CONTROLLING CREATOR EXPOSURE UNDER RECOMMENDATION FEEDBACK

Anonymous authors
Paper under double-blind review

ABSTRACT

Recommendation platforms allocate visibility among creators and shape users' content menus, while engagement feeds back into subsequent allocations. Controlling information oligopoly and information cocoons requires coordinating global exposure concentration and per-user variety while accounting for the randomness of serving. In this paper, we formulate the Constrained Welfare Occupancy Program (CWOP), a request-level allocation model that maximizes expected utility under concentration ceilings and per-user entropy floors. Our analysis separates the utility cost of these targets from optimization error and establishes why feasible plans can still violate targets during finite-window serving. We derive policy-dependent safety margins that provide simultaneous exposure guarantees across adaptive windows under a known request distribution and conditionally independent sampling, and bound the utility cost of certification. An anonymized real-platform audit finds that exposure concentration is approximately eight times audience-stock concentration on a matched creator subset, measured by the Herfindahl–Hirschman index, while the estimated elasticity of audience growth with respect to audience stock is approximately 0.7. These observational findings motivate direct exposure control, while our framework makes its feasibility, reliability, and utility tradeoffs explicit.

1 INTRODUCTION

Recommendation platforms allocate exposure, shaping creators' access to audiences and users' menus (Covington et al., 2016). Exposure generates engagement signals that can attract further exposure (Chaney et al., 2018; Jiang et al., 2019; Mansoury et al., 2020). This feedback can foster two linked problems: *information oligopoly*, where a few creators dominate visibility, and *information cocoons*, where users repeatedly encounter a narrow set of creators (Anwar et al., 2024; Sukiennik et al., 2024).

Classical reinforcement models explain how concentration[1] can persist. They do not determine how reliably a platform can achieve a target exposure distribution. Under independent single-item serving, a creator whose planned share equals the cap exceeds that cap with probability tending to one-half as the window grows (Proposition 2). This holds even for an exact optimizer: optimization accuracy alone cannot prevent sampling violations at the boundary (Section 4). We introduce a safety margin between planned exposure and the target (Figure 1).

The **Constrained Welfare Occupancy Program (CWOP)** addresses these two problems through complementary exposure constraints: aggregate HHI and share ceilings curb creator dominance, while per-user entropy floors broaden each user's served menu. CWOP maximizes expected utility subject to these targets. Our analysis separates utility lost to the targets, optimization error, and safety margins for random serving. PD and the conic reference implement the same program; the latter is more accurate on our benchmarks, and both benefit from safety margins.

Our contributions are:

[1]The Herfindahl–Hirschman index (HHI) is the sum of squared shares (U.S. Department of Justice, Antitrust Division, 2024). For fractional exposure shares, $\mathrm{HHI}(p) = \sum_c p_c^2$ ranges from $1/K$ for K equally exposed creators to 1 for a monopoly.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

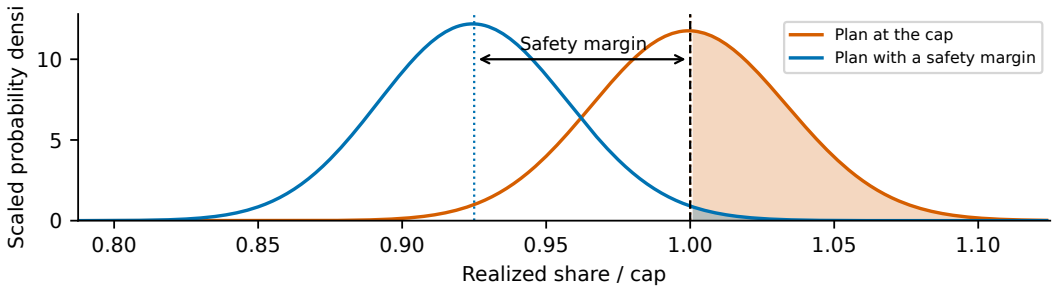

Figure 1: Illustration using exact binomial probabilities: $N = 10{,}000$ requests, cap 0.08, and planned shares equal to 1.000 or 0.925 times the cap. Both axes are transformed to share/cap units. Shading marks the probability of exceeding the cap. The illustrated margin is not a joint HHI/entropy certificate.

- We formulate joint finite-window exposure control for two coupled recommendation problems: information oligopoly across creators and information cocoons within users' served menus. An explicit request–candidate model represents global and per-user exposure as affine functions of a randomized allocation policy. CWOP maximizes expected utility under concentration ceilings and entropy floors, distinguishing a feasible plan from compliant randomized delivery (Sections 2–3).
- We derive optimization and serving certificates for this model: feasible-witness bounds separate target cost from solver error, and policy-dependent safety margins control caps, HHI, and entropy simultaneously across adaptive serving windows. A necessity result establishes the local $N^{-1/2}$ cap-margin order, while witness mixing bounds the score cost of certified policies (Sections 3–4).
- We evaluate both exposure objectives and their utility cost using independent conic solves, closed-loop feedback simulations, and repeated frozen-policy sampling. Matched comparisons on synthetic graphs and a public Last.fm graph with learned preference proxies assess numerical accuracy, target violations, and the safety–utility tradeoff. Relative reporting separates the cost of the constraints from optimization error and additional sampling protection (Section 5).

Related work. Exposure fairness includes amortized attention (Biega et al., 2018), fair policy learning (Singh & Joachims, 2019), joint user/item diversity (Wang & Joachims, 2021), Lorenz-efficient rankings (Do et al., 2021), and calibration (Steck, 2018). The distinction between fairness in expectation and realized rankings is established. Singh et al. (2021) study randomized ranking under merit uncertainty; Gorantla et al. (2023) sample rankings satisfying group-representation constraints ex post. Our targets instead concern nonlinear aggregate and per-type exposure over random serving windows. Mladenov et al. (2020) connect constrained matching to provider viability and long-term ecosystem welfare.

FairCo (Morik et al., 2020), FORM (Chen et al., 2024), and Offr (Usunier et al., 2022) combine allocation with online control or learning. We compare explicit batch adaptations under the same known-score protocol. Constraint tightening is established in chance-constrained optimization (Nemirovski & Shapiro, 2006) and safe constrained RL (Liu et al., 2021). Slater-based multiplier bounds and primal averaging are also standard (Nedić & Ozdaglar, 2009; Zinkevich, 2003). We apply these tools to joint HHI, share, and entropy targets, derive policy-dependent certificates, and measure their utility cost under adaptive feedback. Appendix I develops these comparisons and their mathematical foundations; Appendix A reviews the motivating urn models.

2 Preliminaries and Problem Formulation

For each serving window, the platform receives a request distribution, eligible slates, and utility scores. It chooses a randomized policy, serves from that policy, and may update the next window's

scores using observed outcomes. We first formalize the within-window decision; Section 4 treats repeated planning.

2.1 Request Occupancy and Exposure

Let $\mathcal{U}$ be a finite set of request types with probabilities $q_u > 0$, $\sum_{u\in\mathcal{U}} q_u = 1$. Each type has a finite candidate set $\mathcal{C}_u$ and a finite set $\mathcal{A}_u$ of admissible slates $a = (a_1, \ldots, a_L)$ with distinct creators $a_\ell \in \mathcal{C}_u$. Position weights satisfy $v_\ell \geq 0$ and $\sum_{\ell=1}^{L} v_\ell = 1$. Define the exposure incidence

$$z_{u,a,c} := \sum_{\ell=1}^{L} v_\ell \mathbf{1}\{a_\ell = c\}. \tag{1}$$

The request occupancy is $d_{u,a} = q_u x_{u,a}$, where x_u is a conditional distribution over $\mathcal{A}_u$. Its finite polytope is

$$\mathcal{D} := \left\{ d \geq 0 : \sum_{a\in\mathcal{A}_u} d_{u,a} = q_u \ \forall u \right\}. \tag{2}$$

Let $\mathcal{C} = \bigcup_u \mathcal{C}_u$. Aggregate and per-user planned exposure are the affine marginals

$$R_c(d) := \sum_{u,a} d_{u,a} z_{u,a,c}, \qquad S_{u,c}(d) := q_u^{-1} \sum_{a} d_{u,a} z_{u,a,c}. \tag{3}$$

Thus $\sum_c R_c(d) = 1$ and $S_u(d)$ is a probability vector on creators accessible to type u. For a specified per-request utility score $r_{u,a}$, let $W(d) := \sum_{u,a} d_{u,a} r_{u,a}$. We use "welfare" to mean this expected score and "occupancy" to mean the joint request–action distribution. The capitalized $R(d)$ and $S_u(d)$ denote planned marginals and are distinct from the time-indexed feedback shares ρ_t and $\sigma_{u,t}$. An entropy floor is per user only when the type identifies that user; aggregating users into one request type does not guarantee each individual's variety.

