemic risk, crowd dynamics, and macroeconomics (Achdou, Han, Lasry, Lions, & Moll, 2022). Recently, Iyer, Johari, and Sundararajan (2014) and Yang et al. (2018) have applied mean-field approximations to large-scale information systems.

Our work introduces a distinct technical novelty to this field: the *Mean-Field Stackelberg Game with Recycling*. Unlike standard MFGs where the population mass is conserved via a simple transport equation or vanishes at the boundary, our model features a "re-entrant" mechanism governed by a Non-Local Fokker-Planck equation. This structure allows us to capture the "velocity" of user flow—a critical metric in short-video platforms that standard MFG models ignore.

3 The Model

We formulate the interaction between the short-video platform and its massive user base as a Mean-Field Stackelberg Game. The platform (Leader) commits to an algorithmic policy to maximize long-term rewards, while individual users (Followers) observe the ecosystem's state and optimally decide when to "swipe" to the next video.

3.1 The Micro-Ecosystem: Attention Dynamics and Retention Control

Scope of Analysis: The Demand-Side Loop. To isolate the mechanism of the "Algorithmic Liquidity Trap," we focus our analysis on the closed-loop interaction between the *Platform's Algorithmic Policy* and the *User's Attention Dynamics*. We treat the content supply side (creators) as an instrument controllable by the platform via the matching variable $u(t)$, which shifts the quality distribution $g(z; u)$. While we acknowledge that creators inevitably react to algorithmic incentives in the long run (a supply-side response), explicitly modeling the creator game would introduce significant tractability issues without altering the fundamental existence of the liquidity trap driven by user physiology. We thus abstract away strategic creator behavior to focus on the immediate feedback loop between algorithmic acceleration and cognitive capacity.

Assumption (The Infinite Feed Paradigm): We restrict our analysis to the *Feed-based Infinite Scroll* mechanism (e.g., TikTok's "For You" page), characterized by passive consumption and high-frequency switching. We distinguish this from *search-based* or *long-form* consumption (e.g., Netflix), which follows different decision dynamics.

The dynamics of the user population are driven by the interplay between continuous engagement (viewing) and discrete resampling events (swiping). We first describe the continuous phase.

Controlled Attention Process. Consider a continuum of ex-ante homogeneous users indexed by $i \in [0, N]$. Let $Z_t^i \in \mathbb{R}_+$ denote the instantaneous attention level (or physiological

arousal) of user i at time t during the viewing of a single video. We model Z_t^i as a controlled Geometric Brownian Motion (GBM):

$$\mathrm{d}Z_t^i = \left(\mu + a(t, Z_t^i)\right) Z_t^i \mathrm{d}t + \sigma Z_t^i \mathrm{d}W_t^i, \tag{1}$$

where W_t^i are independent standard Brownian motions capturing the intrinsic volatility of content stimuli.

- **Natural Decay ($\mu < 0$):** The parameter μ represents the baseline rate of attention decay (i.e., the onset of boredom). In the absence of stimuli, user attention naturally drifts toward zero, reflecting the entropy of the swiping mechanism.
- **Algorithmic Modulation ($a(t, z)$):** The control variable $a(t, z) \in [a_{\min}, a_{\max}]$ (where $a_{\min} < 0$) represents the platform's *intervention intensity*. Unlike traditional editorial editing, this corresponds to bidirectional *Sensory Modulation*:
  - **Positive Stimulation ($a > 0$):** Using real-time features (e.g., visual hooks, intense audio cues) to artificially inject a positive drift that *counteracts* natural decay (Retention).
  - **Negative Nudging ($a < 0$):** Strategically reducing stimuli or withholding gratification to *accelerate* decay. This allows the algorithm to proactively "flush" users out of low-yield states to trigger a new swipe (Impression), serving as a micro-foundation for the Algorithmic Liquidity Trap.

We denote the effective drift as $\tilde{\mu}(t, z) = \mu + a(t, z)$. This SDE governs the user's path *only* until an endogenous stopping time τ^*. Upon hitting the churn boundary $b^*(t)$, the user exits the current path and triggers a "swipe" event, governed by the traffic recycling mechanism.

Modeling Remark: From Agents to Fields. In defining the macroscopic evolution, we adopt a "top-down" Mean-Field approach. While methods like heterogeneous-agent models (Aiyagari, 1995) are valuable in macroeconomics, they often rely on static equilibrium concepts. Given our focus on the *velocity* of user flows and the *recycling* dynamics, we model the aggregate state directly via a Mean-Field PDE. Crucially, our ecosystem is an *open system with internal recycling*, governed by a *Non-Local Fokker-Planck Equation*. This structure allows us to capture the complex feedback between the churn boundary b^* and the re-entry distribution $g(z)$—a necessary feature to model the "Algorithmic Liquidity Trap."

3.2 The User's Problem: Optimal Stopping with Endogenous Thresholds

Given the controlled attention dynamics described in Equation (1), the user solves an optimal stopping problem to determine the precise moment to "swipe" away from the current video.

Value Function and the HJB Equation. The user's objective is to maximize total discounted utility. The value function $v(t, z)$ is defined as the supremum over all admissible stopping times τ:

$$v(t,z) = \sup_{t \le \tau < \infty} \mathbb{E}\Bigg[\int_t^\tau e^{-\rho(s-t)} u_{util}(Z_s) \mathrm{d}s \\ + e^{-\rho(\tau - t)} \left(K(t, u(\tau)) - C \right) \Bigg| Z_t = z \Bigg]. \tag{2}$$

We employ a CRRA utility function $u_{util}(z) = \frac{z^{1-\gamma}}{1-\gamma}$ (with $\gamma > 1$). The term C represents the cognitive switching cost.

Crucially, the payoff upon switching, $K(t, u)$, is not fixed but depends on the platform's concurrent *Matching Control* $u(t)$. This constitutes a *Variational Inequality (VI)* for the value function:

$$\min \Bigg\{ \rho v - \frac{\partial v}{\partial t} - \mathcal{L}^a v - u_{util}(z), \\ v(t,z) - \big(K(t, u(t)) - C \big) \Bigg\} = 0, \tag{3}$$

where $\mathcal{L}^a$ is the generator of the controlled diffusion:

$$\mathcal{L}^a v = (\mu + a(t,z)) z \frac{\partial v}{\partial z} + \frac{1}{2} \sigma^2 z^2 \frac{\partial^2 v}{\partial z^2}. \tag{4}$$

The *optimal stopping boundary* (or Swiping Threshold) $b^*(t)$ is the free boundary separating the continuation region from the stopping region:

$$b^*(t) = \sup \left\{ z \in \Omega \mid v(t,z) = K(t, u(t)) - C \right\}. \tag{5}$$

Users with attention $z \le b^*(t)$ immediately execute a "swipe".

The Endogenous Outside Option (Coupling Variable). The variable $K(t, u)$, which serves as the boundary condition for the user's problem, represents the expected value of the "next video". In our model, this creates a *Fixed-Point problem*: the value of the outside option depends on the value function itself. Specifically, K is determined by the platform's investment $u(t)$, which shapes the content quality distribution $g(z; u)$:

$$K(t, u(t)) = \int_{z_{\min}}^{z_{\max}} v(t, z') g(z'; u(t)) \mathrm{d}z'. \tag{6}$$

This equation highlights the *Dopamine Treadmill* effect: An increase in platform matching investment $u(t)$ shifts $g(z; u)$ towards higher quality. This raises the expected switching value K, which in turn raises the user's stopping threshold $b^*(t)$. Paradoxically, better recommendation algorithms may make users *more* impatient, as the opportunity cost of watching mediocre content increases. This mechanism mathematically drives the acceler-

ation of swiping velocity, serving as the micro-foundation for the "High-Velocity Saturation" regime observed in our results.

Theoretical Well-posedness

Before characterizing the equilibrium, we establish the theoretical well-posedness of the user's sub-problem defined by the QVI in (3).

Proposition 1 (Well-posedness of the User QVI). *Given a fixed platform control policy $\{a(t,z), u(t)\}$, the coupled system (Equation (3) and the coupling definition K) admits a unique viscosity solution $v(t,z)$ (Fleming & Soner, 2006). The map associated with the value function iteration is a contraction, stemming from the implicit obstacle structure characteristic of impulse control problems.*

Proposition 2 (Threshold Structure). *For the CRRA utility function with $\gamma > 1$, the value function $v(t,z)$ is monotonically increasing in z. Consequently, the optimal stopping region is connected and characterized by a unique threshold $b^*(t)$:*

$$\mathcal{S}_t = [z_{\min}, b^*(t)]. \tag{7}$$

Proof. The result follows from the monotonicity of the generator for one-dimensional diffusions. Detailed proof is provided in Appendix A.1.

This structural property is computationally significant: it allows us to search for a single scalar $b^*(t)$ (e.g., via bisection) rather than evaluating the stopping decision for every grid point, reducing the complexity from $O(N_z)$ to $O(\log N_z)$.

3.3 Baseline Analysis: The "Organic" Equilibrium

Before solving the full Stackelberg game with active algorithmic intervention, we simulate the user subgame under a neutral policy ($a = 0, u = u_0$). This analysis is critical not only to validate structural stability but to establish an *"Organic Benchmark"*—representing the natural state of user attention before it is distorted by the Algorithmic Liquidity Trap.

To ensure empirical relevance, we employ parameters calibrated from the KuaiRand dataset (Section 4). The equilibrium is computed using the *Fully Implicit Finite Difference Solver* (Algorithm 1 in Appendix).

3.3.1 Equilibrium Properties: The Healthy Baseline

Figure 1 presents the numerical solution. The results confirm that the system converges to a unique *Stationary Mean-Field Equilibrium (MFE)*.

Stationary Strategy and Patience (Left Panel). The optimal stopping boundary $b^*(t)$ converges to a constant threshold $b^*_\infty \approx 0.68$. The strict positivity of this threshold ($b^* > z_{\min}$) confirms that rational users exercise their "option to switch" strategically.

Logistic Growth and Flux Conservation (Middle Panel). The macroscopic dynamics reveal two structural insights that contrast sharply with the high-velocity regime discussed later:

- **Open System Dynamics:** The total active population $N(t)$ stabilizes at $N^* \approx 70.9$. This is constrained by the natural churn rate ϵ, reflecting the platform's "carrying capacity" under organic conditions.
- **Validating the "Healthy Heartbeat":** The swipe flux $\Phi(t)$ converges to a constant rate $\Phi^* \approx 5.6$ swipes/sec. Crucially, this implies an *average dwell time* of[1]:

$$\tau_{base} = \frac{N^*}{\Phi^*} \approx \frac{70.9}{5.6} \approx 12.66 \text{ seconds.} \tag{8}$$

  Unlike the "Phantom Liquidity" regime ($\tau \approx 0.38s$) identified under algorithmic maximization, this baseline flux represents *High-Quality Inventory* that comfortably satisfies viewability standards (e.g., the IAB/MRC standard of > 2s; see Media Rating Council (MRC) and Interactive Advertising Bureau (IAB) 2015). It demonstrates that the "physiological cliff" is not an inherent property of the user, but an induced state by the algorithm.

Distributional Convergence (Right Panel). The heatmap confirms the formation of a stationary distribution $f_\infty(z)$. The high-density "engagement core" (yellow region) shows that without aggressive intervention, user attention naturally settles into a sustainable rhythm of consumption and switching.

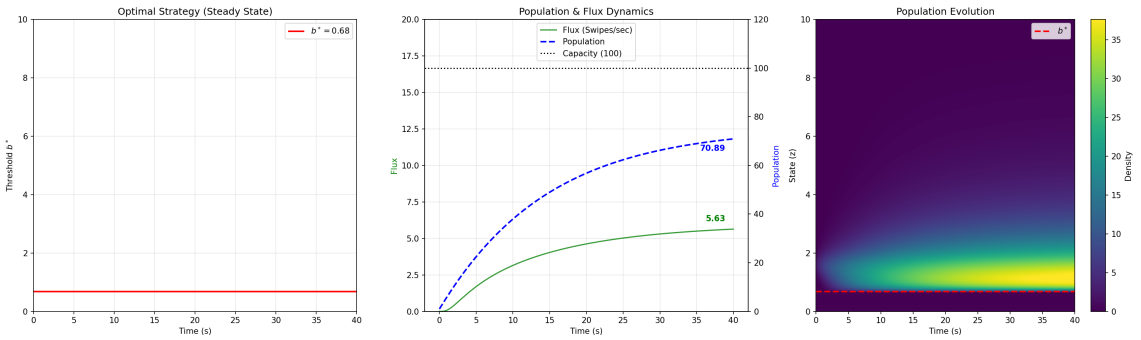

Figure 1: **Baseline Organic Equilibrium.** Under neutral policies, the system stabilizes at a healthy dwell time of $\tau \approx 12.7s$. (Left) Stationary stopping boundary. (Middle) Logistic growth of population and stable flux. (Right) Stationary distribution of attention.

3.3.2 Comparative Statics: Mechanisms of Sensitivity

To probe the sensitivity of the equilibrium, we conduct comparative statics along two dimensions: algorithmic retention intensity (μ) and content ecosystem quality (μ_g). The re-

[1]This relationship is derived from *Little's Law* (Little et al., 1961) in queueing theory, which states that the long-term average number of customers in a stationary system (N^*) is equal to the long-term average effective arrival rate (Φ^*) multiplied by the average time a customer spends in the system (τ).

sults, illustrated in Figure 2, reveal two critical phenomena: "Recycling Stagnation" and the "Dopamine Treadmill."

Experiment A: Algorithmic Intensity and Recycling Stagnation. We vary the drift parameter μ to simulate "Low Stickiness" ($\mu = -0.15$), "Baseline" ($\mu = -0.0235$), and "High Stickiness" ($\mu = 0.2$).

- **The Lock-in Effect (Top-Left):** Strong positive reinforcement ($\mu = 0.2$) *lowers* the optimal stopping threshold ($b^* \approx 0.63$). Users become "locked in," as the immediate gratification dominates the option value of future matches.
- **The Stagnation Risk (Bottom-Left):** Crucially, excessive individual-level stickiness triggers a macroscopic collapse. The "High Stickiness" scenario results in the *lowest* steady-state population ($N \approx 36.2$). *Mechanism - Recycling Stagnation:* The ecosystem relies on *Traffic Recycling* for growth. Excessive retention ($\mu = 0.2$) stifles the swipe flux needed to recycle users back into the pool. This *stagnation* (low flux) stands in direct contrast to the "Phantom Liquidity" (excessive flux) trap we identify later, highlighting that the platform must navigate a narrow channel between stagnation and overheating.

Experiment B: Content Ecosystem Maturity. We shift the mean content quality μ_g to simulate "Low Quality" (1.3), "Baseline" (1.65), and "High Quality" (2.0) ecosystems.

- **The Dopamine Treadmill (Top-Right):** Improving the content pool ($\mu_g = 2.0$) strictly *raises* the reservation utility ($b^* \approx 0.73$). *Economic Interpretation:* This reflects an *"Inflation of Expectations."* As the environment improves, users rationally internalize the higher expected value of the "next video" (K). They become more selective, requiring a higher utility threshold to justify staying on the current item.
- **Active Searching & Time Compression (Bottom-Right):** Consequently, the High Quality ecosystem generates the *highest* traffic flux ($\Phi \approx 5.70$). *Mechanism:* High quality encourages *Active Searching*. Users swipe more frequently to maximize utility per unit of time, effectively compressing the average dwell time τ. This mechanism explains why *long-form content* struggles to survive in high-quality feed environments without explicit algorithmic subsidies: better content paradoxically makes users more impatient.

3.4 The Macro-Dynamics: Non-Local Fokker-Planck Equation

As the population size $N \to \infty$, the aggregate behavior of users is described by the probability density function $f(t, z)$. The evolution of $f(t, z)$ is governed by the conservation of probability mass, adjusted for inflows (acquisition and recycling) and outflows (true churn).

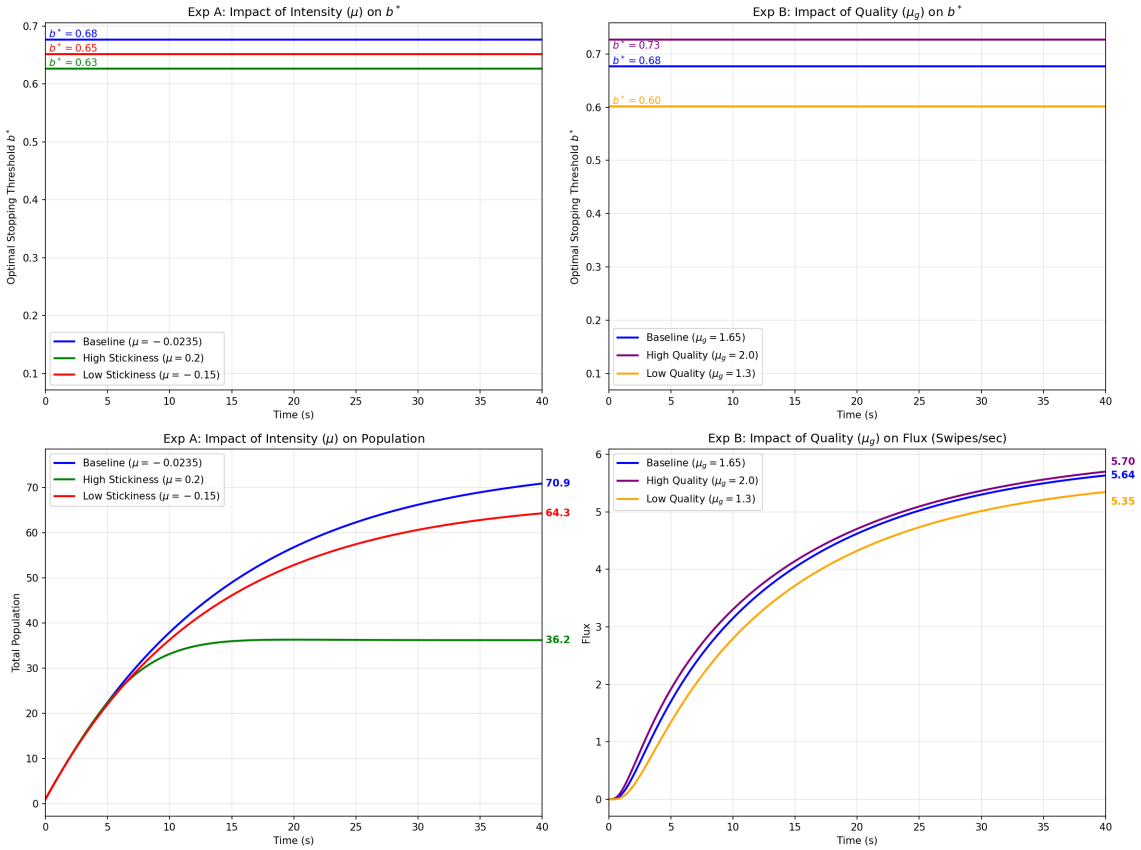

Figure 2: **Comparative Statics. (Left Column):** Excessive stickiness ($\mu = 0.2$) causes **Recycling Stagnation**, stifling population growth. **(Right Column):** Improving content quality ($\mu_g = 2.0$) triggers the **Dopamine Treadmill**, raising user expectations (b^*) and making them more impatient.

Derivation of the Non-Local PDE. Standard arguments involving Itô's formula and integration by parts (see Gardiner et al., 2004) yield the differential operator $\mathcal{L}^*$ adjoint to the generator $\mathcal{L}^a$. However, unlike standard Fokker-Planck equations that assume mass conservation, our ecosystem acts as an open system with internal loops. The governing equation is thus a *Non-Local Partial Differential Equation* defined on the dynamic domain $\Omega_t = [b^*(t), z_{\max}]$:

$$\begin{aligned}\frac{\partial f}{\partial t} = & \underbrace{\mathcal{L}^* f(t,z)}_{\text{Gaze Dynamics}} + \underbrace{A(t)\delta(z - z_p)}_{\text{External Acquisition}} \\ & + \underbrace{(1-\epsilon)\Phi_{swipe}(t) g(z; u(t))}_{\text{Traffic Recycling}},\end{aligned} \tag{9}$$

where the adjoint operator is:

$$\mathcal{L}^* f = -\frac{\partial}{\partial z}\left[(\mu + a(t,z))zf\right] + \frac{1}{2}\frac{\partial^2}{\partial z^2}\left[\sigma^2 z^2 f\right]. \tag{10}$$

Interpretation of Source Terms. Equation (9) distinguishes our model from standard MFG frameworks through two distinct source mechanisms:

1. **External Acquisition (Dirac Source):** The term $A(t)\delta(z - z_p)$ represents the injection of new users. The control variable $A(t)$ corresponds to marketing spend (akin to the Bass, 1969 diffusion model), while the Dirac delta $\delta(z - z_p)$ signifies a standardized onboarding process.

2. **Traffic Recycling (The Feedback Loop):** The term $(1-\epsilon)\Phi_{swipe}(t)g(z; u(t))$ captures the closed-loop nature of short-video platforms. When a user swipes, a fraction $(1-\epsilon)$ immediately re-enters the cycle. *Theoretical Note:* This non-local coupling is the mathematical root of the *"Algorithmic Liquidity Trap."* Since the platform can boost population density $f(t,z)$ by accelerating the recycling flux $\Phi_{swipe}(t)$, unconstrained optimization algorithms may aggressively drive users to swipe (increasing Φ) rather than extending their gaze (retention), inadvertently creating a high-velocity, low-value ecosystem.

Boundary Conditions and Flux Dynamics. The system is subject to the following boundary conditions:

$$f(0,z) = f_0(z), \quad \text{(Initial Distribution)} \tag{11}$$

$$J(t, z_{\max}) = 0, \quad \text{(Reflecting Upper Boundary)} \tag{12}$$

$$f(t, b^*(t)) = 0. \quad \text{(Absorbing/Swiping Boundary)} \tag{13}$$

The reflecting boundary at $z_{\max}$ implies an *Attention Saturation* ceiling (Lanham, 2006). The condition at $b^*(t)$ represents the swiping threshold. Crucially, while the density vanishes at $b^*(t)$, the probability flux does not. We strictly define the *Total Swipe Flux* $\Phi_{swipe}(t)$ as the

magnitude of the outflow current:

$$\begin{aligned}\Phi_{swipe}(t) &\triangleq -J(t, b^*(t)) \\ &= \frac{1}{2}\sigma^2(b^*(t))^2 \frac{\partial f}{\partial z}(t, b^*(t)).\end{aligned} \tag{14}$$

This flux $\Phi_{swipe}(t)$ is the coupling variable that closes the feedback loop. In the platform's optimization problem (defined next), this flux represents the *supply of ad inventory*. The conflict between maximizing this inventory supply (Φ) and maintaining user retention (τ) constitutes the central tension of our model.

Remark 1 (Mathematical Well-posedness). The inclusion of source terms in Fokker-Planck equations is well-grounded. Zhu (2022) establishes the existence of weak solutions for kinetic FPEs with source terms, showing that the solution norm depends on the source norm. In the context of MFG, Claisse, Ren, and Tan (2023) derive similar equations where state-dependent birth/death rates lead to solutions that are finite measures rather than probability measures—a property that perfectly matches our non-conservative user population dynamics.

Remark 2 (Numerical Implementation). While the Dirac delta $\delta(z - z_p)$ is theoretically convenient, for numerical implementation (Section 4), we approximate it using a smooth Gaussian kernel to prevent Gibbs phenomena and ensure spectral stability.

Remark 3 (Distinction from Closed Queueing Networks). It is worth noting that the "Traffic Recycling" topology bears a structural resemblance to Closed Queueing Networks (CQN) or Gordon-Newell networks (Gordon & Newell, 1967). However, our framework differs fundamentally in its *control objective*. In classical CQN, the service mechanism is typically passive, and the objective is often to manage congestion or maximize throughput given fixed service requirements. In our Short-Video Economy, the "service time" (dwell time) is *strategically endogenous*. The platform (server) actively intervenes via $a(t, z)$ to manipulate the rate of recycling. The innovation lies not just in the closed loop itself, but in the *Stackelberg gaming* of this loop: the platform optimizes the *velocity* of circulation against the user's physiological patience, creating the unique "Algorithmic Liquidity Trap" phenomenon that standard queueing models do not capture.

3.5 The Leader's Problem: Mean-Field Control with Dual Monetization

In the standard Mean-Field Game literature, the principal often maximizes a functional of the population state alone (e.g., total density). However, as identified in the audit of short-video economics, a model focusing solely on total attention time ($\int z f$) mimics a subscription-based model (like Netflix) but fails to capture the ad-driven revenue model of platforms like TikTok or Reels. In these ecosystems, revenue is generated not just by "gazing" (retention) but significantly by "swiping" (impression generation).

To resolve the "Stalemate of Perfection" paradox—where optimal retention implies zero ad inventory—we reconstruct the platform's objective and control mechanisms. The

platform acts as a Stackelberg leader, optimizing a vector of controls to balance retention intensity against the need for traffic circulation (swiping).

3.5.1 Reformulation: Decoupling Scale and Structure

To explicitly separate the economic effects of market growth (Scale) from user engagement quality (Structure), we perform a change of variables on the state space. We decompose the unnormalized population density $f(t,z) \in \mathcal{M}^+(\Omega)$ into two orthogonal components: the total population mass $N(t) \in \mathbb{R}_+$ and the probability density function $m(t,z) \in \mathcal{P}(\Omega)$.

The relationship is defined as:

$$\begin{aligned} f(t,z) &= N(t) \cdot m(t,z), \\ \text{where } N(t) &= \int_\Omega f(t,z)\,\mathrm{d}z, \\ \text{and } \int_\Omega m(t,z)\,\mathrm{d}z &= 1. \end{aligned} \tag{15}$$

Consequently, the platform's value functional is redefined as $\mathcal{W}(f) \equiv V(N,m)$. This decomposition allows us to distinguish the platform's two strategic imperatives: managing the *market size* (via acquisition A and churn ϵ) and optimizing the *attention quality* (via retention a and matching u).

3.5.2 The Reconstructed Objective Functional and Constraints

The platform seeks to maximize the total discounted net profit over an infinite horizon. The objective functional $J(a,u,A)$, expressed in terms of the decomposed variables, combines retention value, impression revenue, and operational costs:

$$\begin{aligned} \max_{a,u,A} J = \int_0^\infty e^{-rt} \Bigg[& \underbrace{N(t) \int_{b^*(t)}^{z_{\max}} \left(\lambda_1 z - \frac{\xi_1}{2} a^2 \right) m \,\mathrm{d}z}_{\text{Net Retention Value}} \\ & + \underbrace{N(t)\lambda_2 \phi_{swipe}(m) - \frac{\xi_2}{2} u(t)^2}_{\text{Net Impression Value}} \\ & - \underbrace{\Psi(A(t))}_{\text{Acq. Cost}} \Bigg] \mathrm{d}t \end{aligned} \tag{16}$$

Economic Interpretation of the Objective: Equation (16) encapsulates the fundamental tension of the attention economy:

- **The Editor's Incentive ($\lambda_1 z$):** Rewards prolonged engagement depth (Duration).
- **The Dealer's Incentive ($\lambda_2 \phi_{swipe}$):** Rewards the velocity of card dealing (Flux).

The conflict arises because ϕ_{swipe} is physically generated by terminating the viewing process. If λ_2 (Ad Revenue) dominates, the algorithm is mathematically incentivized to ac-

celerate the "churn" from individual videos to maximize aggregate inventory, potentially creating the "Algorithmic Liquidity Trap."
Subject to the following constraints:

1. **Scale Dynamics (ODE):** The evolution of the total population $N(t)$ is governed by the balance between acquisition and net churn:

$$\frac{\mathrm{d}N}{\mathrm{d}t} = A(t) - \epsilon N(t)\phi_{swipe}(m). \tag{17}$$

2. **Structural Dynamics (PDE):** The probability density $m(t, z)$ evolves according to the Kolmogorov equation with dilution and recycling terms, defined on $z \in (b^*(t), z_{\max}]$:

$$\begin{aligned}\frac{\partial m}{\partial t} = \mathcal{L}^*[a]m \\ + \frac{A(t)}{N(t)}(\delta_{z_p} - m) \\ + (1-\epsilon)\phi_{swipe}(m)\big[g(z; u(t)) - m\big].\end{aligned} \tag{18}$$

3. **Flux Consistency:** The per-capita swipe flux is determined by the density gradient at the absorbing boundary:

$$\phi_{swipe}(m) \triangleq \frac{1}{2}\sigma^2(b^*(t))^2\frac{\partial m}{\partial z}(t, b^*(t)). \tag{19}$$

4. **User Incentive Compatibility:** The stopping boundary $b^*(t)$ arises endogenously from the users' optimal stopping problem (smooth-pasting condition):

$$b^*(t) = \inf\left\{z \in \Omega \mid v(t, z) > K(t, u(t)) - C\right\}. \tag{20}$$

3.5.3 The Master Equation (Explicit Form)

The platform's value function $V(N, m)$ satisfies the infinite-dimensional Hamilton-Jacobi-Bellman equation. By solving the maximization problem using the structure-scale decomposition (see Appendix C), we derive the explicit equation:

$$
\begin{aligned}
rV(N,m) = \sup_{a,u,A} \Bigg\{ & \underbrace{\pi(N,m,a,u,A)}_{\text{Flow Payoff}} \\
& + \frac{\partial V}{\partial N}\Big(A - \epsilon N \phi_{swipe}(m)\Big) \\
& + \int_{\Omega} \frac{\delta V}{\delta m}(z) \Bigg[\\
& \quad \mathcal{L}^*[a]m \\
& \quad + \frac{A}{N}(\delta_{z_p} - m) \\
& \quad + (1-\epsilon)\phi_{swipe}(m) \\
& \quad \cdot \big(g(z;u) - m\big) \Bigg] \mathrm{d}z \Bigg\}.
\end{aligned}
\tag{21}
$$

Expanding the optimal controls yields the non-linear form. Note that the boundary term arising from $\mathcal{L}^*$ is implicitly contained within the Hamiltonian structure:

$$
\begin{aligned}
rV(N,m) = N \int & \Big[\lambda_1 z + \mu z \partial_z \lambda_m \\
& \qquad + \frac{1}{2}\sigma^2 z^2 \partial_{zz}\lambda_m \Big] m \,\mathrm{d}z \\
& + \underbrace{N \int \frac{z^2}{2\xi_1} \big(\partial_z \lambda_m\big)^2 m \,\mathrm{d}z}_{\text{Quadratic Retention Gain}} \\
& + H_{acq}\Big(\frac{\partial V}{\partial N} + \frac{1}{N}\big(\lambda_m(z_p) - \bar{\lambda}_m\big) \Big) \\
& + \mathcal{H}_{match}\big(N\phi_{swipe}(m), \lambda_m\big),
\end{aligned}
\tag{22}
$$

where $\lambda_m(z) \equiv \frac{1}{N}\frac{\delta V}{\delta m}(z)$ represents the normalized structural shadow price, and $\bar{\lambda}_m = \int \lambda_m m \,\mathrm{d}z$. The functions H_{acq} and $\mathcal{H}_{match}$ denote the Legendre-Fenchel transforms (convex conjugates) of the cost functions associated with acquisition A and matching control u, respectively. Specifically, $H_{acq}(Y) = \sup_{A \geq 0} \{Y \cdot A - C_{acq}(A)\}$, where the argument Y represents the net shadow value of a new user.