Interpreting concentration and variety. For K creators with exposure shares $p_c \geq 0$ summing to one, $\mathrm{HHI}(p) = \sum_c p_c^2$ ranges from $1/K$ (equal exposure) to 1 (all exposure goes to one creator). For example, four equal shares yield HHI 0.25, whereas shares $(0.7, 0.1, 0.1, 0.1)$ yield 0.52. Its reciprocal, $1/\mathrm{HHI}(p)$, is the *effective creator count*: the number of equally exposed creators with the same concentration, not the actual number reached. To measure per-user variety, we use entropy $H(p) = -\sum_c p_c \log p_c$, with $0 \log 0 = 0$ and natural logarithms. Entropy ranges from zero (one creator) to $\log K$ (equal shares).

2.2 Feasible Exposure Targets

A *strict feasible witness* is a policy satisfying every target with positive slack. It certifies that the chosen targets are jointly attainable on the candidate graph; uniform exposure need not be attainable when eligibility is restricted.

Assumption 1 (Request-level Slater condition)**.** Every $\mathcal{A}_u$ is nonempty, and there exist $d^0 \in \mathcal{D}$ and $\delta_0 > 0$ such that $\mathrm{HHI}(R(d^0)) \leq \eta - \delta_0$, $R_c(d^0) \leq \kappa - \delta_0$ for every creator, and $H(S_u(d^0)) \geq h_0 + \delta_0$ for every request type. Necessary checks include $h_0 < \log|\mathcal{C}_u|$, $\eta > 1/|\mathcal{C}|$, and $\kappa > 1/|\mathcal{C}|$. These strict checks are necessary for positive slack, but candidate restrictions can make them insufficient.

We define the feasible set

$$\begin{aligned} \mathcal{F} = \Big\{ d \in \mathcal{D} :\ & \mathrm{HHI}(R(d)) \leq \eta,\ \ \|R(d)\|_\infty \leq \kappa, \\ & H(S_u(d)) \geq h_0\ \ \forall u \Big\}, \end{aligned} \tag{4}$$

The ceilings (η, κ) operationalize control of information oligopoly; the floor h_0 controls exposure-based information cocoons. They require at least $1/\eta$ effective creators globally, at most share κ for any creator, and entropy at least h_0 for each type. These chosen levels specify the degree of control. **CWOP** solves the constrained program

$$\hat{d} = \arg\max_{d\in\mathcal{F}} W(d), \tag{5}$$

which is convex because W is linear, the marginals are affine, HHI and the maximum share are convex, and entropy super-level sets are convex (Appendix B.3).

Under Assumption 1, compactness and continuity give an optimizer. Write $W_g^* = \max_{d\in\mathcal{D}} W(d)$ and $W_c^* = W(\hat{d})$. For any feasible computed policy d, the identity

$$W_g^* - W(d) = \underbrace{W_g^* - W_c^*}_{\text{cost of the targets } \Lambda(\mathcal{F})} + \underbrace{W_c^* - W(d)}_{\text{controller gap}}$$

separates two different questions: how costly are the targets, and how well are they optimized? Neither a small controller gap nor this identity implies a small target cost.

3 CWOP: Optimization and Feasibility Repair

CWOP specifies a constrained serving distribution. We use a conic reference and a projected primal–dual solver (PD), applying the same feasibility checks to both.

3.1 Why Constrain Exposure Directly?

In an *urn model*, each creator is a color and each selection increases that color's count, reinforcing future selection. Classical nonlinear urns can concentrate permanently (Pemantle, 2007), motivating direct targets rather than an estimated feedback exponent. Appendix A provides this background. The serving theorem covers arbitrary history-dependent scores, including reinforcing scores, without requiring the urn model to be correct.

Why two constraints? A two-user example. Consider two equally frequent users and two eligible creators. Serving each user only a different creator gives aggregate HHI $1/2$ but zero per-user entropy: balanced global exposure can coexist with individual cocoons. Serving both users the distribution $(0.9, 0.1)$ instead gives entropy 0.325 each but aggregate HHI 0.82: an entropy floor 0.3 is met while an HHI ceiling 0.6 is violated. Equal shares satisfy both targets. Controlling information oligopoly and exposure cocoons therefore requires both global and per-user constraints.

The aggregate ceilings also address different aspects of dominance: HHI limits concentration across the distribution, whereas the share cap bounds each creator's individual allocation. Neither determines which creators appear in a user's menu, so the entropy floor supplies a separate condition on local variety.

Numerical domain and residuals. Write $(Qx)_{u,a} = q_u x_{u,a}$ and collect residuals as

$$g(d) = \big(\,\mathrm{HHI}(R(d)) - \eta,\ (R_c(d) - \kappa)_c,\ (h_0 - H(S_u(d)))_u\big).$$

The implementations use $L = 1$ and a small exploration floor, $\mathcal{X}_\epsilon = \{x_u \in \Delta(\mathcal{A}_u) : x_{u,a} \geq \epsilon\}$, with $\mathcal{D}_\epsilon = Q\mathcal{X}_\epsilon$. The PD algorithm and its convergence guarantee are in Appendix B.1.

3.2 Certifying and Repairing the Planned Policy

An averaged iterate can violate a target even when its optimization error is small. The following proposition uses a strict witness to separate exploration bias, certify the dual box, and repair the policy. Write W_ϵ^* for the constrained optimum on $\mathcal{D}_\epsilon$ and ε_T for the optimization bound in Theorem 6.

Proposition 1 (Computable error accounting)**.** *Let x^0 satisfy $g(Qx^0) \leq -\delta_0\mathbf{1}$ and have minimum admissible coordinate $\epsilon_0 > 0$. For $\epsilon \leq \epsilon_0$, write $W_c^* = \max_{d\in\mathcal{F}} W(d)$, $W_g^* = \max_{d\in\mathcal{D}} W(d)$, and $W_0 = W(Qx^0)$. Then*

$$0 \leq W_c^* - W_\epsilon^* \leq (\epsilon/\epsilon_0)(W_g^* - W_0), \tag{6}$$

$$\|\xi^*\|_1 \leq (W_g^* - W_0)/\delta_0. \tag{7}$$

Thus any B strictly above the latter bound certifies the dual-box assumption in Theorem 6. For a numerical policy $x \in \mathcal{X}_\epsilon$, let $v = \|[g(Qx)]_+\|_\infty$ and $\theta = v/(v + \delta_0)$. Then $x^{\mathrm{rep}} = (1-\theta)x + \theta x^0$ is feasible, and

$$W_c^* - W(Qx^{\mathrm{rep}}) \leq \frac{\epsilon}{\epsilon_0}(W_g^* - W_0) + \varepsilon_T + \theta\big(W(Qx) - W_0\big)$$

when x is the averaged iterate of Theorem 6.

Convexity makes the mixture feasible and linearity accounts for its utility cost (proof in Appendix B.2). The implementation checks every residual and reduces the mixing weight by bisection along the witness segment. Section 4 accounts for the randomness of serving a feasible policy.

4 Adaptive Serving and Utility Cost

We reserve slack inside the exposure targets to control finite-window sampling error. The policy can change with the observed history between windows.

4.1 Finite-Window Exposure Guarantees

The serving guarantee applies to single-item policies ($L = 1$) chosen from the history before each window. It permits reinforcing scores and any optimizer.

Theorem 1 (Exposure control under adaptive replanning)**.** *Consider J windows of N requests. Conditional on the history before a window, its policy is frozen, request types are i.i.d. with known probabilities $q_u > 0$, and items are sampled independently. Let $C = |\mathcal{C}|$, $U = |\mathcal{U}|$, $K_u = |\mathcal{C}_u| \geq 2$, and $\delta \in (0, 1)$. Define*

$$\ell_C = \log(6JC/\delta), \quad \ell_H = \log(6J/\delta), \quad \ell_U = \log(6JU/\delta), \quad n_u = \left\lfloor Nq_u - \sqrt{2Nq_u\ell_U} \right\rfloor.$$

Assume $n_u \geq 3$. For a strictly positive probability vector p, write $V(p) = \mathrm{Var}_{I\sim p}(\log p_I)$ and $A(p) = \max_i \log p_i - \min_i \log p_i$. Let $V_R = \sum_c R_c^3 - \mathrm{HHI}(R)^2$ and $A_R = \max_c R_c - \min_c R_c$. The following policy-dependent margins suffice:

$$\begin{aligned} m_c &= \sqrt{2R_c(1-R_c)\ell_C/N} + 2\ell_C/(3N), \\ m_H &= 2\sqrt{2V_R\ell_H/N} + 4A_R\ell_H/(3N) + (1+\sqrt{\ell_H})^2/N, \\ m_u &= \min\{\log K_u,\ b_u,\ t_u\}, \\ b_u &= \log(1 + (K_u-1)/n_u) + (\log n_u + 1)\sqrt{\ell_U/(2n_u)}, \\ t_u &= \sqrt{2V(S_u)\ell_U/n_u} + \frac{2A(S_u)\ell_U}{3n_u} + \frac{(K_u-1)\log(n_u+1) + \ell_U}{n_u}. \end{aligned} \tag{8}$$

Set $t_u = +\infty$ if S_u has a zero coordinate; the b_u branch remains valid. If each chosen plan satisfies $\mathrm{HHI}(R) \leq \eta - m_H$, $R_c \leq \kappa - m_c$, and $H(S_u) \geq h_0 + m_u$ for all c, u, every window and the pooled exposure meet the original targets simultaneously with probability at least $1 - \delta$.