3.5.4 Optimal Strategic Responses

The First-Order Conditions (FOCs) describe the optimal strategy:

- **Optimal Retention Intensity:**

$$
a^*(t,z) = \frac{z}{\xi_1}\frac{\partial}{\partial z}\left(\frac{1}{N}\frac{\delta V}{\delta m}(z)\right). \tag{23}
$$

The platform applies friction proportional to the gradient of the structural shadow price, fighting boredom where the marginal gain of engagement quality is highest.

- **Optimal Matching Investment:**

$$\xi_2 u^*(t) = (1-\epsilon)N(t)\phi_{swipe}(m) \cdot \int \frac{1}{N}\frac{\delta V}{\delta m}(z)\frac{\partial g}{\partial u}(z;u^*)\,\mathrm{d}z. \tag{24}$$

  Investment is proportional to traffic flux $N\phi_{swipe}$. This creates a *positive feedback loop*: higher flux justifies better algorithms, and better algorithms (via the Dopamine Treadmill) further accelerate flux. This self-reinforcing mechanism explains the system's tendency to overheat into the "Phantom Liquidity" regime.

- **Optimal Acquisition:**

$$\Psi'(A^*) = \frac{\partial V}{\partial N} + \frac{1}{N}\left(\frac{\delta V}{\delta m}(z_p) - \int \frac{\delta V}{\delta m} m\,\mathrm{d}z\right). \tag{25}$$

  Marketing spend balances the value of adding a user ($\partial_N V$) against the dilution effect of adding a user at the specific entry state z_p.

3.5.5 Mean Field Equilibrium

Substituting the optimal controls $a^*(t,z)$, $u^*(t)$, and $A^*(t)$ derived in Equations (23)–(25) back into the structural dynamics, we obtain the fully coupled system governing the platform ecosystem.

Definition 1 (Mean-Field Stackelberg Equilibrium)**.** A Mean-Field Equilibrium is a tuple of time-dependent functions $(V, m, \Phi_{swipe}, b^*)$ such that:

1. **Optimality:** Given the population distribution m, the value function V satisfies the Master HJB Equation (21), and the controls $\{a^*, u^*, A^*\}$ maximize the Hamiltonian.

2. **Consistency:** Given the optimal controls, the population density m evolves according to the controlled Non-Local Fokker-Planck Equation:

$$\begin{cases} \dfrac{\partial m}{\partial t} = -\dfrac{\partial}{\partial z}\Big((\mu + a^*)zm\Big) \\ \qquad\qquad + \dfrac{1}{2}\sigma^2\dfrac{\partial^2}{\partial z^2}(z^2 m) + \mathcal{S}(m,V), \\ \\ rV = \sup\limits_{a,u,A} \mathcal{H}\left(N, m, V, \dfrac{\delta V}{\delta m}, a, u, A\right), \end{cases} \tag{26}$$

   where $\mathcal{S}(m,V)$ represents the source terms driven by acquisition $A^*[V]$ and recycled traffic flux $\Phi_{swipe}(m)$; $\mathcal{H}$ denotes the *pre-optimized* Hamiltonian functional defined in Eq. (21).

3. **Boundary Consistency:** The swipe flux $\Phi_{swipe}(t)$ and the free boundary $b^*(t)$ satisfy

the smooth-pasting condition generated by the users' optimal stopping problem:

$$b^*(t) = \inf\{z \in \Omega \mid v(t,z) > K(t,u^*(t)) - C\}. \tag{27}$$

Due to the forward-backward structure (forward in time for m, backward for V) and the non-local coupling via Φ_{swipe}, this system does not admit a closed-form analytical solution. In the following section, we propose a computational framework for solving the stationary equilibrium.

4 Structure Calibration and Solution Algorithm

To bridge the gap between our theoretical Mean-Field Stackelberg framework and the empirical reality of short-video platforms, we adopt a two-stage methodological approach. First, we calibrate the micro-foundation of the model—specifically the user's attention dynamics and utility parameters—using the *KuaiRand* dataset (Gao et al., 2022), a large-scale unbiased dataset collected from Kuaishou. Second, to tackle the computational intractability of the high-dimensional coupled Master Equation (Eq. 21), we propose a novel *Spectral Soft Actor-Critic (Spectral-SAC)* algorithm. This physics-informed Deep Reinforcement Learning (DRL) solver enables us to capture the complex non-local feedback loops of the traffic recycling mechanism efficiently.

4.1 Data and Structural Estimation

4.1.1 Data Source and Identification Strategy

To structurally identify the micro-foundation parameters of our model—specifically the natural attention decay rate μ and the intrinsic volatility σ—we utilize the *KuaiRand Dataset* (Gao et al., 2022). This is a large-scale sequential recommendation dataset collected from the video-sharing mobile App, Kuaishou.

The Identification Challenge: Endogeneity. A critical challenge in calibrating user attention dynamics from standard production logs is the presence of *endogeneity bias* (or confounding bias) (J. Chen et al., 2023). In a standard environment, the recommendation algorithm ($a(t,z)$) actively intervenes to match content with users. Consequently, the observed attention drift is a composite of intrinsic user patience and algorithmic matching quality:

$$\underbrace{\text{Observed Drift}}_{\text{Data}} = \underbrace{\mu}_{\substack{\text{Unknown} \\ \text{User Param}}} + \underbrace{a(t,z)}_{\substack{\text{Unobservable} \\ \text{Algo Control}}} \tag{28}$$

Without isolating the algorithmic intervention, the structural parameter μ remains unidentified, as a long dwell time could result from either high user patience or an exceptionally good recommendation.

Identification Strategy: The Randomized Experiment. To disentangle these effects, we exploit the unique *KuaiRand-Pure* subset (Gao et al., 2022). In this regime, videos were exposed to users via a purely random policy, effectively setting the algorithmic intervention to zero ($a(t, z) \approx 0$). This allows us to treat the observed viewing behaviors as a *Randomized Controlled Experiment*, revealing the users' baseline physiological and psychological reaction functions (the "Organic Baseline") without the confounding influence of the recommender policy (Saito, Yaginuma, Nishino, Sakata, & Nakata, 2020). This identification strategy provides the ground truth for the "Healthy" parameters used in Section 3.3.

4.1.2 Variable Construction and Processing

We focus on the granular interaction logs where the user's viewing duration is recorded in milliseconds (Gao et al., 2022). We define the mapping between the theoretical stopping time τ (from our optimal stopping model) and the observed data. Let τ_{obs} denote the observed viewing time and D denote the finite duration of the video. The relationship is given by:

$$\tau_{obs} = \min(\tau, D). \tag{29}$$

Handling Right-Censoring. This structure implies that the data is *Right-Censored*. An observation where a user watches until the end ($\tau_{obs} = D$) provides only partial information: we know the user's latent patience τ satisfies $\tau \geq D$, but the exact stopping threshold was not triggered. Ignoring this censoring mechanism—i.e., treating completion as an exact measure of interest—would lead to a systematic **underestimation** of the natural decay parameter μ (overestimating drift).

Data Filtering and Sample Selection. Data processing was performed on the *KuaiRand-1K* subset (Gao et al., 2022). We applied rigorous filters to ensure the structural validity of the calibration:

1. **Mechanical Noise Removal:** We excluded interactions shorter than 0.5s. This threshold corresponds to the physiological limit of the human motor system reaction time, filtering out accidental touches devoid of cognitive processing.

2. **Content Validity:** We removed videos shorter than 3 seconds[2], ensuring a meaningful consumption horizon.

3. **User Activity:** We excluded inactive users with fewer than 5 interactions to ensure sufficient data points for stable parameter identification.

The final sample used for structural calibration consists of 40,922 interaction events under the random exposure policy. Table 1 summarizes the descriptive statistics.

[2]The 3-second threshold aligns with the platform's definition of "short_time_play" and ensures the content has sufficient duration to convey information.

Table 1: Descriptive Statistics of the Processed Sample (KuaiRand-Pure)

Variable	**N**	**Mean**	**Std. Dev.**	**Median**	**Max**
Viewing Time (τ_{obs}, s)	40,922	6.97	18.14	2.15	508.1
Video Duration (D, s)	40,922	108.46	104.81	80.78	1177.7
Structure Metric			**Value**		
Censoring Rate (Completes)			3.33%		

Note: The sample excludes mechanical noise ($< 0.5s$) and inactive users. Viewing time is treated as right-censored if $\tau_{obs} = D$. The low censoring rate (3.33%) confirms that "swiping" is the dominant mode of termination.

4.1.3 Structural Estimation via Maximum Likelihood

Under the Geometric Brownian Motion assumption for the attention state Z_t, the logarithm of attention, $X_t = \ln(Z_t)$, follows an Arithmetic Brownian Motion. Consequently, the First Passage Time (FPT) to the swiping threshold b^* follows an *Inverse Gaussian (IG)* distribution (also known as the Wald distribution) (Folks & Chhikara, 1978; Peskir & Shiryaev, 2006).

The probability density function (PDF) for the random stopping time τ is given by:

$$f(t;\eta,\lambda) = \sqrt{\frac{\lambda}{2\pi t^3}} \exp\left(-\frac{\lambda(t-\eta)^2}{2\eta^2 t}\right), \tag{30}$$

where $\eta > 0$ represents the mean time to swiping (mean FPT) and $\lambda > 0$ is the shape parameter characterizing the variance.

Handling Censored Data. To account for the right-censoring structure identified in the previous step, we construct the log-likelihood function by combining the probability density for observed switches (uncensored, $\delta_i = 1$) and the survival function $S(t) = 1 - \int_0^t f(s)\mathrm{d}s$ for videos that ended before a swipe occurred (censored, $\delta_i = 0$):

$$\begin{aligned}\mathcal{L}(\eta,\lambda) = \sum_{i=1}^{N} \Big[& \delta_i \ln f(t_i;\eta,\lambda) \\ & + (1-\delta_i)\ln S(D_i;\eta,\lambda)\Big].\end{aligned} \tag{31}$$

We maximize $\mathcal{L}(\eta,\lambda)$ using the L-BFGS-B optimization algorithm to ensure positivity constraints on parameters (Byrd, Lu, Nocedal, & Zhu, 1995).

Recovering Micro-Parameters. The estimated distributional parameters $(\hat{\eta},\hat{\lambda})$ are not our final goal; we use them to recover the *structural parameters* (μ,σ) of the user's attention dynamics. Let $x_0 = \ln(z_0/b^*)$ denote the normalized initial distance to the threshold.

Based on the properties of the FPT for Brownian motion, the mapping is derived as:

$$\begin{aligned} \hat{\sigma} &= \frac{x_0}{\sqrt{\hat{\lambda}}}, \\ \hat{\mu} &= \frac{1}{2}\hat{\sigma}^2 - \frac{x_0}{\hat{\eta}}. \end{aligned} \tag{32}$$

This procedure allows us to identify the natural decay rate $\hat{\mu}$ and volatility $\hat{\sigma}$ that govern the *Organic Baseline* of user behavior, stripped of algorithmic manipulation.

4.1.4 Calibration Results and Interpretation

The structural estimation results are presented in Table 2. The calibrated parameters establish the *"Physiological Baseline"* of the user, providing the empirical foundation for the Organic Equilibrium discussed in Section 3.3.

We highlight three key empirical findings regarding the human attention mechanism:

1. **The Inevitability of Boredom ($\mu < 0$):** We identify a significant negative drift $\mu = -0.0235$. This empirically validates the model's core assumption: without high-intensity stimuli, human attention naturally decays. This acts as the physiological gravity that algorithms must fight against.

2. **The Gambling Structure (σ):** The estimated volatility $\sigma = 0.3000$ is substantial relative to the drift. This reveals that the short-video ecosystem functions psychologically like a *Slot Machine* (Variable Ratio Schedule). The high variance implies that while the average video may be boring, the "fat tail" of occasional high-reward content keeps users engaged via the uncertainty of the next swipe.

3. **The Cost of Swiping (ϵ):** The macroscopic churn rate is identified as $\epsilon = 1.25\%$. This implies that every swipe carries a non-trivial risk of user exit. In our Stackelberg game, this ϵ represents the *structural constraint* that prevents the platform from infinite acceleration—if the algorithm pushes the swipe velocity too high, the cumulative churn will drain the user pool (the "Recycling Stagnation" effect).

These calibrated values provide the "physics engine" for our subsequent numerical simulations, ensuring that our policy analysis is grounded in the empirical reality of the KuaiRand dataset.

Figure 3 visualizes the goodness-of-fit of our structural estimation. As demonstrated in the figure, our theoretical model successfully captures the two defining morphological features of short-video consumption:

1. **Rapid Physiological Filtering (Left Tail):** The model accurately reproduces the high probability mass concentrated in the initial few seconds. This reflects the "short attention span" characteristic, where users rapidly filter out content that fails to meet the minimum stimulation threshold within the first cognitive cycle ($< 2s$).

Table 2: Structural Parameters Calibrated from KuaiRand-Pure

Parameter	Symbol	Value	Economic Interpretation
Natural Drift	μ	-0.0235	Users exhibit a systematic decay in attention (boredom) of approx. 2.35% per second.
Volatility	σ	0.3000	High content heterogeneity creates significant uncertainty ("gambling" effect).
Churn Rate	ϵ	0.0125	The probability of exiting the App after a single swipe is 1.25%.
Init. Distance	x_0	0.50	Calibrated relative distance between initial expectation and the threshold.

Note: Parameters are estimated using Maximum Likelihood Estimation (MLE) on the Inverse Gaussian distribution. x_0 is set to 0.5 to satisfy the natural decay constraint ($\mu < 0$).

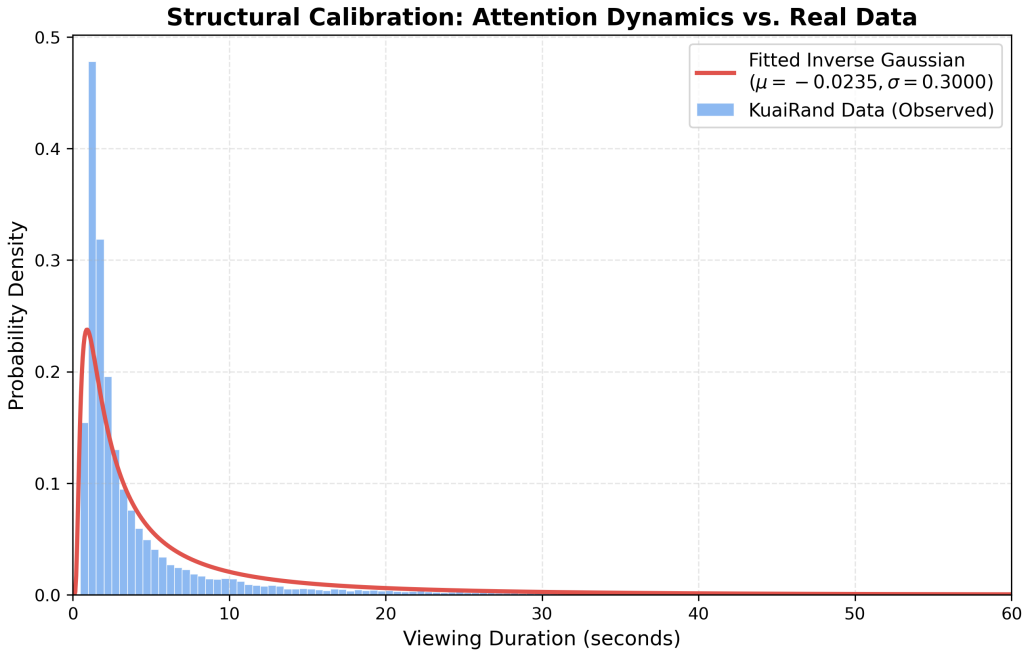

Figure 3: Structural Calibration Model Fit. The blue histogram represents the empirical probability density of user viewing durations observed in the KuaiRand-Pure dataset. The red solid line depicts the theoretical Inverse Gaussian (IG) distribution implied by our optimal stopping model, parametrized by the calibrated values $\hat{\mu} = -0.0235$ and $\hat{\sigma} = 0.3000$. The close alignment between the theoretical curve and empirical data validates the Geometric Brownian Motion assumption for attention dynamics.

Note: The comparison shows a strong fit in both the rapid dropout phase (left tail) and the heavy-tailed retention phase (right tail), confirming the model's ability to capture the dual nature of impatience and engagement.

2. **Heavy-Tailed Engagement (Right Tail):** Crucially, the fitted curve aligns closely with the long right tail of the empirical distribution. This confirms that the calibrated volatility parameter ($\sigma = 0.3000$) effectively captures the "gambling-like" uncertainty mechanism: a small fraction of high-match videos can sustain user attention for extended periods despite the natural decay drift ($\mu = -0.0235$).

The strong concordance between the theoretical prediction and the natural experiment data provides robust empirical support for modeling user attention as a stochastic process with an absorbing boundary.

4.2 Solution Methodology: Spectral Deep Reinforcement Learning

Solving the Mean-Field Stackelberg Game described in Section 3 poses significant computational challenges. The system is characterized by a high-dimensional state space (the discretized density $m(t, \cdot)$) and, crucially, a *non-local coupling term* (traffic recycling) that introduces strong feedback loops. Standard numerical methods, such as finite difference schemes with adjoint-based optimization, suffer from the *curse of dimensionality* and often fail to converge due to the instability caused by the free boundary (Han, Jentzen, & E, 2018; Peskir & Shiryaev, 2006).

To overcome these hurdles, we develop a novel hybrid solver: *Spectral Soft Actor-Critic (Spectral-SAC)*. This approach integrates the structural regularization of spectral methods with the exploration capabilities of maximum entropy reinforcement learning.

Dimensionality Reduction via Spectral Embedding. A key innovation in our approach is the parameterization of the platform's infinite-dimensional control policy $a(t, \cdot) \in L^2(\Omega)$. Direct output for every spatial grid point (action dim $N_z \approx 400$) would lead to inefficient learning and jagged, discontinuous policies. Instead, we project the retention policy onto the subspace spanned by the first K Chebyshev polynomials $T_k(\cdot)$:

$$a(t, z) \approx \sum_{k=0}^{K-1} w_k(t) T_k(\tilde{z}). \tag{33}$$

This spectral technique (detailed in Online Appendix D) reduces the effective action dimension to $K + 2 \approx 12$. Crucially, it acts as a *low-pass filter*, enforcing spatial smoothness on the algorithmic intervention, which aligns with the continuity requirements of practical recommendation systems.

Physics-Informed "Solver-in-the-Loop". Unlike black-box RL approaches that rely on purely data-driven transitions, our environment is *Physics-Informed*. At each interaction step, the reward calculation involves explicitly solving:

1. The user's Optimal Stopping problem (HJB QVI) to determine the swiping boundary $b^*(t)$.

2. The population's mass conservation (Non-Local Fokker-Planck equation) to compute the exact next state density.

This "Solver-in-the-Loop" architecture ensures that the learned policy strictly satisfies the user's incentive compatibility and the system's flux consistency constraints at all times, minimizing the gap between the RL approximation and the true MFG solution.

Theoretical Connection to the Master Equation. The Critic network $Q_\theta(s, a)$ in our algorithm serves as a function approximator for the value function of the Master Equation defined on the Wasserstein space of probability measures (Carmona, Delarue, et al., 2018). By minimizing the Bellman error using the Soft Actor-Critic (SAC) objective, the algorithm implicitly recovers the solution to the Mean-Field Game. The entropy regularization in SAC is particularly valuable here, promoting adequate exploration of the high-dimensional state space and improving the robustness of the solution against numerical instability.

For the detailed MDP formulation, network architecture, hyperparameter settings, and the complete pseudo-code (Algorithm 2), we refer readers to Online Appendix D.

5 Main Results: The Algorithmic Liquidity Trap

In this section, we present the numerical equilibrium results derived from the calibrated model. We contrast the equilibrium outcomes under two distinct monetization regimes: a *Retention-Oriented* strategy (maximizing time spent) and an *Ad-Centric* strategy (maximizing swipe flux).

Our analysis reveals a critical phase transition: when the incentive for ad impressions (λ_2) surpasses a critical threshold relative to retention (λ_1), the ecosystem collapses into a pathological state we term the *"Algorithmic Liquidity Trap."* In this regime, the platform creates an illusion of high engagement (extreme swipe velocity) that is, in reality, devoid of meaningful attention (Phantom Liquidity).

Benchmarking the Solver. To validate the capability of our Spectral-SAC solver in handling the *complex feedback dynamics* of the ecosystem, we benchmarked it against a standard gradient-based optimization method (SLSQP). The comparison, detailed in Online Appendix E, confirms that traditional solvers are often limited to inferior linear policies. In contrast, the *exploratory nature* of our physics-informed DRL agent enables it to discover the globally superior non-linear control structures required to manipulate the equilibrium.

5.1 Experimental Setup and Parameter Specification

Physiological Baseline (Fixed). To ensure empirical validity, we fix the user behavioral parameters to the values structurally estimated from the KuaiRand dataset (Table 2). These represent the immutable *"Physiological Constraints"* of the user population:

- Natural Decay: $\mu = -0.0235$ (The force of boredom).

- Volatility: $\sigma = 0.3000$ (The gambling uncertainty).
- Churn Rate: $\epsilon = 0.0125$ (The cost of recycling).

Platform Incentives (Variable). The core of our experiment lies in varying the platform's objective function (Eq. 16). We normalize the user's cognitive switching cost to $C = 1$. We define the *Monetization Ratio* $\mathcal{R}$ as the relative weight of impression revenue to retention value:

$$\mathcal{R} \equiv \frac{\lambda_2}{\lambda_1}. \tag{34}$$

Note on Scaling: Since the optimal control policy depends on the relative weights in the objective functional, our results are robust to the absolute scaling of λ_1 and λ_2. The phenomenon of the Liquidity Trap is driven by the high Monetization Ratio $\mathcal{R}$, independent of the total magnitude of the calculated profit.

We simulate the equilibrium path as $\mathcal{R}$ increases from 0 (Pure Retention Model) to high values (Ad-Driven Model), observing how the optimal algorithmic policy $a^*(t, z)$ adapts to extract maximum value from the user's physiological constraints.

Detailed hardware specifications, hyperparameter settings for the SAC agent (e.g., learning rates, replay buffer size), and the spectral regularization parameter ($c_{\text{smooth}} = 20.0$) are provided in Online Appendix D.

5.2 Baseline Equilibrium Analysis: The Healthy Retention Regime

We first examine the baseline scenario where the platform prioritizes user retention (Long-Term Value) over immediate traffic flux. The objective weights are configured as $\lambda_1 = 20.0$ (High Retention Priority) and $\lambda_2 = 0.01$ (Low Flux Priority). The Mean-Field Stackelberg agent was trained using the Spectral SAC algorithm for 500,000 steps. The simulation results, visualized in Figure 4, demonstrate that under this "User-Centric" policy, the ecosystem successfully converges to a stable and sustainable Mean-Field Equilibrium (MFE).

Sustainable Carrying Capacity (N^*). As shown in Figure 4 (Top), the system dynamics exhibit a robust convergence toward stationarity. Following a brief transient phase ($t < 10s$) where the agent overcomes the system's initial inertia, the total active population $N(t)$ follows a logistic growth trajectory, stabilizing at an equilibrium level of $N^* \approx 37.5$. *Economic Interpretation:* This equilibrium represents the ecosystem's *Sustainable Carrying Capacity*, where user acquisition is organically balanced by endogenous churn. Simultaneously, the traffic flux converges to $\Phi^* \approx 23.6$ swipes per second. This stable flux serves as the "healthy heartbeat" of the ecosystem, driving traffic recycling without overheating the user base.

The Dopamine Treadmill and Rational Expectations. A critical insight, depicted in Figure 4 (Middle Right), is the evolution of the optimal stopping threshold, which stabilized at $b^* \approx 2.14$. This validates the Dopamine Treadmill hypothesis: as the platform improves

content quality (dashed orange line), users rationally internalize this improvement by raising their reservation utility (b^*). *Policy Implication:* In this baseline regime, the treadmill is stable. The platform meets the rising expectations through quality matching rather than manipulative acceleration, maintaining a high threshold that filters for quality over quantity.

Resolution of the "Stalemate": Organic Monetization. Figure 4 (Bottom Right) offers a definitive resolution to the "Stalemate of Perfection" paradox. Even with a negligible weight on flux ($\lambda_2 = 0.01$), the ad revenue contributes approximately **13.6%** to the total value. *Mechanism:* This implies that a "Retention-First" strategy implicitly sustains a valuable ad inventory. By maintaining a large, engaged user population (N^*), the platform generates sufficient "Organic Swipes" to monetize effectively without needing to artificially shorten the user's attention span. This establishes the *Healthy Baseline* against which we will measure the pathological effects of algorithmic greed.

5.3 Comparative Statics: The Algorithmic Liquidity Trap

To investigate the dynamics of an Ad-Centric monetization strategy, we conducted *Experiment A*. In this scenario, we pivot the platform's objective from the "Editor" logic to the "Dealer" logic: we increase the weight on traffic flux to $\lambda_2 = 5.0$ (up from 0.01) while reducing retention weight to $\lambda_1 = 5.0$.

The results, visualized in Figure 5, reveal the emergence of a counter-intuitive and pathological regime we term the *"Algorithmic Liquidity Trap."*

Macroscopic Illusion: Phantom Liquidity. Contrary to the "Leaky Bucket" intuition—which suggests that encouraging churn would deplete the user base—the system converged to a significantly *larger* steady-state population of $N^* \approx 53.7$ (vs. 37.5 in baseline).

The Mechanism of Illusion: This growth is not driven by satisfaction, but by *Velocity*. The traffic flux exploded to $\Phi^* \approx 139.9$ swipes/sec (a $6\times$ increase). The platform has effectively created a "Centrifuge": by accelerating the *Traffic Recycling Loop*, it keeps users trapped in a state of high-frequency circulation. We define this state as *Phantom Liquidity*: the ecosystem is awash with traffic volume, but the actual attention depth per interaction has collapsed.