For user-specific types and feasible certified plans, this jointly controls the two operational failure modes: global concentration and per-user exposure narrowing stay within their prescribed targets throughout J windows with probability at least $1 - \delta$. It is a finite-horizon guarantee under the stated sampling model.

Proof idea. Bernstein controls each cap using its Bernoulli variance. Expanding the HHI error as $2\langle R, \widehat{R} - R\rangle + \|\widehat{R} - R\|_2^2$ separates a scalar Bernstein term from a bounded-difference remainder. For entropy,

$$H(p) - H(\widehat{p}) = \langle \widehat{p} - p, \log p\rangle + D(\widehat{p}\|p).$$

Bernstein bounds the linear term and a method-of-types inequality bounds the KL term, giving t_u. The alternative b_u uses plug-in bias and bounded differences. The branch with the smaller bound is selected before serving. A Chernoff lower bound on each type count and a union bound complete the argument. Conditioning on each window's history allows adaptive replanning (Appendix C).

Computing a certified plan. The margins are checked on the final policy. We first solve CWOP, then search the segment to a strict witness until all inequalities in Theorem 1 hold. Up to three additional conic solves use the current margins and separate entropy floors for each type. Each candidate is checked and, if needed, mixed with the witness; we retain it only if its current score improves. This constructs certified policies without solving the nonconvex problem of optimizing policy-dependent margins globally. Appendix C.2 also gives a bound for general slates and a policy-independent single-item alternative.

4.2 Why a Safety Margin Is Necessary

Proposition 2 (Local necessity at a share cap)**.** *Fix $0 < \kappa < 1$. In a window of N independent single-item requests, suppose a creator's planned share is $p_N = \kappa - a_N$ with $a_N \geq 0$ and $a_N \to 0$. If $\sqrt{N}a_N/\sqrt{\kappa(1-\kappa)} \to z \in [0, \infty)$, then*

$$\Pr(\widehat{R}_c > \kappa) \longrightarrow 1 - \Phi(z),$$

where Φ is the standard normal cdf. Consequently a boundary plan has limiting violation probability $1/2$. For fixed $\delta_{\text{cap}} \in (0, 1/2)$, $\limsup_N \Pr(\widehat{R}_c > \kappa) \leq \delta_{\text{cap}}$ requires $\liminf_N \sqrt{N}a_N/\sqrt{\kappa(1-\kappa)} \geq \Phi^{-1}(1 - \delta_{\text{cap}})$.

Proof. The creator count is binomial (N, p_N). Its centered, variance-normalized sum converges to a standard normal because $p_N \to \kappa \in (0, 1)$. The cap threshold converges to z on this scale; lattice rounding vanishes. The final claim follows by applying the same argument to any subsequence with a smaller threshold. □

Thus the cap margin has the necessary $N^{-1/2}$ order at a fixed interior cap, up to confidence and union-bound factors. This local result does not establish optimal entropy or HHI margins. Multiple active constraints can produce a much higher probability of violating at least one target than the single-cap limit of $1/2$.

4.3 Utility Cost of Sampling Safety

Write $W_0 = W(Qx^0)$ and $W_g^* = \max_{d \in \mathcal{D}} W(d)$; as above, W_ϵ^* is the constrained optimum on the exploration-restricted domain.

Corollary 1 (Safety-margin utility budget)**.** *For a fixed window, let a feasible witness $x^0 \in \mathcal{X}_\epsilon$ have slack $s_i = -g_i(Qx^0) > 0$. Suppose a required nonnegative safety-margin vector Δ satisfies $\Delta_i \leq s_i$. Let $\vartheta = \max_i \Delta_i/s_i \leq 1$ and let $W_{\epsilon,\Delta}^*$ be the optimum with $g_i(Qx) \leq -\Delta_i$. Then*

$$0 \leq W_\epsilon^* - W_{\epsilon,\Delta}^* \leq \vartheta(W_\epsilon^* - W_0) \leq \vartheta(W_g^* - W_0).$$

The following envelope extends this fixed-target bound to certified utility.

Lemma 1 (Utility bound for policy-dependent certification)**.** *Under the assumptions of Theorem 1, with its logarithms and count bounds, define*

$$\begin{aligned}
\bar{m}_c &= \sqrt{2\kappa\ell_C/N} + 2\ell_C/(3N), \\
\bar{m}_H &= \kappa\sqrt{2\ell_H/N} + 4\kappa\ell_H/(3N) + (1 + \sqrt{\ell_H})^2/N, \\
\bar{m}_u &= \min\{\log K_u, b_u\}.
\end{aligned}$$

For every base-feasible plan, $m_i(x) \leq \bar{m}_i$. If $\bar{\vartheta} = \max_i \bar{m}_i/s_i \leq 1$, the best score $W_{\epsilon,\text{cert}}^*$ among policies passing Theorem 1 satisfies

$$W_{\epsilon,\text{cert}}^* \geq (1 - \bar{\vartheta})W_\epsilon^* + \bar{\vartheta}W_0.$$

For any feasible score $L \leq W_\epsilon^*$, replacing W_ϵ^* by L gives a computable lower bound, attained or exceeded by a witness mixture. The implementation includes this candidate whenever $\bar{\vartheta} \leq 1$. The envelope can be much looser than the final policy's own margins (proof in Appendix C.4). The ratio ϑ measures the fraction of the witness's available slack required by a fixed safety margin. Shorter windows and rarer request types increase this cost. Figure 3 reports both the bound and computed utility as N varies.

5 Experiments

We test the central claim that CWOP can jointly control the exposure-based forms of information oligopoly and information cocoons at an explicit utility cost. **Q1** measures the cost and accuracy of joint constrained optimization; **Q2** tests whether both targets hold during feedback-driven serving; **Q3** tests the same construction on a public listening-derived graph. All outcomes are reported as ratios, percentages, or normalized indices with stated denominators. Public benchmark and synthetic design settings are retained.

Table 1: Base static benchmark. Loss $L = (W_g - W)/(W_g - W_{\mathrm{ref}})$ and gap $G = (W_{\mathrm{ref}} - W)/(W_g - W_{\mathrm{ref}})$ use the greedy-to-conic target cost as one unit. These relative quantities are invariant to a common score offset; negative G does not imply feasibility.

Policy	Relative loss L	Relative gap G
Unconstrained greedy	0.0000	−1.0000
Conic reference	1.0000	0.0000
PD, repaired	1.0042	0.0042

Table 2: Matched feedback: 20 windows per trajectory, 20k requests on five synthetic graphs and 100k on three Last.fm splits. Utility U/U_{ref} and its SD use the mean no-margin conic utility on the same dataset as a fixed denominator. Failure rates are percentages of windows. Baselines are tuned on disjoint instances at matching window sizes.

Method	Synthetic ratio	Failures	Last.fm ratio	Failures
Temperature	0.781 ± 0.008	0.0%	0.466 ± 0.002	1.7%
Log-share	0.805 ± 0.007	47.0%	0.504 ± 0.005	50.0%
FairCo-style batch	0.952 ± 0.012	100.0%	1.835 ± 0.032	100.0%
Two-sided batch FW	0.889 ± 0.008	0.0%	0.625 ± 0.010	0.0%
Conic reference, no margin	1.000 ± 0.008	100.0%	1.000 ± 0.014	100.0%
Conic reference + certificate	0.959 ± 0.008	0.0%	0.829 ± 0.010	0.0%

5.1 Q1: Cost of Joint Exposure Control

Q1 separates the utility cost of joint concentration and variety targets from solver error. The synthetic benchmark has 24 request types (contexts with fixed candidates and scores), 36 creators, and 18 eligible creators per type. Appendix D specifies the graph, score, and request-probability generators. Targets are $(\eta, \kappa, h_0) = (0.055, 0.08, 0.65 \log 18)$. Clarabel solves the independent conic formulation; SCS checks the base instance. Across six instances, feasible lower and tangent-Lagrangian upper bounds differ by at most 7.1×10^{-5} times the instance's greedy-to-conic target cost. PD runs 30,000 iterations before feasibility repair.

Table 1 sets target cost to one unit; PD's gap is 0.0042. Relative gaps are 0.0030–0.0063 across five graphs and 0.0050 on a 48-type, 72-creator instance. Together with the joint feasibility checks (Appendix Table 4), this supports CWOP's cost-accounting claim: utility loss from controlling both exposure problems can be distinguished from solver error. The 512-type, 1,024-creator stress test has a larger PD gap, 0.350 of the conic improvement over the witness after 10,000 steps. We therefore use the more accurate conic implementation to evaluate serving safety.

5.2 Q2: Joint Control under Feedback

Testing the joint-control claim. Q2 evaluates both exposure targets after randomized serving, as popularity and affinity feed back into later scores. All methods receive 20 windows of 20,000 requests on each of five held-out synthetic graphs. Eligible-pair response probabilities are $p_{u,c} = \mathrm{sigmoid}(2r_{u,c} - 0.8)$. Each simulated click increases global popularity f_c and per-type affinity $A_{u,c}$, both initialized at one. Before each window the score is

$$s^j_{u,c} = p_{u,c} + \chi \log(f_c/\bar{f}) + (2\chi/3) \log(A_{u,c}/\bar{A}_u), \qquad \chi = 0.15.$$

The policy is frozen during serving. We evaluate intrinsic response probability p, while controllers optimize the feedback score s^j. Temperature, log-share, FairCo-style batch, and two-sided batch Frank–Wolfe (FW, 10,000 steps) are tuned on two disjoint graphs for highest utility with at most 5% violating windows. If none qualifies, selection orders violation rate, normalized excess, then utility (Appendix D).

Safety with utility retention. Both datasets use policy-dependent margins, separate per-type entropy floors, and checked conic reoptimization (Section 4), with $\delta = 0.05$ per 20-window trajectory. Certification uses only presampling information. On synthetic graphs, this construction retains 0.959

Table 3: Joint target control. Concentration ratios are maxima across test windows; entropy is the minimum over types and windows. Targets are ratios ≤ 1 and normalized entropy ≥ 0.65. Global and local columns give the percentages of windows violating either concentration ceiling or any entropy floor, respectively (raw tolerance 10^{-6}); they may overlap.

Data	Policy	HHI/η	Share/κ	$H/\log K$	Global	Local
Synthetic	No margin	1.013	1.083	0.606	96.0%	100.0%
Synthetic	Certified	0.974	0.946	0.732	0.0%	0.0%
Last.fm	No margin	1.012	1.057	0.581	63.3%	100.0%
Last.fm	Certified	0.978	0.949	0.804	0.0%	0.0%

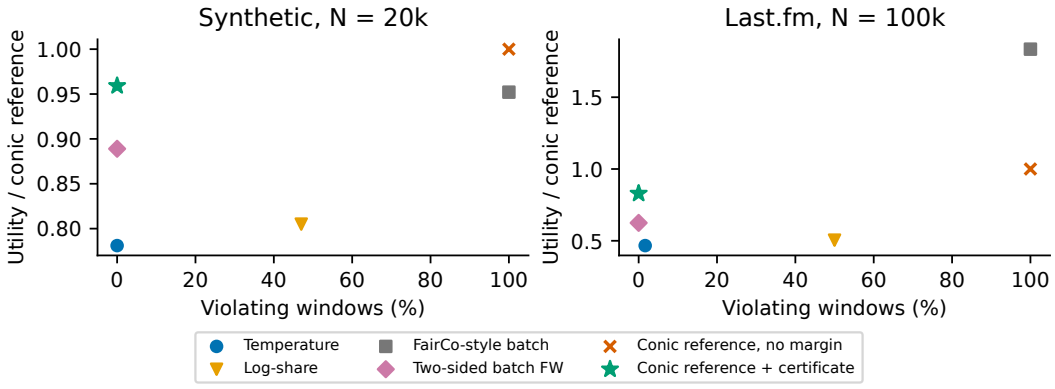

Figure 2: Held-out safety and utility at matched window sizes: synthetic (20k) and Last.fm (100k). Each baseline parameter is selected on separate training instances. Utility is divided by the dataset's no-margin conic mean, as in Table 2. The certified policy uses checked, policy-dependent reoptimization on both datasets. Points summarize dependent windows within independent graph/split trajectories.

of the no-margin conic reference's utility, with zero observed failures versus 100% violating windows without margins. At the same window size, the best selected baseline with zero failures is FW at ratio 0.889 (Table 2, Figure 2).

Evidence for both exposure objectives. Table 3 separates the two outcomes. Across every certified window, both concentration ratios stay below one and minimum normalized creator entropy stays above 0.65: 0.732 on synthetic graphs and 0.804 on Last.fm. Without margins, concentration fails in 96.0% and 63.3% of windows, respectively, and the entropy floor fails in every window. This directly supports the thesis: certified CWOP controls creator dominance and per-user narrowing together, as operationalized by the chosen targets. A threshold violation alone does not establish severe oligopoly or a complete cocoon.

Reoptimization improves utility over the certified witness-segment ratio 0.904 (Appendix Table 6), strengthening the empirical safety–utility tradeoff. These comparisons follow separate feedback trajectories; Lemma 1 instead bounds the score at a fixed history, and the implementation checks that bound whenever available.

Sample size and certification cost. Figure 3 quantifies when joint control also retains useful utility. Relative to $N_0 = 20$k requests, the policy-independent certificate admits the uniform witness at $N/N_0 = 0.617$, using nearly all its entropy slack; its mixing coefficient falls to 0.409 at $N/N_0 = 5$. Policy-dependent certification admits the witness at $N/N_0 = 0.254$. With witness-to-conic score gain normalized to one, the construction achieves gain 0.061 there and 0.848 at $N = N_0$. Thus larger windows can reduce the cost of protecting both targets. These are attainable constructions, not certified global optima.

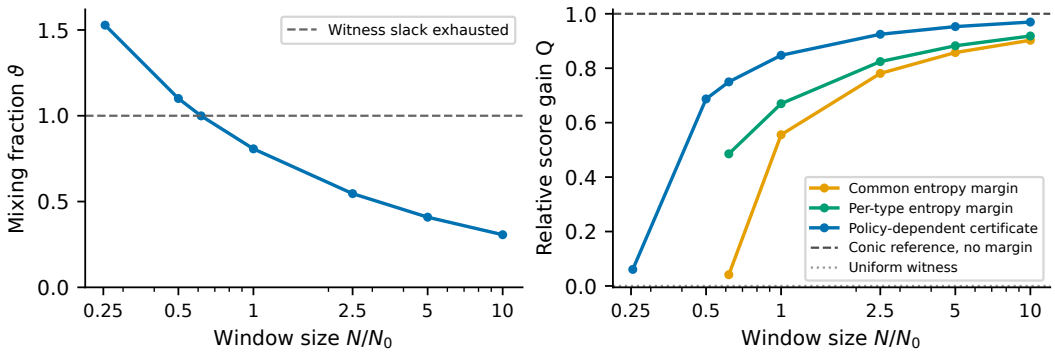

Figure 3: Base synthetic graph ($N_0 = 20$k). Left: witness mixing fraction for policy-independent per-type margins. Right: relative score gain Q (Appendix D.1); the uniform witness has $Q = 0$ and the no-margin conic reference $Q = 1$. Missing points mean this construction lacks a certificate, not global infeasibility.

Testing sampling protection. Each saved window on the first graph of each dataset is independently resampled 2,000 times. Certified plans have a 0.0025% cap-failure rate in synthetic replications and zero in Last.fm replications; all no-margin tests fail. Pointwise one-sided 95% failure upper bounds are 0.24% for the one-failure window and 0.15% for each zero-failure window. These tests assess conditional risk; Theorem 1 gives the simultaneous adaptive guarantee. Scaling each candidate's own margins along a witness segment gives first zero-failure grid points at $c = 0.8$ (synthetic) and $c = 0.6$ (Last.fm; Appendix Figure 9). Only the full margin carries the certificate.

5.3 Q3: Joint Control on a Listening-Derived Graph

Q3 tests joint control on listening-derived candidates. A rank-16 SVD of public HetRec Last.fm counts (Cantador et al., 2011) supplies preference proxies and 64 candidates per user (128 users, 256 artists). Held-out Recall@20 is 0.539–0.583, versus 0.286–0.327 for popularity. PD's static gaps are 1.06%–1.24% of the corresponding conic utility across three test splits (Appendix Table 10).

Feedback uses the proxies clipped to $[0, 1]$ as Bernoulli response probabilities. All methods use 100,000 requests per window, with baseline tuning on two separate SVD splits. Certification achieves both targets (Table 3), with utility ratio 0.829 to the no-margin conic reference, a 17.1% loss. The uniform-witness threshold is 0.337 of the 100k window on each split: more types leave fewer observations per entropy floor. Different candidates and preference proxies extend Q2's joint control and certification to listening-derived exposure (Appendix D.6). Consumption and viewpoint diversity remain separate outcomes.

An anonymized audit motivates concentration control: matched exposure HHI is approximately eight times audience-stock HHI. Matching fixes the creator universe, but the ratio compares a serving flow with an accumulated audience stock. It therefore motivates exposure control without identifying a causal allocation effect. Coverage is limited, and the exposure-to-growth interval includes parity (Appendix E). Urn diagnostics isolate concentration and narrowing mechanisms (Appendix F).

6 Conclusion

CWOP addresses exposure-based information oligopoly and information cocoons through joint concentration ceilings and per-user entropy floors. Optimization certificates separate target cost from solver error; safety margins provide probabilistic finite-window control under adaptive feedback. Both targets hold throughout the certified synthetic and listening-derived feedback runs, at relative utility costs of 4.1% and 17.1%.

THE USE OF LARGE LANGUAGE MODEL

We used large language models (LLMs) only to aid or polish the writing; no content generation, ideation, or data analysis was performed by LLMs.