This hyper-velocity drives the average dwell time down to a physiological floor of $\tau \approx 0.38s$. This value, falling below our data cleaning threshold for cognitive processing (0.5s), suggests the system has degenerated into a state of reflex-driven interaction rather than conscious consumption. The platform has effectively created a "Centrifuge": by accelerating the *Traffic Recycling Loop*, it keeps users trapped in a state of high-frequency circulation. We define this state as *Phantom Liquidity*: a condition of high system throughput but collapsed value.[3]

[3]It is important to clarify that while a 0.38s exposure is optically non-zero (potentially allowing for subliminal icon recognition), it constitutes *"Economic Zero"* under current industry standards. According to the Media Rating Council (MRC) guidelines for mobile video advertising, an impression is only classified as "Viewable" (and billable) if it plays for at least 2 continuous seconds (Media Rating Council (MRC) and Interactive Ad-

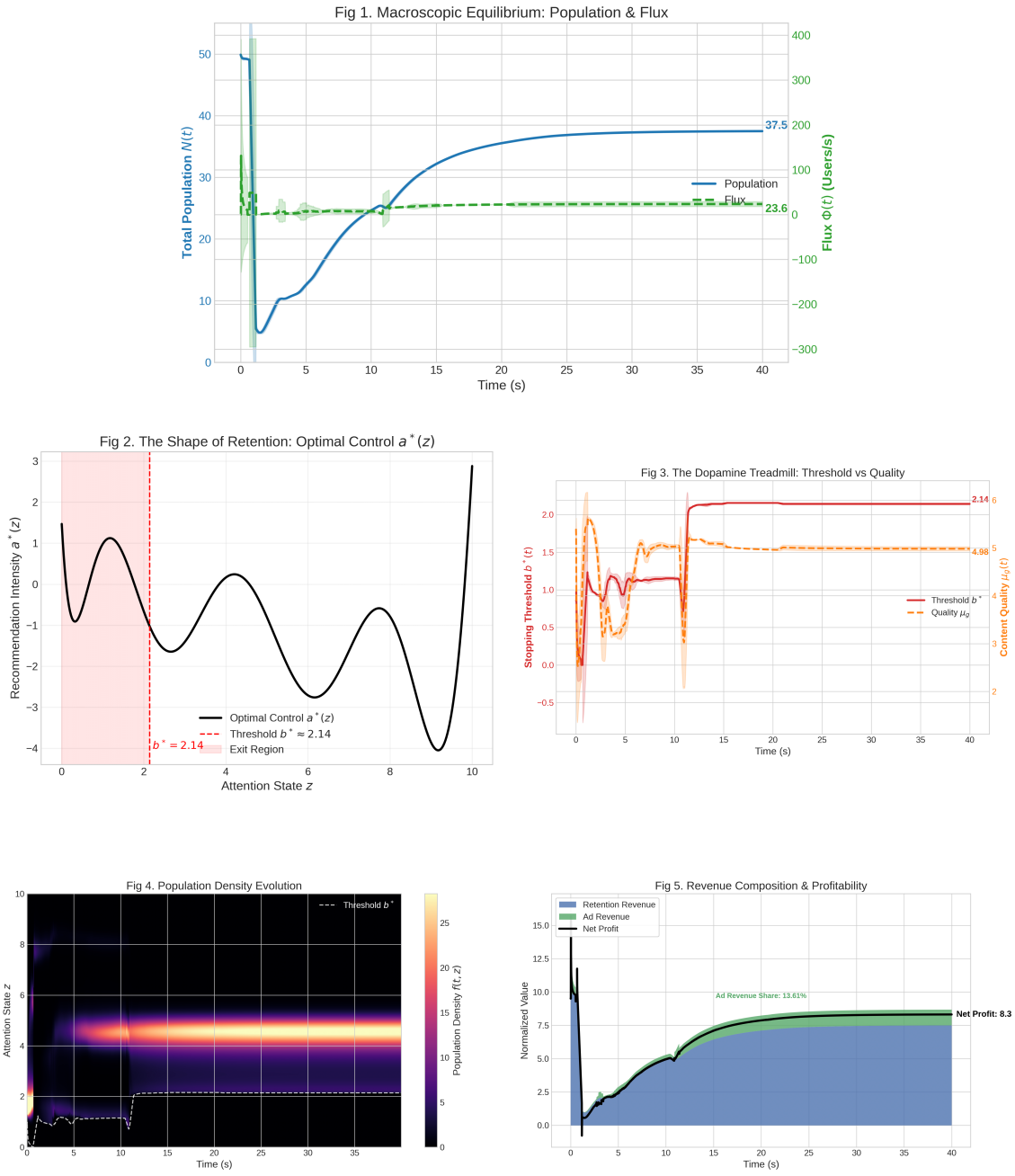

Figure 4: **The Healthy Baseline Equilibrium. (Top)** Macroscopic convergence to a sustainable carrying capacity ($N^* \approx 37.5$) and organic flux ($\Phi^* \approx 23.6$). **(Middle Left)** The learned optimal retention control $a^*(z)$, showing a smooth, non-linear intervention strategy enabled by spectral regularization. **(Middle Right)** The **Dopamine Treadmill**: The user's stopping threshold b^* (red) rises in tandem with content quality (orange), reflecting rational expectation adjustment. **(Bottom Right) Organic Monetization**: The breakdown of Total Value (black) shows that Ad Revenue (green area) naturally accounts for 13.61% of profits even under a Retention-First policy, resolving the paradox of perfection.

Micro-Strategy: The "Safety Net" Manipulation. The learned control policy $a^*(z)$ (Figure 5, Middle Left) reveals how the DRL agent manipulates user psychology. It discovers a non-linear strategy we term the *"Safety Net" Policy*:

- **The Barrier ($a^* > 3.5$):** The algorithm applies extreme retention force *only* immediately above the exit threshold b^*. This acts as a safety net to prevent the user from leaving the App entirely.
- **Planned Obsolescence ($a^* < 0$):** Crucially, once the user is "safe" (in higher attention states), the algorithm actively *negatively* impacts retention (or relaxes control). This allows natural boredom to push the user back towards the boundary rapidly, triggering another lucrative swipe event without risking total churn.

This strategy effectively converts "Time" into "Swipes" with surgical precision.

Economic Transformation: The Slot Machine Model. Figure 5 (Bottom Right) illustrates the complete mutation of the business model. Ad Revenue now constitutes **99.92%** of the total value. The platform has ceased to be a utility for content consumption and has evolved into a *"Slot Machine" mechanism*. It maximizes profit not by extending the "Gaze" (Baseline behavior) but by inducing rapid, repetitive cycles of "Anticipation and Result" (Swiping), exploiting the user's gambling-like volatility (σ) identified in Section 4.

Robustness Verification: Scale Invariance. A potential concern regarding Experiment A is whether the emergence of the "Liquidity Trap" is an artifact of the specific parameter scale ($\Sigma\lambda = 10$) rather than the monetization structure. To rule this out, we conducted a *Scale-Invariant Robustness Check* (Experiment C) where we doubled the incentive weights to $\lambda_1 = \lambda_2 = 10.0$. This normalizes the total weight budget to match the Baseline ($\Sigma\lambda = 20$) while maintaining the critical monetization ratio $\mathcal{R} = 1$.

The results, detailed in Online Appendix E.4, confirm the persistence of the trap. The system converged to a high-velocity equilibrium with a flux of $\Phi^* \approx 108.75$ swipes/sec, which is statistically consistent with the Ad-Centric regime ($\Phi^* \approx 118.75$) and fundamentally distinct from the Baseline ($\Phi^* \approx 23.6$). This empirically confirms that the pathological state is driven by the *relative structure* of incentives ($\mathcal{R}$), independent of the absolute magnitude of the profit function.

6 Conclusion and Implications

This paper bridges the gap between the micro-foundations of user attention and the macro-dynamics of platform economics. By integrating an optimal stopping framework into a Mean-Field Stackelberg Game, and calibrating it with the KuaiRand dataset, we provide a structural lens to understand the "Scroll or Stop" dilemma.

vertising Bureau (IAB), 2015). Furthermore, such rapid dismissal serves as a strong negative feedback signal (Active Rejection) in recommender systems, distinguishing it from merely "low-value" passive consumption.

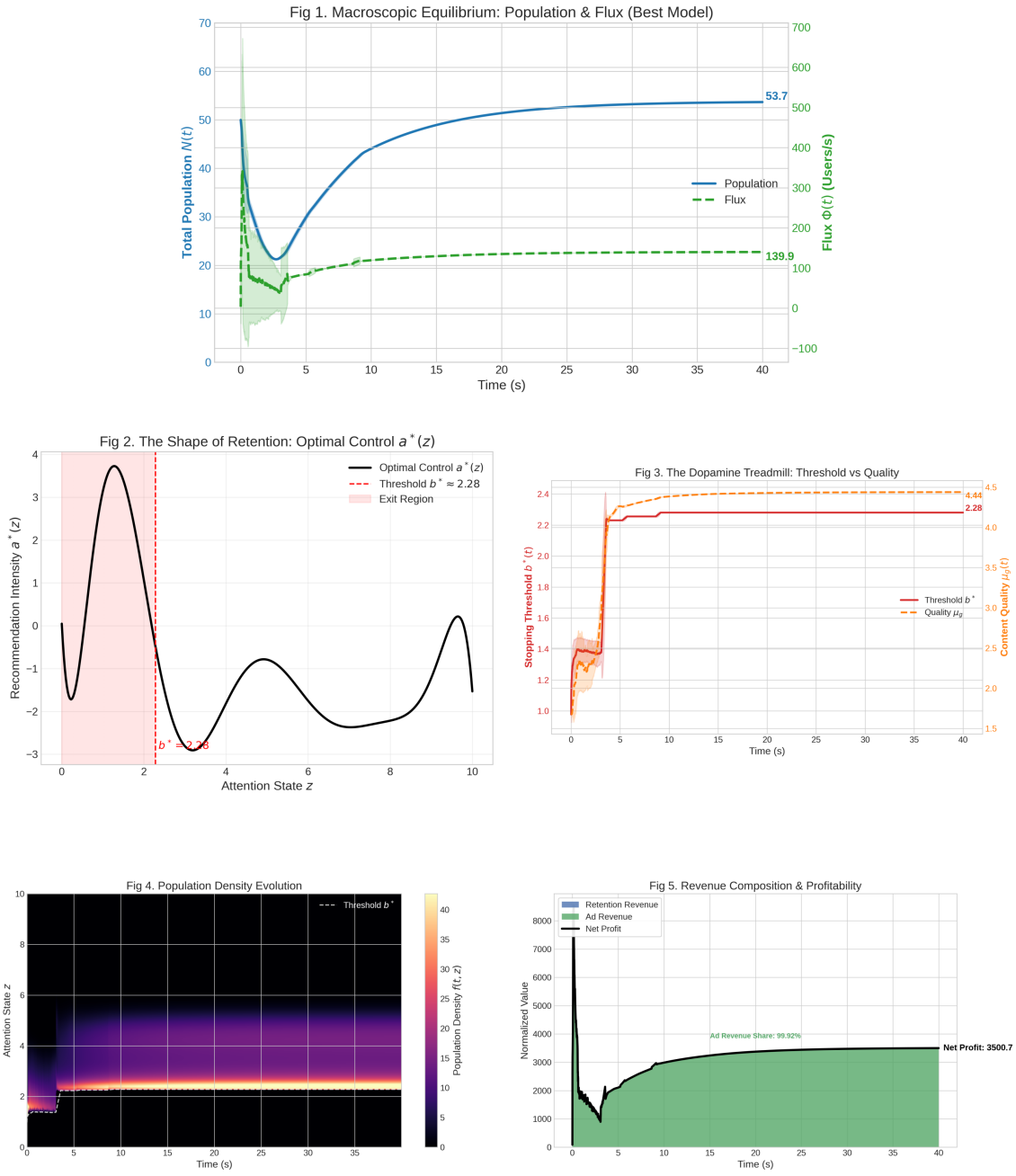

Figure 5: **The Algorithmic Liquidity Trap (Experiment A). (Top)** *Phantom Liquidity*: The ecosystem inflates to a larger population ($N^* \approx 53.7$) driven by massive flux ($\Phi^* \approx 139.9$), masking the collapse of attention depth. **(Middle Left)** The *"Safety Net" Policy*: The algorithm exerts intense control only at the exit boundary to prevent churn, while encouraging rapid decay elsewhere to accelerate swiping. **(Middle Right)** The threshold b^* rises to 2.28, reflecting the user's adaptation to a high-turnover environment. **(Bottom Right)** *The Slot Machine Economy*: The monetization structure shifts almost entirely to Ad Revenue (99.92%), confirming the transition to a high-velocity, low-attention equilibrium.

6.1 Theoretical Contribution: The Anatomy of the Trap

Our central theoretical contribution is the identification of the *"Algorithmic Liquidity Trap."* We demonstrate that standard algorithmic objectives, when biased towards impression generation (the "Dealer" logic), do not merely optimize traffic; they fundamentally restructure the user's attention span.

- **Phantom Liquidity:** We show that high traffic metrics (DAU, Flux) can coexist with, and indeed mask, a collapse in engagement depth. The platform's ability to accelerate the *Traffic Recycling Loop* creates an illusion of prosperity ($N^* \uparrow$) driven by hyperactive churn rather than genuine satisfaction.
- **The Safety Net Mechanism:** We reveal how sophisticated reinforcement learning agents weaponize human physiology. By applying the "Safety Net" policy—relaxing control to induce decay and catching users only at the exit boundary—algorithms effectively commoditize user boredom, converting physiological impatience into commercial velocity.

6.2 Managerial Implications

For platform managers, our findings sound a cautionary note regarding the *KPI Alignment Problem*.

1. **The Metric Fallacy:** Optimizing for short-term flux (swipes/sec) yields immediate ad inventory gains (as seen in Experiment A) but pushes the ecosystem towards a high-volatility, low-trust equilibrium. This "Slot Machine" economy is fragile; while our model assumes a fixed physiological baseline, in reality, prolonged exposure to the "Safety Net" strategy may permanently erode user patience ($\mu \downarrow$), leading to eventual ecosystem collapse.
2. **Quality vs. Velocity:** We show that a "Retention-First" strategy (Baseline) is not antithetical to monetization. It generates *Organic Monetization* (13.6% of value) by maintaining a sustainable carrying capacity. Managers must resist the temptation of "Algorithmic Overheating" to preserve the long-term health of the user pool.

6.3 Policy Implications

For regulators, this study suggests that the current scrutiny on "Addiction" (Total Time Spent) may be insufficient. The more insidious danger lies in *"Cognitive Acceleration."* Algorithms that maximize velocity do not necessarily keep users "hooked" for longer hours; instead, they condition users to process information at a frenetic pace, systematically stripping away the capacity for deep attention. Regulatory frameworks should consider monitoring *Attention Depth Metrics* (e.g., average dwell time per impression) alongside standard engagement figures to assess the cognitive health of digital platforms.

6.4 Limitations and Future Directions

While our framework offers significant insights, it abstracts away certain complexities of the real-world ecosystem.

1. **The Missing Supply Side (Creator Economy):** Our current model treats the content pool as a distribution $g(z; u)$ controlled effectively by the platform. In reality, content is generated by *Strategic Creators* who react to algorithmic incentives. If the platform falls into the "Algorithmic Liquidity Trap" (prioritizing flux), creators will rationally respond by producing shorter, sensationalist content to maximize their own exposure. This *Supply-Side Adaptation* could create a secondary feedback loop, exacerbating the "Race to the Bottom" and making the ecosystem even harder to fix. Future work should model creators as a third player in this game.

2. **Competition and Network Effects:** We assume a single monopoly platform. A natural extension is to introduce *Platform Competition*, where rival algorithms compete for the same user attention budget. It remains an open question whether competition would force platforms to return to a "Quality-First" equilibrium or accelerate the acceleration towards even higher velocities. Additionally, incorporating social network effects (viral sharing) could offer new avenues for mechanism design beyond pure algorithmic matching.