ETHICS STATEMENT

This study combines controlled simulations, a public artist-listening benchmark, and an observational production audit; it does not report a deployed CWOP intervention. The audit uses pseudonymized user and creator identifiers and reports aggregate results; raw production records are not distributed. The Last.fm benchmark uses numeric listening relations from the HetRec research release, not names, tags, or social links. The data owner has authorized research use of the platform exports and publication of aggregate results. Raw user, creator, and brand identifiers are not included in the manuscript or submission package. The proposed policy uses eligibility, relevance, and competition signals rather than sensitive attributes. Absolute platform levels, sample counts, proprietary labels, and source identifiers are omitted from the release. Any production evaluation requires appropriate data-access permissions and institutional review, privacy safeguards, fairness monitoring, and an interference-aware design. Appendix G describes the data handling and deployment considerations.

REPRODUCIBILITY STATEMENT

Sections 2–4 specify the model, controller, and guarantees; Appendices A, B, and C provide the assumptions and proofs. Appendix D documents candidate construction, utility models, tuning and test splits, comparator budgets, solver certificates, and runtime measurement. Appendices E and F describe the audit and urn simulations, respectively.

The submission source package contains the manuscript, relative-result tables, and vector figures. Experimental settings and public dataset citations are retained; table captions and Appendix D.1 define every reporting scale. The authors retain experiment runners, tests, fixed seeds, full-precision outputs, and checksum records separately. These computational records support checking the relative transformations but are not bundled with the anonymized manuscript. Raw Last.fm inputs are retrieved from the official research release, not redistributed; the derived arrays retain its noncommercial-use restriction. Confidential production results require governed access to the original records and audit pipeline. Only coarsened relative platform summaries are released; source-specific metadata and fine-grained private distributions are withheld. This limits independent audit reproduction and does not establish causal effects.

REFERENCES

Md Sanzeed Anwar, Grant Schoenebeck, and Paramveer S. Dhillon. Filter bubble or homogenization? disentangling the long-term effects of recommendations on user consumption patterns. In *Proceedings of the ACM Web Conference 2024*, pp. 123–134, 2024. doi: 10.1145/3589334.3645497.

Koenraad M. R. Audenaert. A sharp continuity estimate for the von Neumann entropy. *Journal of Physics A: Mathematical and Theoretical*, 40(28):8127–8136, 2007. doi: 10.1088/1751-8113/40/28/S18.

Asia J. Biega, Krishna P. Gummadi, and Gerhard Weikum. Equity of attention: Amortizing individual fairness in rankings. In *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pp. 405–414, 2018. doi: 10.1145/3209978.3210063. URL https://asiabiega.github.io/papers/equity_of_attention_sigir2018.pdf.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013. doi: 10.1093/acprof:oso/9780199535255.001.0001. URL https://academic.oup.com/book/26549.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. doi: 10.1017/CBO9780511804441. URL https://web.stanford.edu/~boyd/cvxbook/.

Iván Cantador, Peter Brusilovsky, and Tsvi Kuflik. 2nd workshop on information heterogeneity and fusion in recommender systems (HetRec 2011). In *Proceedings of the 5th ACM Conference on Recommender Systems*. ACM, 2011. URL https://grouplens.org/datasets/hetrec-2011/.

Allison J. B. Chaney, Brandon M. Stewart, and Barbara E. Engelhardt. How algorithmic confounding in recommendation systems increases homogeneity and decreases utility. In *Proceedings of the 12th ACM Conference on Recommender Systems*, pp. 224–232, 2018. doi: 10.1145/3240323.3240370. URL https://arxiv.org/abs/1710.11214.

Qinyi Chen, Jason Cheuk Nam Liang, Negin Golrezaei, and Djallel Bouneffouf. Interpolating item and user fairness in multi-sided recommendations. In *Advances in Neural Information Processing Systems 37*, 2024. doi: 10.52202/079017-1589. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/59d4e18a60490b9ed9913f3be2b14839-Abstract-Conference.html.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley, 2 edition, 2006. doi: 10.1002/047174882X. URL https://onlinelibrary.wiley.com/doi/book/10.1002/047174882X.

Paul Covington, Jay Adams, and Emre Sargin. Deep neural networks for YouTube recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pp. 191–198, 2016.

Virginie Do, Sam Corbett-Davies, Jamal Atif, and Nicolas Usunier. Two-sided fairness in rankings via lorenz dominance. In *Advances in Neural Information Processing Systems*, volume 34, pp. 8596–8608, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/48259990138bc03361556fb3f94c5d45-Abstract.html.

Rick Durrett. *Probability: Theory and Examples*. Cambridge University Press, 5 edition, 2019. doi: 10.1017/9781108591034. URL https://www.cambridge.org/core/books/probability/DD9A1907F810BB14CCFF022CDFC5677A.

Itay Eilat and Nir Rosenfeld. Performative recommendation: Diversifying content via strategic incentives. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 9082–9103, 2023. URL https://proceedings.mlr.press/v202/eilat23a.html.

Sruthi Gorantla, Amit Deshpande, and Anand Louis. Sampling ex-post group-fair rankings. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pp. 409–417, 2023. doi: 10.24963/ijcai.2023/46. URL https://www.ijcai.org/proceedings/2023/46.

Meena Jagadeesan, Nikhil Garg, and Jacob Steinhardt. Supply-side equilibria in recommender systems. In *Advances in Neural Information Processing Systems 36*, 2023. doi: 10.52202/075280-0642. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/2f1486343c2c942a617e4f5bb0cc64c8-Abstract-Conference.html.

Ray Jiang, Silvia Chiappa, Tor Lattimore, András György, and Pushmeet Kohli. Degenerate feedback loops in recommender systems. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 383–390, 2019.

Haruka Kiyohara, Fan Yao, and Sarah Dean. Policy design for two-sided platforms with participation dynamics. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 30966–30985. PMLR, 2025. URL https://proceedings.mlr.press/v267/kiyohara25a.html.

Tao Liu, Ruida Zhou, Dileep Kalathil, Panganamala R. Kumar, and Chao Tian. Learning policies with zero or bounded constraint violation for constrained MDPs. In *Advances in Neural Information Processing Systems*, volume 34, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/8ec2ba5e96ec1c050bc631abda80f269-Abstract.html.

Masoud Mansoury, Himan Abdollahpouri, Mykola Pechenizkiy, Bamshad Mobasher, and Robin Burke. Feedback loop and bias amplification in recommender systems. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pp. 2145–2148, 2020.

Martin Mladenov, Elliot Creager, Omer Ben-Porat, Kevin Swersky, Richard Zemel, and Craig Boutilier. Optimizing long-term social welfare in recommender systems: A constrained matching approach. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 6987–6998, 2020. URL https://proceedings.mlr.press/v119/mladenov20a.html.

Marco Morik, Ashudeep Singh, Jessica Hong, and Thorsten Joachims. Controlling fairness and bias in dynamic learning-to-rank. In *Proceedings of the 43rd International ACM SIGIR Conference on Research & Development in Information Retrieval*, pp. 429–438, 2020. doi: 10.1145/3397271.3401100. URL https://www.cs.cornell.edu/people/tj/publications/morik_etal_20a.pdf.

Angelia Nedić and Asuman Ozdaglar. Approximate primal solutions and rate analysis for dual subgradient methods. *SIAM Journal on Optimization*, 19(4):1757–1780, 2009. doi: 10.1137/070708111.

Arkadi Nemirovski and Alexander Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4):969–996, 2006. doi: 10.1137/050622328.

Arkadi Nemirovski, Anatoli Juditsky, Guanghui Lan, and Alexander Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization*, 19(4):1574–1609, 2009. doi: 10.1137/070704277.

Liam Paninski. Estimation of entropy and mutual information. *Neural Computation*, 15(6):1191–1253, 2003. doi: 10.1162/089976603321780272. URL https://sites.stat.columbia.edu/liam/research/pubs/info_est-nc.pdf.

Robin Pemantle. A survey of random processes with reinforcement. *Probability Surveys*, 4:1–79, 2007. doi: 10.1214/07-PS094. URL https://www2.math.upenn.edu/~pemantle/papers/reinforce.pdf.

Juan Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative prediction. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 7599–7609, 2020. URL https://proceedings.mlr.press/v119/perdomo20a.html.

Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. Recommendations as treatments: Debiasing learning and evaluation. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 1670–1679, 2016. URL https://proceedings.mlr.press/v48/schnabel16.html.

Ashudeep Singh and Thorsten Joachims. Fairness of exposure in rankings. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2219–2228, 2018. URL https://www.cs.cornell.edu/~tj/publications/singh_joachims_18a.pdf.

Ashudeep Singh and Thorsten Joachims. Policy learning for fairness in ranking. In *Advances in Neural Information Processing Systems 32*, pp. 5426–5436, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/9e82757e9a1c12cb710ad680db11f6f1-Abstract.html.

Ashudeep Singh, David Kempe, and Thorsten Joachims. Fairness in ranking under uncertainty. In *Advances in Neural Information Processing Systems*, volume 34, pp. 11896–11908, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/63c3ddcc7b23daa1e42dc41f9a44a873-Abstract.html.