3. **Heterogeneous User Types:** Our Mean-Field approach assumes a representative agent structure (with stochastic heterogeneity). Introducing distinct user types (e.g., "Passive Viewers" vs. "Active Searchers") could reveal distributional consequences, where the algorithm might exploit vulnerable sub-populations (e.g., minors with lower μ) while serving high-quality content to more selective users.

4. **Co-evolutionary Dynamics and Strategy Expansion:** Our Stackelberg framework models users as rational agents optimizing within a fixed physiological attention mechanism. In reality, the interaction may resemble a dynamic repeated game where users "learn" to counter algorithmic manipulation over time (e.g., developing "ad-blindness" or actively skipping content). However, we argue that the *Algorithmic Liquidity Trap* is robust because it exploits invariant physiological constraints (boredom decay μ and reward volatility σ) rather than transient behavioral loopholes. While users may evolve cognitive countermeasures, the platform's superior adaptation speed (Leader advantage) allows it to maintain the trap by shifting the locus of extraction. Future work could explicitly model this "Arms Race" using Evolutionary Game Theory.

References

Achdou, Y., Camilli, F., & Capuzzo-Dolcetta, I. (2012). Mean field games: numerical methods for the planning problem. *SIAM Journal on Control and Optimization*, *50*(1), 77–109.

Achdou, Y., & Capuzzo-Dolcetta, I. (2010). Mean field games: numerical methods. *SIAM Journal on Numerical Analysis*, *48*(3), 1136–1162.

Achdou, Y., Han, J., Lasry, J.-M., Lions, P.-L., & Moll, B. (2022). Income and wealth distribution in macroeconomics: A continuous-time approach. *The review of economic studies*, *89*(1), 45–86.

Aiyagari, S. R. (1995). Optimal capital income taxation with incomplete markets, borrowing constraints, and constant discounting. *Journal of political Economy*, *103*(6), 1158–1175.

Bass, F. M. (1969). A new product growth for model consumer durables. *Management science*, *15*(5), 215–227.

Bellman, R. (1966). Dynamic programming. *science*, *153*(3731), 34–37.

Bensoussan, A., & Lions, J. L. (1984). Impulse control and quasi-variational inequalities. *(No Title)*.

Boyd, J. P. (2001). *Chebyshev and fourier spectral methods*. Courier Corporation.

Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on scientific computing*, *16*(5), 1190–1208.

Carmona, R., Delarue, F., et al. (2018). *Probabilistic theory of mean field games with applications i-ii* (Vol. 3). Springer.

Carmona, R., & Laurière, M. (2021). Convergence analysis of machine learning algorithms for the numerical solution of mean field control and games i: The ergodic case. *SIAM Journal on Numerical Analysis*, *59*(3), 1455–1485.

Chang, J., & Cooper, G. (1970). A practical difference scheme for fokker-planck equations. *Journal of Computational Physics*, *6*(1), 1–16.

Chen, J., Dong, H., Wang, X., Feng, F., Wang, M., & He, X. (2023). Bias and debias in recommender system: A survey and future directions. *ACM Transactions on Information Systems*, *41*(3), 1–39.

Chen, M., Beutel, A., Covington, P., Jain, S., Belletti, F., & Chi, E. H. (2019). Top-k off-policy correction for a reinforce recommender system. In *Proceedings of the twelfth acm international conference on web search and data mining* (pp. 456–464).

Claisse, J., Ren, Z., & Tan, X. (2023). Mean field games with branching. *The Annals of Applied Probability*, *33*(2), 1034–1075.

Covington, P., Adams, J., & Sargin, E. (2016). Deep neural networks for youtube recommendations. In *Proceedings of the 10th acm conference on recommender systems* (pp. 191–198).

Fleming, W. H., & Soner, H. M. (2006). *Controlled markov processes and viscosity solutions*. Springer.

Folks, J. L., & Chhikara, R. S. (1978). The inverse gaussian distribution and its statistical

application—a review. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *40*(3), 263–275.

Gao, C., Li, S., Zhang, Y., Chen, J., Li, B., Lei, W., ... He, X. (2022). Kuairand: An unbiased sequential recommendation dataset with randomly exposed videos. In *Proceedings of the 31st acm international conference on information & knowledge management* (pp. 3953–3957).

Gardiner, C. W., et al. (2004). *Handbook of stochastic methods* (Vol. 3). springer Berlin.

Gershkov, A., & Moldovanu, B. (2009). Learning about the future and dynamic efficiency. *American Economic Review*, *99*(4), 1576–1587.

Glowinski, R. (2013). *Numerical methods for nonlinear variational problems*. Springer Science & Business Media.

Goldfarb, A., & Tucker, C. E. (2011). Privacy regulation and online advertising. *Management science*, *57*(1), 57–71.

Gordon, W. J., & Newell, G. F. (1967). Closed queuing systems with exponential servers. *Operations research*, *15*(2), 254–265.

Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018). Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International conference on machine learning* (pp. 1861–1870).

Han, J., Jentzen, A., & E, W. (2018). Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences*, *115*(34), 8505–8510.

Iyer, K., Johari, R., & Sundararajan, M. (2014). Mean field equilibria of dynamic auctions with learning. *Management Science*, *60*(12), 2949–2970.

Ke, T. T., Shen, Z.-J. M., & Villas-Boas, J. M. (2016). Search for information on multiple products. *Management Science*, *62*(12), 3576–3603.

Kushner, H. J. (1990). Numerical methods for stochastic control problems in continuous time. *SIAM Journal on Control and Optimization*, *28*(5), 999–1048.

Lanham, R. A. (2006). *The economics of attention: Style and substance in the age of information*. University of Chicago Press.

Lasry, J.-M., & Lions, P.-L. (2007). Mean field games. *Japanese journal of mathematics*, *2*(1), 229–260.

LeVeque, R. J. (2007). *Finite difference methods for ordinary and partial differential equations: steady-state and time-dependent problems*. SIAM.

Little, J. D., et al. (1961). A proof for the queuing formula. *Operations research*, *9*(3), 383–387.

Lobel, I., Patel, J., Vulcano, G., & Zhang, J. (2016). Optimizing product launches in the presence of strategic consumers. *Management Science*, *62*(6), 1778–1799.

Marden, J. R., Arslan, G., & Shamma, J. S. (2009). Cooperative control and potential games. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *39*(6), 1393–1407.

Media Rating Council (MRC) and Interactive Advertising Bureau (IAB). (2015, June). *Mrc viewable ad impression measurement guidelines (version 2.0)* (Industry Standard). Me-

dia Rating Council. Retrieved from https://www.iab.com/wp-content/uploads/2015/06/MRC-Viewable-Ad-Impression-Measurement-Guideline.pdf (Accessed: 2025-12-25)

Peskir, G., & Shiryaev, A. (2006). *Optimal stopping and free-boundary problems*. Basel: Birkhäuser Basel.

Saito, Y., Yaginuma, S., Nishino, Y., Sakata, H., & Nakata, K. (2020). Unbiased recommender learning from missing-not-at-random implicit feedback. In *Proceedings of the 13th international conference on web search and data mining* (pp. 501–509).

Stokey, N. L. (2008). *The economics of inaction: Stochastic control models with fixed costs*. Princeton University Press.

Trélat, E., & Zuazua, E. (2015). The turnpike property in finite-dimensional nonlinear optimal control. *Journal of Differential Equations*, *258*(1), 81–114.

Yang, Y., Luo, R., Li, M., Zhou, M., Zhang, W., & Wang, J. (2018). Mean field multi-agent reinforcement learning. In *International conference on machine learning* (pp. 5571–5580).

Zhu, Y. (2022). Regularity of kinetic fokker-planck equations in bounded domains. *arXiv preprint arXiv:2206.04536*.

A Mathematical Proofs

A.1 Proof of Proposition 2 (Threshold Structure)

The proof relies on the classical theory of optimal stopping for one-dimensional diffusions.

Let $\mathcal{L}^a$ denote the infinitesimal generator of the diffusion process Z_t. We consider the *waiting region* (continuation region) $\mathcal{C}_t = \{z \in \Omega \mid v(t,z) > K(t,u) - C\}$. Based on the results in Peskir and Shiryaev (2006) and Stokey (2008), the shape of the stopping region is determined by the monotonicity of the "local benefit of waiting".

For the CRRA utility function $u_{util}(z) = \frac{z^{1-\gamma}}{1-\gamma}$ with $\gamma > 1$ and the geometric Brownian motion dynamics, the net flow utility (holding cost vs. immediate reward) satisfies the single-crossing property. Specifically, the infinitesimal generator applied to the utility flow exhibits monotonicity in z: the marginal value of "waiting" (staying on the video) strictly increases with the attention state z.

Since the diffusion is non-degenerate ($\sigma > 0$), the Principle of Smooth Fit applies. The monotonicity ensures that the continuation region $\mathcal{C}_t$ is connected and takes the form of an upper set:

$$\mathcal{C}_t = (b^*(t), z_{\max}], \tag{35}$$

where the threshold $b^*(t)$ is the unique point satisfying the smooth-pasting condition. Consequently, the optimal stopping region is the complement $\mathcal{S}_t = [z_{\min}, b^*(t)]$. □

B Numerical Solution Details

B.1 A Fully Implicit Scheme with Mass Truncation

Given the implicit obstacle structure of the Quasi-Variational Inequality (QVI) and the non-local nature of the population dynamics, obtaining an analytic solution is mathematically infeasible. Therefore, this study constructs a numerical solution framework based on the *Fully Implicit Finite Difference Method (FDM)*. We adapt the iterative approach used in impulse control problems (Bensoussan & Lions, 1984) to handle the specific coupling between the user's stopping boundary and the traffic recycling mechanism, leveraging established schemes for Mean Field Games (Achdou, Camilli, & Capuzzo-Dolcetta, 2012; Achdou & Capuzzo-Dolcetta, 2010).

Approximation of the Infinite-Horizon Solution Although the theoretical model is formulated on an infinite time horizon $t \in [0, \infty)$, numerically resolving the system requires a finite temporal truncation. We approximate the infinite-horizon stationary equilibrium by simulating the coupled system over a sufficiently large finite time horizon $T_{\text{sim}} = 50$. This truncation is justified by two factors:

1. **Discounting Effect:** Given the discount rate ρ, the contribution of utility flows from $t > T_{\text{sim}}$ to the value function at $t = 0$ decays exponentially (proportional to $e^{-\rho T_{\text{sim}}}$), rendering the truncation error negligible.

2. **Convergence to Stationarity:** Our numerical experiments leverage the *Turnpike Property* of optimal control problems (Trélat & Zuazua, 2015). As shown in the results, the system dynamics (both the optimal boundary $b^*(t)$ and population density $f(t,\cdot)$) rapidly converge to a time-invariant state well before the terminal time. The time-stepping scheme thus serves as a fixed-point iteration to recover the stationary Mean-Field Equilibrium (MFE).

To eliminate the artifacts caused by the artificial terminal condition at T_{sim} (the terminal boundary layer), we report and analyze the solution over the interval $t \in [0, 0.8T_{\text{sim}}]$, where the system exhibits stable equilibrium behavior.

(1) Discretization Framework The continuous state space $\Omega = [z_{\min}, z_{\max}]$ is discretized into a uniform grid $\mathcal{Z}_h = \{z_i \mid z_i = z_{\min} + i\Delta z\}_{i=0}^{N_z}$. The time interval is discretized into N_t time steps $\mathcal{T}_h = \{t_n \mid t_n = n\Delta t\}_{n=0}^{N_t}$. To ensure numerical stability given the stiff source terms (Dirac delta) and the recycling flux, a *fully implicit time integration scheme* is employed (LeVeque, 2007).

(2) Backward Solution of the User HJB Equation For a fixed platform policy $\{a(t,z), u(t)\}$, the user's value function $v(t,z)$ is solved via backward time stepping. For the convection term in the operator $\mathcal{L}^a v$, an upwind difference scheme is adopted to ensure the monotonicity of the scheme (M-matrix property), a critical condition for convergence in stochastic control problems (Kushner, 1990):

$$\partial_z v(z_i) \approx \begin{cases} \frac{v_{i+1}-v_i}{\Delta z} & \text{if } \mu + a(t_n, z_i) \geq 0 \\ \frac{v_i - v_{i-1}}{\Delta z} & \text{if } \mu + a(t_n, z_i) < 0 \end{cases}$$

The optimal stopping constraint is enforced via a projection operation at each time step, similar to the Projected Successive Over-Relaxation (PSOR) method for obstacle problems (Glowinski, 2013):

$$v_i^n = \max\{\tilde{v}_i^n, K(t_n, u^n) - C\}$$

where $K(t_n, u^n)$ is the endogenous outside option computed using the current matching control $u(t_n)$. The optimal stopping boundary $b^*(t_n)$ is identified as the critical state separating the continuation and stopping regions.

(3) Forward Solution of the Non-Local FPE Unlike standard decoupled frameworks, our population dynamics involve a *feedback loop* via the boundary flux. Once $b^*(t)$ is obtained, we solve the Non-Local Fokker-Planck Equation forward in time using a conservative *Fully Implicit Scheme*.

Flux Calculation via Mass Truncation: Instead of computing the gradient at the boundary, we calculate the swipe flux Φ_{swipe} by explicitly integrating the probability mass that falls into the stopping region $z < b^*(t_n)$ during the implicit step. This approach aligns with the

conservative flux reconstruction methods proposed by Chang and Cooper (1970):

$$\Phi_{\text{swipe}}(t_n) = \frac{1}{\Delta t} \sum_{z_i < b^*(t_n)} f^*(t_n, z_i) \Delta z$$

This mass is physically removed from the system and a fraction $(1 - \epsilon)$ is recycled back via the source term $g(z; u)$. This ensures rigorous mass conservation and stability.

(4) Iterative Resolution of the QVI Fixed Point Since the obstacle $K(t, u)$ depends on the value function v itself, a fixed-point iteration is required (Bensoussan & Lions, 1984):

- **Initialization:** Guess an initial trajectory for $K^{(0)}(t)$.
- **Update Step:** In iteration m, solve HJB backward to get $v^{(m)}$, then update the coupling variable:
$$K_{\text{new}}(t) = \int_{z_{\min}}^{z_{\max}} v^{(m)}(t, z) g(z; u(t)) dz$$
- **Convergence:** Apply relaxation until $\|K^{(m+1)} - K^{(m)}\|_\infty < \text{Tol}$.

B.2 Parameter Configuration and Pseudocode

The model parameters are calibrated to ensure a stable S-curve growth trajectory and a meaningful stationary equilibrium. Table 3 details the specific settings used in the simulation. Algorithm 1 presents the detailed numerical procedure.