Harald Steck. Calibrated recommendations. In *Proceedings of the 12th ACM Conference on Recommender Systems*, pp. 154–162, 2018. doi: 10.1145/3240323.3240372.

Nicholas Sukiennik, Nian Li, and Chen Gao. Uncovering the deep filter bubble: Narrow exposure in short-video recommendation. In *Proceedings of the ACM Web Conference 2024*, pp. 4727–4735, 2024. doi: 10.1145/3589334.3648159.

U.S. Department of Justice, Antitrust Division. Herfindahl–Hirschman Index, 2024. URL https://www.justice.gov/atr/herfindahl-hirschman-index. Updated January 17, 2024; accessed September 25, 2026.

Nicolas Usunier, Virginie Do, and Elvis Dohmatob. Fast online ranking with fairness of exposure. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, 2022. doi: 10.1145/3531146.3534633. URL https://facctconference.org/static/pdfs_2022/facct22-3534633.pdf.

Lequn Wang and Thorsten Joachims. User fairness, item fairness, and diversity for rankings in two-sided markets. In *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval*, 2021. doi: 10.1145/3471158.3472260. URL https://www.cs.cornell.edu/people/tj/publications/wang_joachims_21a.pdf.

Tsachy Weissman, Erik Ordentlich, Gadiel Seroussi, Sergio Verdú, and Marcelo J. Weinberger. Inequalities for the L1 deviation of the empirical distribution. Technical Report HPL-2003-97R1, HP Laboratories, 2003.

Fan Yao, Chuanhao Li, Denis Nekipelov, Hongning Wang, and Haifeng Xu. How bad is top-k recommendation under competing content creators? In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 39674–39701. PMLR, 2023. URL https://proceedings.mlr.press/v202/yao23b.html.

Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th International Conference on Machine Learning*, pp. 928–936, 2003.

A Reinforcement Background

These classical results motivate exposure constraints and interpret the diagnostic simulations. They are separate from the finite-window serving guarantee. For a fixed catalog of $n \geq 2$ creators, let $f_t(c) > 0$ be reinforced counts, $F_t = \sum_c f_t(c)$, and $\rho_t(c) = f_t(c)/F_t$.

Theorem 2 (Classical concentration trichotomy (Pemantle, 2007))**.** *With unit initial counts and selection proportional to $f_t(c)^\beta$, $\beta \geq 0$: (i) $\rho_t \to (1/n, \ldots, 1/n)$ almost surely for $\beta < 1$; (ii) $\rho_t \to \rho_\infty \sim \mathrm{Dirichlet}(\mathbf{1})$ for $\beta = 1$, with $\mathbb{E}\,\mathrm{HHI}(\rho_\infty) = 2/(n+1)$; and (iii) for $\beta > 1$ one creator is selected at all but finitely many steps almost surely, so $\mathrm{HHI}(\rho_t) \to 1$.*

Proof outline. For $0 \leq \beta < 1$, the mean ODE is $\dot{\rho}_c = \rho_c^\beta/S - \rho_c$, $S = \sum_c \rho_c^\beta$. For $M = \max_c \rho_c$, $m = \min_c \rho_c$, $D^+ \log(M/m) \leq (M^{\beta-1} - m^{\beta-1})/S < 0$ away from uniformity. Boundary faces repel interior trajectories. The sublinear-urn stochastic-approximation result gives almost-sure convergence (Pemantle, 2007). At $\beta = 1$, the Pólya limit is Dirichlet and $\mathbb{E}\rho_\infty(c)^2 = 2/[n(n+1)]$.

For $\beta > 1$, Rubin's embedding assigns independent waiting times $E_{c,k}/k^\beta$, $E_{c,k} \overset{\text{iid}}{\sim} \mathrm{Exp}(1)$. Each explosion time $T_c = \sum_k E_{c,k}/k^\beta$ is finite almost surely. The times are independent, continuous, and hence distinct. Before the smallest explosion time, every other clock fires only finitely often. This gives (iii). □

Remark 1. The linear limit is random non-uniformity, rather than a few-winner oligopoly. A few-winner configuration in a finite simulation is a transient; the superlinear asymptotic limit is monopoly. The expected factor $2n/(n+1)$ is not a pathwise HHI ratio.

Theorem 3 (Per-user exposure narrowing)**.** *For a user's fixed $K_u \geq 2$ candidates and unit initial counts, the same urn has $H(\sigma_{u,t})$ converging almost surely to a random value below $\log K_u$ at $\beta = 1$, and to zero at $\beta > 1$.*

Proof. Apply Theorem 2 to this user's served counts and use continuity of entropy. The continuous Dirichlet law gives probability zero to exact uniformity. For positive request weights, $\rho = \sum_u q_u \sigma_u$ also implies $\mathrm{HHI}(\rho) \leq \sum_u q_u \,\mathrm{HHI}(\sigma_u)$ by Jensen. If global exposure is supported on S, then each positive-weight user's exposure is supported on S and $H(\sigma_u) \leq \log|S|$. High HHI alone does not restrict support. □

Theorem 4 (Bounded count-only shaping in the superlinear regime)**.** *For $\beta > 1$, a time-homogeneous bonus $b_t(c) = g_c(f_t(c))$ with $|g_c(k)| \leq B_{\mathrm{sh}}$ preserves monopoly under Boltzmann temperature $\tau > 0$.*

Proof. The weights are $\widetilde{w}_c(k) = k^\beta e^{g_c(k)/\tau}$ and $\sum_k 1/\widetilde{w}_c(k) \leq e^{B_{\mathrm{sh}}/\tau} \sum_k k^{-\beta} < \infty$. The creator-specific exponential embedding above applies. □

Count-only dependence matters. At $\beta = 1$, the bounded state-dependent two-creator bonus $b_i(\rho) = -\gamma(\rho_i - 1/2)$ creates mean reversion; boundedness alone is therefore insufficient for this conclusion.

Proposition 3 (Marginal cumulative-entropy bonus)**.** *For $b_t(c) = \lambda\, \partial H(\rho_t)/\partial f_t(c)$, $b_t(c) - b_t(c') = -(\lambda/F_t)\log[f_t(c)/f_t(c')]$.*

Proof. Differentiate $H(f/F) = \log F - F^{-1}\sum_c f_c \log f_c$. The creator-independent terms cancel. Positive unit counts bound the difference by $O(\log F_t/F_t) \to 0$. This derivative of cumulative-count entropy differs from ordinary policy entropy regularization. □

Theorem 5 (Log-share threshold)**.** *For $\beta \geq 0$, $\lambda \geq 0$, $\tau > 0$, and positive initial counts, $b_t(c) = -\lambda \log \rho_t(c)$ gives exponent $\beta_{\mathrm{eff}} = \beta - \lambda/\tau$. Balance holds for $\beta_{\mathrm{eff}} < 1$, a random non-uniform limit for $\beta_{\mathrm{eff}} = 1$, and monopoly for $\beta_{\mathrm{eff}} > 1$. Thus $\lambda^\star = \tau(\beta - 1)$ is the boundary.*

Proof. The factor $F_t^{\lambda/\tau}$ cancels in $f_t(c)^\beta \rho_t(c)^{-\lambda/\tau}$. Apply Theorem 2 for nonnegative effective exponents and Lemma 2 below otherwise. With non-unit initial counts, the linear Dirichlet parameters are those counts; the other regimes are unchanged. □

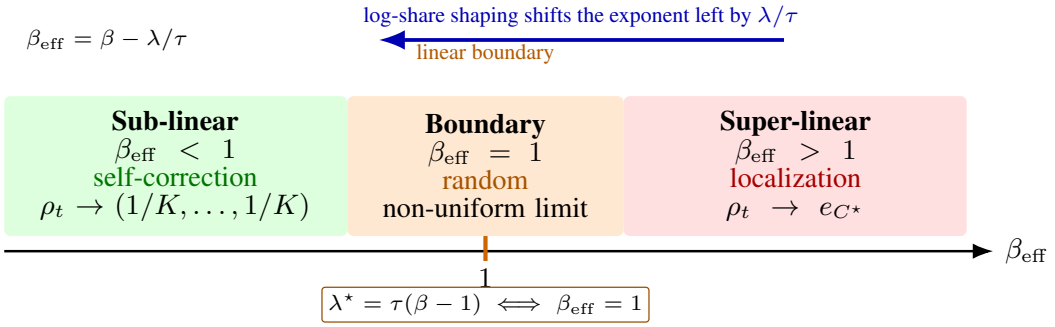

Figure 4: Three feedback regimes and the log-share threshold (schematic). Within the single-state power-law model, log-share shaping shifts the exponent exactly by λ/τ. The limit at the linear boundary is non-uniform but need not be an oligopoly.

Lemma 2 (Nonpositive reinforcement)**.** *For finitely many creators, positive initial counts and weights f_i^ν, $\nu \le 0$, the normalized counts converge almost surely to uniformity.*

Proof. Write $\pi_i = \rho_i^\nu / \sum_k \rho_k^\nu$, $u = (1/n, \ldots, 1/n)$, and $V = \|\rho - u\|_2^2$. Opposite ordering yields $\sum_{i,k}(\rho_i - \rho_k)(\rho_i^\nu - \rho_k^\nu) \le 0$, hence $\langle \rho, \pi \rangle \le 1/n$. For the next draw I, $\rho' = (F\rho + e_I)/(F+1)$, so

$$\mathbb{E}[V' \mid \rho] \le (F/(F+1))^2 V + (1 - 1/n)/(F+1)^2.$$

Adding the tail sum of the summable last term makes a nonnegative supermartingale. It converges, and the sum of its expected decreases is finite. Thus V converges and $\sum_t [1-(F_t/(F_t+1))^2] V_t < \infty$ almost surely. Since the coefficients have divergent sum, the limit of V is zero. □

Two-state scope. If popularity f_c and exposure e_c are distinct, cancellation is no longer an identity. Aggregate conversion $\sum_c f_c / \sum_c e_c \to q_{\text{conv}}$ does not establish creator-wise locking. Appendix F reports the coupled simulations as numerical diagnostics.

B Controller Guarantees and Proofs

B.1 PD: Projected Primal–Dual Optimization

For an executable controller, collect the convex constraint residuals in

$$g(d) := \big(\text{HHI}(R(d)) - \eta,\ (R_c(d) - \kappa)_c,\ (h_0 - H(S_u(d)))_u\big). \tag{9}$$

Write $(Qx)_{u,a} = q_u x_{u,a}$ and $\mathcal{X}_\epsilon = \{x_u \in \Delta(\mathcal{A}_u) : x_{u,a} \ge \epsilon\}$, where $0 < \epsilon < 1/\max_u |\mathcal{A}_u|$ is small enough to preserve strict feasibility. Let $\mathcal{D}_\epsilon = Q\mathcal{X}_\epsilon$ and $\mathcal{L}_x(x, \xi) = W(Qx) - \xi^\top g(Qx)$, with $\xi \in \Xi_B = [0, B]^{M_g}$, $M_g = 1 + |\mathcal{C}| + |\mathcal{U}|$. The analyzed and implemented Euclidean updates use conditional probabilities x:

$$\begin{aligned} x_{t+1} &= \Pi_{\mathcal{X}_\epsilon}[x_t + \gamma_t^x \nabla_x \mathcal{L}_x(x_t, \xi_t)], \\ \xi_{t+1} &= \Pi_{\Xi_B}[\xi_t + \gamma_t^\xi g(Qx_t)]. \end{aligned} \tag{10}$$

The returned occupancy is $\bar{d}_T = Q(T^{-1} \sum_{t=1}^T x_t)$. The factors q_u enter welfare and aggregate gradients, but not the per-type entropy gradient. Euclidean projection in d would be a different algorithm.

Optimization guarantee. Under bounded subgradients and steps proportional to $t^{-1/2}$, Theorem 6 gives saddle gap $\varepsilon_T = O(T^{-1/2})$. If the dual box contains an optimal multiplier with coordinate slack $r_\xi > 0$, then $W_\epsilon^* - W(\bar{d}_T) \le \varepsilon_T$ and $\|[g(\bar{d}_T)]_+\|_1 \le \varepsilon_T / r_\xi$. Here W_ϵ^* is the constrained optimum on $\mathcal{D}_\epsilon$. The full statement, constants, and independent-serving tracking bounds are in Appendix B.5. The worst-case constant is vacuous here: the base bound is 1.2×10^7 times the

Algorithm 1 PD: a projected primal–dual solver for CWOP

Require: Request probabilities q_u, admissible slates $\mathcal{A}_u$, exposure incidences $z_{u,a,c}$, welfare $r_{u,a}$, targets (η, κ, h_0), exploration floor ϵ, dual bound B
Ensure: Averaged occupancy $\bar{d}_T$ and conditional serving policy
Define $\mathcal{X}_\epsilon$, $d = Qx$ and residual vector $g(d) = (\mathrm{HHI}(R(d)) - \eta, (R_c(d) - \kappa)_c, (h_0 - H(S_u(d)))_u)$
Initialize $x_1 \in \mathcal{X}_\epsilon$, dual vector $\xi_1 \leftarrow 0$, and running average $\bar{x}_0 \leftarrow 0$
for $t = 1, \ldots, T$ **do**
    Compute the request-slate marginal score $G_t \leftarrow \nabla_x [W(Qx_t) - \xi_t^\top g(Qx_t)]$
    $x_{t+1} \leftarrow \Pi_{\mathcal{X}_\epsilon}(x_t + \gamma_t^x G_t)$ ▷ primal ascent
    $\xi_{t+1} \leftarrow \Pi_{[0,B]^{M_g}}(\xi_t + \gamma_t^\xi g(Qx_t))$ ▷ dual descent
    $\bar{x}_t \leftarrow (1 - 1/t)\bar{x}_{t-1} + x_t/t$
end for
Set $\bar{d}_T \leftarrow Q\bar{x}_T$ and obtain x^{serve} by repairing $\bar{x}_T$ (Proposition 1)
for all arriving requests of type u **do**
    Sample an admissible slate $a \sim x_u^{\mathrm{serve}}$
    Log realized position-weighted exposure for tracking
end for
Return $\bar{d}_T$, x^{serve}, and the empirical constraint report
The controller is request-aware: entropy dual effects depend on u and slate contents. Theorem 6 bounds optimization and sampling error.

greedy-to-conic target cost at $T = 30{,}000$, whereas the measured feasible gap is 0.0042 on that scale. We therefore assess accuracy with independent constrained solves (Appendix B.6).

Dual adjustments depend on the request (Appendix B.4). We implement the length-one special case of (10): Section 5.1 evaluates optimization, utility and runtime, and Section 5.2 tests adaptive replanning. General production-scale slate optimization is not claimed.

B.2 Exploration, Dual Certificates, and Feasibility Repair

We prove Proposition 1.

Proof. Mix an unrestricted constrained optimizer with x^0 using weight ϵ/ϵ_0. Every admissible coordinate is at least ϵ, and convexity preserves feasibility. Linearity gives (6). Slater's condition and compactness imply strong duality. For any optimal multiplier, $W_\epsilon^* = \max_{x \in \mathcal{X}_\epsilon} \mathcal{L}_x(x, \xi^*) \geq W_0 + \delta_0 \|\xi^*\|_1$, giving (7). Finally, convexity gives $g_j(Qx^{\mathrm{rep}}) \leq (1-\theta)v - \theta\delta_0 = 0$ for each constraint. The welfare identity $W(Qx^{\mathrm{rep}}) = (1-\theta)W(Qx) + \theta W_0$, combined with the restricted optimization bound and (6), proves the last claim. In computation we add a small positive safety cushion. Bisection along the same segment can reduce the sufficient mixing weight, provided every residual is checked directly; this is not an optimal projection onto the full feasible set. □

B.3 Convexity of the CWOP Program and Welfare-Gap Accounting

Convexity. The request-occupancy polytope $\mathcal{D}$ is a product of scaled simplices and hence is compact and convex. Every q_u is fixed and positive, making $R(d)$ and $S_u(d)$ affine in d. Therefore $\mathrm{HHI}(R(d)) = \|R(d)\|_2^2$ and $\|R(d)\|_\infty$ are convex, while $H(S_u(d))$ is concave. The HHI and cap sub-level sets and entropy super-level sets are consequently convex. The objective W is linear, so (5) is a convex optimization problem. □

Existence and accounting. By Assumption 1, $\mathcal{F}$ is nonempty. As a closed subset of compact $\mathcal{D}$, it is compact. Continuity of W gives an optimizer by the extreme-value theorem. Since $\mathcal{F} \subseteq \mathcal{D}$,

$$W(\hat{d}) \leq W_g^*, \qquad \Lambda(\mathcal{F}) := W_g^* - W(\hat{d}) \geq 0.$$

This definition gives the claimed identity; it does not supply a numerical welfare bound without specifying and estimating W. □

B.4 Barrier Limit and Dual Adjustments

Barrier characterization. On the strictly feasible exploration-restricted domain, the logarithmic barrier has objective gap at most $M_g\mu$. Its induced dual multipliers remain bounded and every accumulation point is primal–dual optimal. The derivation also identifies how each constraint changes the request-level score.