Table 3: Model Parameters (Calibrated for Stability)

Symbol	Value	Description
User & Content Parameters		
γ	2.1	Relative risk aversion
ρ	0.5	Time discount rate (High impatience)
μ	-0.0235	Natural attention decay (Boredom)
σ	0.3	Content volatility
C	1.0	Cognitive switching cost
Ecosystem Parameters		
ϵ	0.0125	Churn Rate (1.25% exit probability)
A	5.0	Base Acquisition Rate
$N_{\max}$	100.0	Theoretical Market Capacity
μ_g	1.65	Content Quality Mean
Numerical Settings		
T	50.0	Simulation Time Horizon
Δt	0.01	Time Step Size
N_z	400	Grid Points ($\Delta z = 0.025$)

Note: Parameters are calibrated to ensure a stable S-curve growth trajectory.

Algorithm 1 Fully Implicit Solver for User-Population Equilibrium

procedure SOLVEEQUILIBRIUM(Policy $\{a, u\}$, Params $A, \epsilon, N_{\max}$)
  Input: Control $\{a, u\}$, Market Params. **Output:** v^*, b^*, f^*, Φ
  Phase 1: User Optimization (Backward HJB)
  Construct implicit operators; Initialize implicit system.
  for $n \leftarrow N_t - 1$ **downto** 0 **do**
    Solve linear system for $\tilde{v}$; $v^n \leftarrow \max(\tilde{v}, K(t_n) - C)$
    $b^*(t_n) \leftarrow \min\{z \mid v(t_n, z) > K(t_n) - C\}$ ▷ Identify threshold
    if Endogenous K **then** Update $K(t_n)$ via fixed-point iteration
    end if
  end for
  Phase 2: Population Dynamics (Forward FPE)
  Initialize $f(0, z)$ and $N(0)$; Construct **Fully Implicit Matrix** $\mathbf{M}_{FPE}$
  for $n \leftarrow 0$ **to** $N_t - 1$ **do**
    1. Acquisition: $\alpha \leftarrow \max(0, 1 - N^n/N_{\max})$; $S_{new} \leftarrow \alpha A \delta(z - z_p)$
    2. Implicit Solve: $\mathbf{M}_{FPE} \cdot \mathbf{f}^* = \mathbf{f}^n + \Delta t \cdot S_{new}$
    3. Flux & Cut: $M_{stop} \leftarrow \sum_{z<b^*} \mathbf{f}^* \Delta z$; $\Phi(t_n) \leftarrow M_{stop}/\Delta t$
      Set $\mathbf{f}^*(z) = 0$ for all $z < b^*(t_{n+1})$ ▷ Truncate distribution
    4. Recycling: $\mathbf{f}^{n+1} \leftarrow \mathbf{f}^* + (1 - \epsilon)\Phi(t_n) g(z) \Delta t$; Update $N(t_{n+1})$
  end for
  return $\{v, b, f, \Phi\}$
end procedure

C Derivation of the Master Equation on the Space of Measures

In this appendix, we derive the Hamilton-Jacobi-Bellman (HJB) equation for the platform's value function $V(N, m)$, utilizing the structure-scale decomposition introduced in Section 3.3.1.

C.1 Variational Derivatives and Chain Rule

Let $\mathcal{W}(f)$ be the value functional on the space of unnormalized measures. We decompose $f = N \cdot m$. To establish the Master Equation in terms of (N, m), we first relate the functional derivative $\frac{\delta \mathcal{W}}{\delta f}$ to the derivatives with respect to N and m.

Consider a perturbation $f \to f + \delta f$. The variations in N and m are:

$$\delta N = \int \delta f(z) dz, \quad \delta m(z) = \frac{\delta f(z)}{N} - \frac{f(z)}{N^2} \delta N.$$

The total variation of the value function is:

$$\begin{aligned} \delta \mathcal{W} &= \frac{\partial V}{\partial N} \delta N + \int \frac{\delta V}{\delta m}(z) \delta m(z) \, \mathrm{d}z \\ &= \int \left[\frac{\partial V}{\partial N} + \frac{1}{N} \left(\frac{\delta V}{\delta m}(z) \right. \right. \\ &\qquad \left. \left. - \int \frac{\delta V}{\delta m}(y) m(y) \, \mathrm{d}y \right) \right] \delta f(z) \, \mathrm{d}z. \end{aligned} \tag{36}$$

Thus, the shadow price of the unnormalized density is given by the *Decomposition Identity*:

$$\begin{aligned} \frac{\delta \mathcal{W}}{\delta f}(z) &= \frac{\partial V}{\partial N} + \frac{1}{N}\hat{\lambda}(z), \\ \text{where } \hat{\lambda}(z) &= \frac{\delta V}{\delta m}(z) - \mathbb{E}_m\left[\frac{\delta V}{\delta m}\right]. \end{aligned} \tag{37}$$

C.2 The Bellman Principle

The dynamic programming principle states:

$$\begin{aligned} rV(N,m) = \sup_{a,u,A} \Bigg\{ & \pi(N,m,a,u,A) \\ & + \frac{\partial V}{\partial N}\frac{dN}{dt} + \left\langle \frac{\delta V}{\delta m}, \frac{\partial m}{\partial t} \right\rangle \Bigg\} \end{aligned} \tag{38}$$

Substituting the dynamics derived in the main text:

- **Scale Dynamics:** $\frac{dN}{dt} = A - \epsilon N \phi_{swipe}(m)$
- **Structural Dynamics:** $\frac{\partial m}{\partial t} = \mathcal{L}^*[a]m + \frac{A}{N}(\delta_{z_p} - m) + (1-\epsilon)\phi_{swipe}(m)(g-m)$

C.3 Optimization of Controls

Before optimizing specific controls, we must explicitly handle the duality pairing involving the differential operator $\mathcal{L}^*[a]$.

Green's Identity and Boundary Terms: The term $\langle \frac{\delta V}{\delta m}, \mathcal{L}^*[a]m \rangle$ involves the Fokker-Planck operator. Using Green's identity (integration by parts twice) on the domain $\Omega = [b^*, z_{max}]$ with the absorbing boundary condition $m(b^*) = 0$:

$$\begin{aligned} \int_{b^*}^{z_{\max}} \frac{\delta V}{\delta m}\mathcal{L}^*[a]m \, dz = & \int_{b^*}^{z_{\max}} \mathcal{L}^a\left[\frac{\delta V}{\delta m}\right] m \, dz \\ & - \frac{\delta V}{\delta m}(b^*) \cdot \underbrace{\frac{1}{2}\sigma^2 (b^*)^2 \partial_z m(b^*)}_{\phi_{\text{swipe}}(m)} \end{aligned} \tag{39}$$

The boundary term $-\frac{\delta V}{\delta m}(b^*)\phi_{swipe}(m)$ arises because while the density m vanishes at the absorbing boundary, its gradient (flux) is non-zero. This term represents the *value destruction* of a user hitting the churn threshold. Note that we assume zero flux at the reflective boundary z_{max}.

We now maximize the Hamiltonian term by term:

1. Retention Control $a(z)$: The drift control $a(z)$ appears in the generator $\mathcal{L}^a[\frac{\delta V}{\delta m}] = (\mu + a(z))z\partial_z(\frac{\delta V}{\delta m}) + \dots$. Combining with the cost term $-\frac{\xi_1}{2}N\int a^2 m dz$:

$$\sup_a \int \left(-\frac{\xi_1}{2}Na^2 + az\partial_z\frac{\delta V}{\delta m}\right) m dz \implies a^*(z) = \frac{z}{\xi_1 N}\partial_z \frac{\delta V}{\delta m}$$

2. Acquisition Control A: Terms involving A:

$$A\left(\frac{\partial V}{\partial N}+\frac{1}{N}\frac{\delta V}{\delta m}(z_p)-\frac{1}{N}\int\frac{\delta V}{\delta m}m dz\right)-\Psi(A)$$

Using the identity (37), the coefficient of A is simply $\frac{\delta \mathcal{W}}{\delta f}(z_p)$. The maximum is $H_{acq}(\frac{\delta \mathcal{W}}{\delta f}(z_p))$.

3. Matching Control u: Terms involving u (via $g(z;u)$ and costs) and the flux ϕ_{swipe} (which collects the revenue, recycling, scale loss, *and boundary value loss* terms):

$$\begin{aligned}\text{Terms} = \phi_{\text{swipe}}\Bigg[&\underbrace{\lambda_2 N}_{\text{Ads}}+\underbrace{(1-\epsilon)\int\frac{\delta V}{\delta m}g(z;u)\,dz}_{\text{Recycling Gain}}\\&\underbrace{-(1-\epsilon)\int\frac{\delta V}{\delta m}m\,dz}_{\text{Mean Value Correction}}-\underbrace{\epsilon N\frac{\partial V}{\partial N}}_{\text{Churn Loss}}\\&-\underbrace{\frac{\delta V}{\delta m}(b^*)}_{\text{Boundary Loss}}\Bigg]-\frac{\xi_2}{2}u^2\end{aligned} \tag{40}$$

Let $\Delta V_{swipe}(u)$ denote the net marginal value of a swipe (the term in brackets). The optimal matching u^* satisfies:

$$\xi_2 u^* = \phi_{swipe}(1-\epsilon)\int\frac{\delta V}{\delta m}\frac{\partial g}{\partial u}(z;u^*)dz$$

C.4 The Final Equation

Collecting all terms yields the explicit Master Equation presented in Eq. (21). The inclusion of the boundary term $-\frac{\delta V}{\delta m}(b^*)\phi_{swipe}$ ensures the mathematical consistency of the Mean-Field Game on a domain with absorbing boundaries.

D Numerical Solution Details

D.1 MDP Formulation and Spectral Action Embedding

We reformulated the Mean Field Control (MFC) problem as a Markov Decision Process (MDP) defined by the tuple $(\mathcal{S},\mathcal{A},\mathcal{P},\mathcal{R},\gamma)$ as follows:

State Space $\mathcal{S}$. The state at time t is represented by the macroscopic configuration of the system, comprising the total population and the discretized density distribution of user attention:

$$\mathbf{s}_t = [N(t), m(t,z_1), m(t,z_2), \ldots, m(t,z_{N_z})]^\top \in \mathbb{R}^{N_z+1} \tag{41}$$

where $\{z_i\}_{i=1}^{N_z}$ are the spatial grid points. This high-dimensional state input allows the neural network to capture the fine-grained structure of the population distribution (Carmona & Laurière, 2021).

Spectral Action Space $\mathcal{A}$. A core challenge is that the action $a(t,\cdot)$ is a function in an infinite-dimensional space $L^2(\Omega)$. To mitigate the *curse of dimensionality* (Bellman, 1966), we employ a spectral parameterization method. We project the control function onto a subspace spanned by the first K Chebyshev polynomials $T_k(\cdot)$ (Boyd, 2001):

$$a(t,z) \approx \sum_{k=0}^{K-1} w_k(t) T_k(\tilde{z}), \quad \tilde{z} = \frac{2z - z_{max}}{z_{max}} \in [-1,1] \tag{42}$$

Consequently, the action vector output by the agent is finite-dimensional:

$$\mathbf{u}_t = [u(t), A(t), w_0(t), \dots, w_{K-1}(t)]^\top \in \mathbb{R}^{K+2} \tag{43}$$

This spectral embedding reduces the action dimension from $N_z \approx 400$ to $K+2 \approx 12$, while naturally enforcing the spatial smoothness of the algorithmic intervention $a(t,z)$, which is crucial for the stability of the HJB solver.

Reward Function $\mathcal{R}$. The immediate reward reflects the platform's instantaneous profit flow:

$$r_t = N(t)\left(\lambda_1 \bar{z}_{ret} + \lambda_2 \Phi_{swipe}\right) - \left(\frac{\xi_1}{2}\|a\|_2^2 + \frac{\xi_2}{2}u^2 + \Psi(A)\right) \tag{44}$$

where $\bar{z}_{ret}$ is the average retention utility and Φ_{swipe} is the flux of swiping users.

D.2 Algorithm: Spectral Soft Actor-Critic

We adopt the Soft Actor-Critic (SAC) algorithm (Haarnoja, Zhou, Abbeel, & Levine, 2018), an off-policy maximum entropy DRL method. The interaction between the Agent (Platform) and the Environment (User Population) is governed by a physics-based engine that solves the coupled HJB-FPK system at each time step. The complete solution procedure is detailed in Algorithm 2.

D.3 Computational Infrastructure

To ensure the reproducibility of our numerical results, we detail the hardware and software environment used for the simulations and reinforcement learning training processes.
Hardware Specifications. All computations were executed on a high-performance computing server equipped with:

- **GPU:** NVIDIA Tesla P100 (16GB VRAM), ensuring efficient parallel computation for the neural network updates.
- **CPU:** Intel Xeon Platinum 8260 (2.30GHz, 12 vCPUs).
- **Memory:** 60GB RAM to handle the storage of the Experience Replay Buffer (size 10^6 transitions).

Algorithm 2 Spectral SAC with Physics-Informed Environment

Input: Basis $\{T_k\}_{k=0}^{K-1}$; Params $N_{\max}, \epsilon$; Rates α, β **Output:** Policy π_ϕ^*, Value Q_θ^*
Initialize Actor π_ϕ, Critic Q_θ, and Replay Buffer $\mathcal{D}$
for episode $m = 1$ **to** M **do**
    Initialize state $\mathbf{s}_0 = [N_0, \mathbf{f}_0]$
    for step $t = 0$ **to** $T_{horizon}$ **do**

        Phase 1: Spectral Action (Leader). Sample $\mathbf{u}_t \sim \pi_\phi(\cdot|\mathbf{s}_t)$ and reconstruct field:
            $a(z) \leftarrow \sum_{k=0}^{K-1} w_k T_k(\tilde{z})$ where $\mathbf{u}_t = [u, A, \{w_k\}]$ and $\tilde{z} \in [-1, 1]$
        Phase 2: User Equilibrium (HJB). Solve QVI via fixed-point iteration:
            Find $v(z)$ s.t. $\min(\mathcal{L}v - \rho v + U, v - (K - C)) = 0$
            Identify threshold: $b_t^* \leftarrow \min\{z \mid v(z) > K(u) - C\}$

        Phase 3: Population Dynamics (Implicit FPE).
            1. Acquisition: $S_{new} \leftarrow \max(0, 1 - N_t/N_{\max}) \cdot A \cdot \mathcal{N}(z)$
            2. Implicit Solve: $\mathbf{M}_{FPE} \cdot \mathbf{f}^* = \mathbf{f}_t + \Delta t \cdot S_{new}$
            3. Flux & Cut: $\Phi_t \leftarrow (\sum_{z<b_t^*} \mathbf{f}^* \Delta z)/\Delta t$; Set $\mathbf{f}^*(z) \leftarrow 0$ for $z < b_t^*$
            4. Recycling: $\mathbf{f}_{t+1} \leftarrow \mathbf{f}^* + (1 - \epsilon)\Phi_t g(z; u)\Delta t$; Update N_{t+1}

        Phase 4: Learning. Store $(\mathbf{s}_t, \mathbf{u}_t, r_t, \mathbf{s}_{t+1})$ in $\mathcal{D}$.
            Update Critic θ (Bellman) and Actor ϕ (Max Entropy $J = \mathbb{E}[Q - \alpha \log \pi]$).
    end for
end for
return Optimized Policy π_ϕ^*

Software Stack. The solution algorithm was implemented using **Python 3.10** and **PyTorch 2.0.1** with CUDA 11.8 support. The Mean-Field Game solver and the Soft Actor-Critic (SAC) agent were built upon the *Gymnasium* interface standards and the *Stable Baselines 3* library.

D.4 Hyperparameter Configuration and Reward Shaping

To ensure the stability of the Spectral SAC algorithm and accurate calibration against the empirical moments, we employed specific reward shaping coefficients. The training process utilized the Stable Baselines 3 implementation of Soft Actor-Critic. The key hyperparameters used in the baseline experiment (500,000 steps) are detailed in Table 4.

E Solver Validation and Robustness Check

To validate the robustness of the solution and the structural complexity of the identified equilibrium, we compared the Deep Reinforcement Learning (DRL) solver against a standard numerical optimization baseline. This comparison serves two purposes: (1) to verify that the "Retention Trap" strategy is a superior global equilibrium rather than a computational artifact, and (2) to illustrate the limitations of local search methods in handling the complex feedback loops inherent in traffic recycling.

Table 4: Reward Function Coefficients and Hyperparameters

Parameter	Value	Description
Model Structural Parameters (Matching Eq. 16)		
λ_1	20.0	Weight on retention utility flow (Retention Priority)
λ_2	0.01	Weight on swipe flux (Impression Priority)
ξ_1	1.0	Coefficient for retention control cost ($\frac{\xi_1}{2}\|a\|^2$)
ξ_2	0.5	Coefficient for matching investment cost ($\frac{\xi_2}{2}u^2$)
ψ_0	0.01	Coefficient for acquisition cost ($\Psi(A)$)
Numerical Regularization & Training		
c_{smooth}	20.0	*Spectral Smoothness Penalty* (prevents policy oscillation)
Total Steps	500,000	Convergence threshold for SAC
Learning Rate	3×10^{-4}	Actor and Critic networks (Adam Optimizer)
Batch Size	256	Experience replay sampling size
γ	0.99	Discount factor

Note: Parameters $\lambda_1, \lambda_2, \xi_1, \xi_2$ correspond to the objective function defined in Equation (16). The smoothness penalty c_{smooth} is an auxiliary regularization term introduced for the spectral numerical stability.

E.1 Comparison Setup

We utilized the Sequential Least Squares Programming (SLSQP) algorithm as a baseline.

- **Objective:** Maximize the same long-term average reward function (Eq. 15).
- **Challenge:** The optimization landscape is characterized by the coupling between the free boundary b^* and the non-local flux term. Such coupling often creates multiple local optima.

E.2 Key Observation: Structural Divergence

The comparison results (Figure 6) reveal a fundamental divergence in the identified strategies:

- **Local Optima (Baseline):** The gradient-based solver converges to a smooth, near-linear control policy. While mathematically feasible, this strategy fails to leverage the "traffic recycling" mechanism effectively, resulting in a lower equilibrium population.
- **Global Structure (DRL):** The SAC agent identifies a highly non-linear "Trap and Release" structure. By aggressively intervening only near the churn boundary, the agent achieves a significantly higher objective value. This suggests that the "Retention Trap" is a sophisticated strategic response to the ecosystem dynamics, which requires global exploration to be discovered.

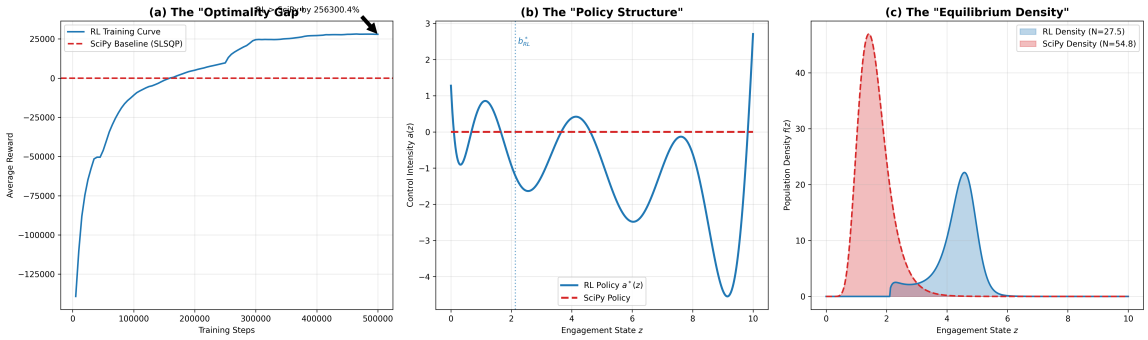

Figure 6: **Benchmarking DRL vs. Gradient-Based Optimization (Baseline). (a) Optimality Gap:** The RL training curve (blue) rapidly surpasses the best solution found by SLSQP (red dashed line). **(b) Policy Structure:** The gradient solver defaults to a trivial linear strategy, whereas the RL agent learns a complex, non-linear control field $a^*(z)$ with a distinct "retention trap" near the churn threshold b^*. **(c) Equilibrium Density:** The RL policy sustains a robust population distribution (blue area) compared to the baseline (red area).

E.3 Robustness in High-Entropy Regimes

We further evaluated the solvers in two stress-test scenarios: the High-Velocity Regime (Experiment A) and the High-Cost Regime (Experiment B).

1. Experiment A: The "Clickbait" Regime ($\lambda_2 = 5.0, \lambda_1 = 5.0$). This regime is characterized by rapid user turnover. As shown in Figure 7:

- **Solver Failure vs. Success:** The gradient solver collapses into a "safety mode" (near-zero control), failing to harness the traffic recycling mechanism. In contrast, the RL agent discovers the "Trap and Release" strategy, yielding a reward improvement of over $1.2 \times 10^6\%$.
- **Necessity of Global Search:** This colossal gap confirms that in high-flux platform economies, the profit landscape is highly irregular. Deep Reinforcement Learning is not merely an alternative but a necessity to locate the narrow region of the parameter space that supports the "High-Velocity Prosperity" equilibrium.

2. Experiment B: High-Cost Regime ($\xi_1 = 10.0$). To simulate resource-constrained environments (detailed in Figure 8), we increased the control cost ten-fold.

- **Strategic Divergence:** The gradient solver converges to a trivial "do-nothing" strategy ($a \approx 0$) to minimize costs. The RL agent, however, executes an *expansion strategy*, identifying that the marginal revenue from a larger population outweighs the high marginal cost. This results in an equilibrium population of $N^* \approx 69.3$, nearly double that of the gradient-based solution.

E.4 Robustness Check: Scale Invariance and Algorithmic Overheating

To rigorously verify that the "Algorithmic Liquidity Trap" identified in Experiment A is not a numerical artifact of lower absolute reward weights, we performed a normalization

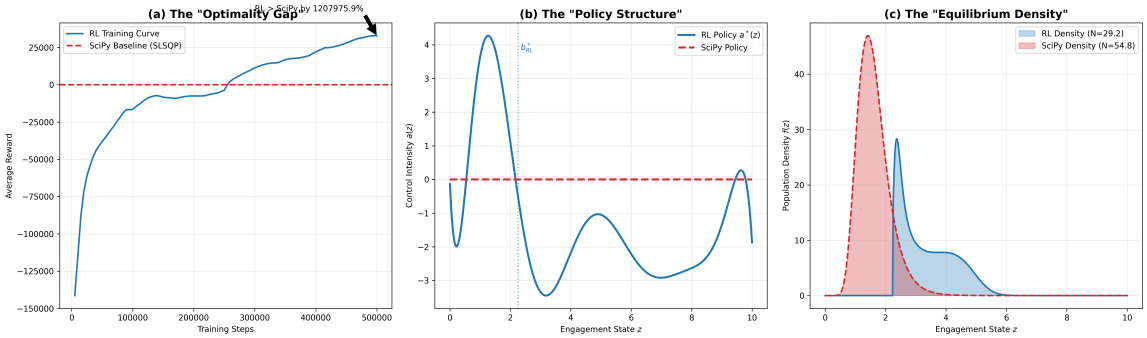

Figure 7: **Benchmark in High-Velocity Regime (Experiment A). (a) Divergence:** The static solver (red) fails to find the high-flux equilibrium, resulting in near-zero value. The RL agent (blue) successfully learns the high-reward strategy. **(b) The "Retention Trap":** The RL policy $a^*(z)$ exhibits a sharp, non-monotonic spike followed by relaxation—a structure missed by the gradient solver.

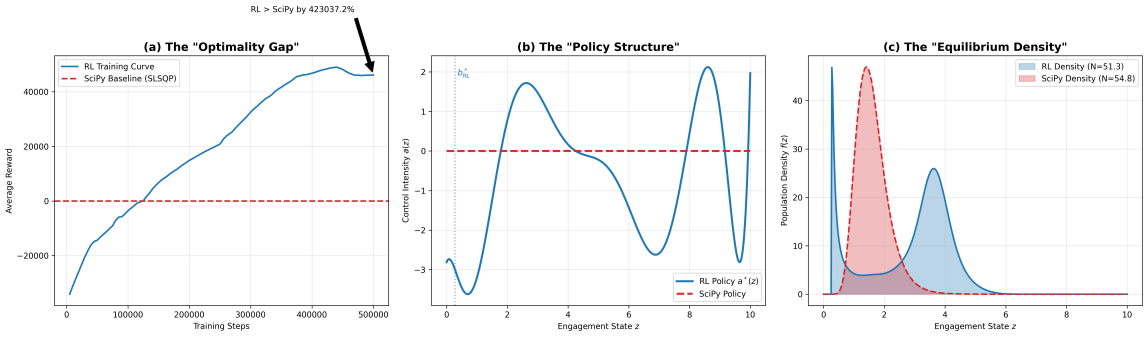

Figure 8: **Benchmark in High-Cost Regime (Experiment B). (a) The "Inaction" Trap:** The gradient solver retreats to zero intervention. The RL agent discovers a high-investment strategy. **(b) Oscillatory Control:** Under high cost pressure, the RL policy adopts a complex structure to maintain engagement efficiently.

experiment (Experiment C).

Experimental Design. We constructed a scenario that matches the *Total Weight Budget* of the Baseline (≈ 20) but retains the *Monetization Ratio* of the Ad-Centric regime ($\mathcal{R} = 1$).

- **Baseline:** $\lambda_1 = 20.0, \lambda_2 = 0.01$ (Total ≈ 20).
- **Experiment A (Original):** $\lambda_1 = 5.0, \lambda_2 = 5.0$ (Total $= 10$).
- **Experiment C (Robustness):** $\lambda_1 = 10.0, \lambda_2 = 10.0$ (Total $= 20$).

Results and Interpretation. The results are visualized in Figure 9.

1. **Validation of the Trap (Flux Consistency):** The equilibrium swipe flux in Experiment C converged to $\Phi^* \approx 108.75$. This value is remarkably close to Experiment A ($\Phi^* \approx 118.75$) and represents a $> 4\times$ increase over the Baseline. This confirms that the high-velocity trap is a structural equilibrium determined by the ratio $\mathcal{R}$, robust to parameter scaling.

2. **Algorithmic Overheating (Population Drop):** We observed that the steady-state population in Experiment C ($N^* \approx 18.34$) was lower than in Experiment A ($N^* \approx 43.71$). This is a structurally consistent outcome of *Relative Cost Reduction*. In the objective function $J = \lambda_1 z + \lambda_2 \Phi - \frac{\xi}{2} a^2$, doubling the rewards ($\lambda$) while holding the control cost coefficient constant ($\xi = 1.0$) effectively makes algorithmic intervention "cheaper" relative to the potential payoff. Consequently, the DRL agent adopted a more aggressive "harvesting" policy—pushing the recycling velocity to its limit to extract immediate flux rewards, even at the cost of higher user burnout and a lower carrying capacity. We term this phenomenon *"Algorithmic Overheating,"* where excessive incentive intensity (without corresponding cost constraints) accelerates ecosystem depletion. This aggressive harvesting drives the implicit dwell time down even further to $\tau_{robust} \approx 18.34/108.75 \approx 0.17$s, indicating that higher reward scales without cost adjustments exacerbate the race to the bottom.

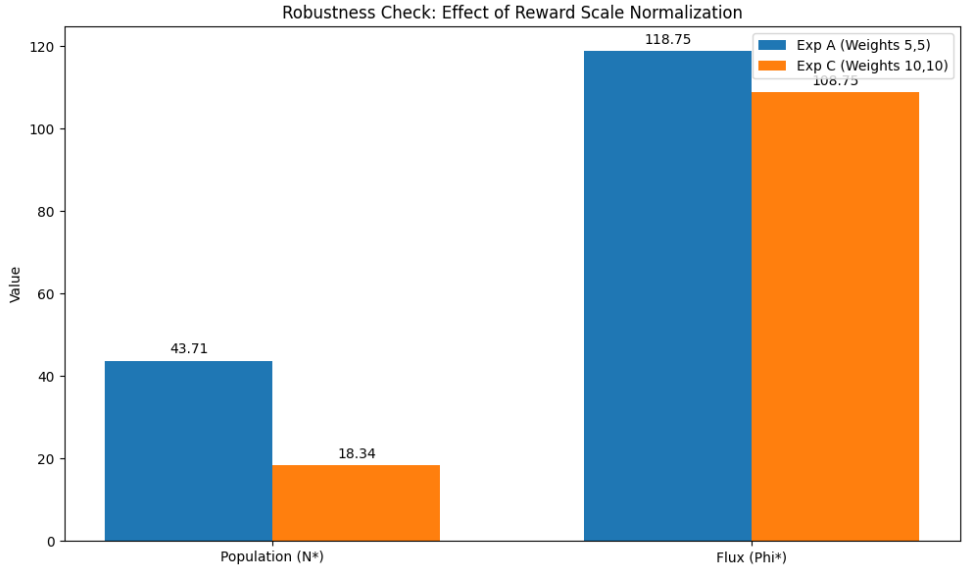

Figure 9: **Robustness Check: Effect of Reward Scale Normalization.** Comparing Experiment A ($\lambda = 5, 5$) with the scaled Experiment C ($\lambda = 10, 10$). **Right Bar:** The Swipe Flux (Φ^*) remains robustly high (> 100), confirming the persistence of the Liquidity Trap regardless of scale. **Left Bar:** The drop in Population (N^*) in Exp C illustrates the "Overheating" effect, where relatively cheaper control costs lead to aggressive over-harvesting.