Proof. Work on $\mathcal{X}_\epsilon$ and remove identically zero exposure coordinates before differentiating entropy. Write $f(x) = W(Qx)$ and $g_x(x) = g(Qx)$. A barrier maximizer maximizes $f(x) + \mu \sum_j \log(-g_{x,j}(x))$ over $\mathcal{X}_\epsilon$ with $g_x < 0$. Existence follows from compactness and strict feasibility: the barrier tends to minus infinity at a constraint boundary. Setting $\xi_{\mu,j} = \mu/(-g_{x,j}(x_\mu))$, first-order concavity gives $\sup_{x\in\mathcal{X}_\epsilon} \mathcal{L}_x(x, \xi_\mu) = \mathcal{L}_x(x_\mu, \xi_\mu) = f(x_\mu) + M_g\mu$. Thus $0 \le W^*_\epsilon - f(x_\mu) \le M_g\mu$. A strict witness bounds $\delta_0\|\xi_\mu\|_1 \le f(x_\mu) + M_g\mu - f(x^0)$, so multipliers remain bounded. Every accumulation point is primal-dual optimal; convergence of the entire pair requires uniqueness, not merely strict complementarity. In x coordinates the HHI and cap gradient terms are $2q_u \sum_c R_c z_{u,a,c}$ and $q_u z_{u,a,c}$, whereas the entropy gradient is $\sum_c z_{u,a,c}(-\log S_{u,c} - 1)$. Consequently entropy adjustments depend on the request and slate; they are not generally a single creator-only price. The barrier limit concerns the restricted problem; the unrestricted comparator is covered by Proposition 1. □

B.5 Finite-Time Optimization and Serving Tracking

Theorem 6 (Finite-time optimization, constraint violation, and serving tracking)**.** *Assume the strict witness of Proposition 1 has minimum admissible coordinate $\epsilon_0 \ge \epsilon$, so Slater's condition holds on $\mathcal{X}_\epsilon$. On $\mathcal{X}_\epsilon \times \Xi_B$, suppose the primal and dual subgradient norms are bounded by G_x, G_ξ and the Euclidean diameters are D_x, D_ξ. Projected saddle updates with steps $\gamma^x_t = \gamma_x/\sqrt{t}$ and $\gamma^\xi_t = \gamma_\xi/\sqrt{t}$ have ergodic averages $(\bar{x}_T, \bar{\xi}_T)$, with $\bar{d}_T = Q\bar{x}_T$, whose saddle gap is at most*

$$\varepsilon_T \le \frac{1}{\sqrt{T}} \left(\frac{D_x^2}{2\gamma_x} + \gamma_x G_x^2 + \frac{D_\xi^2}{2\gamma_\xi} + \gamma_\xi G_\xi^2 \right). \tag{11}$$

If an optimal multiplier ξ^* satisfies $\xi^* + r_\xi \mathbf{1} \le B\mathbf{1}$ for some $r_\xi > 0$, then

$$W^*_\epsilon - W(\bar{d}_T) \le \varepsilon_T, \qquad \|[g(\bar{d}_T)]_+\|_1 \le \varepsilon_T/r_\xi, \tag{12}$$

where W^*_ϵ is the constrained optimum on $\mathcal{D}_\epsilon$.

The optimization comparator is restricted to $\mathcal{D}_\epsilon$, not all $\mathcal{F}$. If N independent request-slate pairs are sampled from the conditional policy represented by $\bar{d}_T$, then with probability at least $1 - \delta$,

$$\max_c |\widehat{R}_c - R_c(\bar{d}_T)| \le \sqrt{\frac{\log(2|\mathcal{C}|/\delta)}{2N}}. \tag{13}$$

Let $C_\Sigma = \sum_u |\mathcal{C}_u|$. Separately, conditional on the type counts $N_u > 0$, with probability at least $1-\delta$ a simultaneous coordinate bound gives $\|\widehat{S}_u - S_u(\bar{d}_T)\|_1 \le |\mathcal{C}_u|\sqrt{\log(2C_\Sigma/\delta)/(2N_u)}$. Consequently maximum-share and HHI errors vanish at $O(N^{-1/2})$. The Fannes-Audenaert bound implies that every entropy error vanishes as $N_u \to \infty$ (Audenaert, 2007).

Proof. The saddle function $\mathcal{L}_x(x, \xi) = W(Qx) - \xi^\top g(Qx)$ is concave in x and affine in ξ. The standard projected-gradient inequality (Zinkevich, 2003; Nemirovski et al., 2009; Nedić & Ozdaglar, 2009), summed for primal ascent and dual descent with steps proportional to $t^{-1/2}$, gives for every $(x, \xi) \in \mathcal{X}_\epsilon \times \Xi_B$

$$\frac{1}{T} \sum_{t=1}^{T} \big(\mathcal{L}_x(x, \xi_t) - \mathcal{L}_x(x_t, \xi) \big) \le \varepsilon_T,$$

with ε_T as stated. Concavity, convexity, and Jensen's inequality then yield

$$\max_{x\in\mathcal{X}_\epsilon} \mathcal{L}_x(x, \bar{\xi}_T) - \min_{\xi\in\Xi_B} \mathcal{L}_x(\bar{x}_T, \xi) \le \varepsilon_T.$$

Let $d_\epsilon^\star$ be a constrained optimum. Since it is feasible, $\mathcal{L}_x(Q^{-1}d_\epsilon^\star, \bar{\xi}_T) \geq W_\epsilon^*$. Selecting $\xi = 0$ in the gap bound gives $W_\epsilon^* - W(\bar{d}_T) \leq \varepsilon_T$.

For constraint violation, let $v_j = \mathbf{1}\{g_j(\bar{d}_T) > 0\}$ and select $\xi = \xi^* + r_\xi v$, which lies in Ξ_B by assumption. Strong duality gives $W_\epsilon^* - W(\bar{d}_T) + (\xi^*)^\top g(\bar{d}_T) \geq 0$. Substitution in the saddle-gap inequality leaves $r_\xi \sum_j [g_j(\bar{d}_T)]_+ \leq \varepsilon_T$, proving the second bound.

For serving, the exposure credited to creator c on each independent request is a random variable in $[0, 1]$ with mean $R_c(\bar{d}_T)$. Hoeffding's inequality and a union bound over creators give the stated aggregate coordinate bound. Conditional on N_u, applying the same argument to every creator coordinate for type u and taking a union bound over types gives the ℓ_1 bound for $\widehat{S}_u$. The maximum-share error is at most the maximum coordinate error, and

$$|\operatorname{HHI}(\widehat{R}) - \operatorname{HHI}(R)| = \left|\sum_c (\widehat{R}_c - R_c)(\widehat{R}_c + R_c)\right| \leq 2\|\widehat{R} - R\|_\infty.$$

Finally, with $\Delta_u = \frac{1}{2}\|\widehat{S}_u - S_u\|_1$, the Fannes-Audenaert bound $|H(\widehat{S}_u) - H(S_u)| \leq \Delta_u \log(|\mathcal{C}_u| - 1) + h_2(\Delta_u)$ (or the trivial $\log|\mathcal{C}_u|$ bound outside its usual range) proves entropy tracking as $N_u \to \infty$. □

B.6 Numerical Size of the Optimization Bound

For the base single-item graph, $T = 30{,}000$, $B = 100$, $\epsilon = 10^{-7}$, $\gamma_x = 0.75$, and $\gamma_\xi = 1.25$. A valid uniform eligible-coordinate bound is

$$|\nabla_{x_{u,c}} \mathcal{L}_x| \leq q_u(|r_{u,c}| + 3B) + B\max\{1, |\log\epsilon + 1|\}.$$

The $3B$ covers HHI and cap terms; the last term covers entropy. Squaring and summing these bounds and applying the domain diameters yields a worst-case bound of 1.2×10^7 times the greedy-to-conic target cost, versus an observed repaired gap of 0.0042 on the same relative scale. This valid but vacuous worst-case constant does not certify practical accuracy; independent primal/dual objective bounds serve that role.

B.7 Controller Architecture

Figure 5 connects the aggregate and per-user constraints to the request-aware controller and subsequent serving windows.

C Adaptive Serving: Proofs and Scope

C.1 Proof of Theorem 1

Proof. Condition on the history before a window, fixing its policy. Single-item exposures are independent categorical vectors with mean R; write $e = \widehat{R} - R$. Allocate $\delta/6$ to each of six event families over all windows: caps, the two HHI terms, type counts, and two entropy terms.

Maximum share. Each creator indicator has variance $R_c(1-R_c)$ and range one. One-sided Bernstein gives $e_c \leq m_c$ over all JC coordinates, outside probability $\delta/6$.

HHI linear term. Expand

$$\operatorname{HHI}(\widehat{R}) - \operatorname{HHI}(R) = 2\langle R, e\rangle + \|e\|_2^2.$$

Each summand R_{C_i} has variance V_R and range A_R. One-sided Bernstein bounds $2\langle R, e\rangle \leq 2\sqrt{2V_R\ell_H/N} + 4A_R\ell_H/(3N)$ outside total probability $\delta/6$ over J windows.

HHI quadratic term. Jensen gives $\mathbb{E}\|e\|_2 \leq \sqrt{(1 - \operatorname{HHI}(R))/N} \leq 1/\sqrt{N}$. Replacing one observation changes $\|e\|_2$ by at most $\sqrt{2}/N$. Bounded differences imply $\|e\|_2 \leq (1 + \sqrt{\ell_H})/\sqrt{N}$ outside another total probability $\delta/6$. Combining the HHI events gives m_H.

Type counts. The binomial Chernoff bound gives $N_{j,u} \geq n_u$ except with probability at most $e^{-\ell_U}$ for each pair, hence at most $\delta/6$ overall.

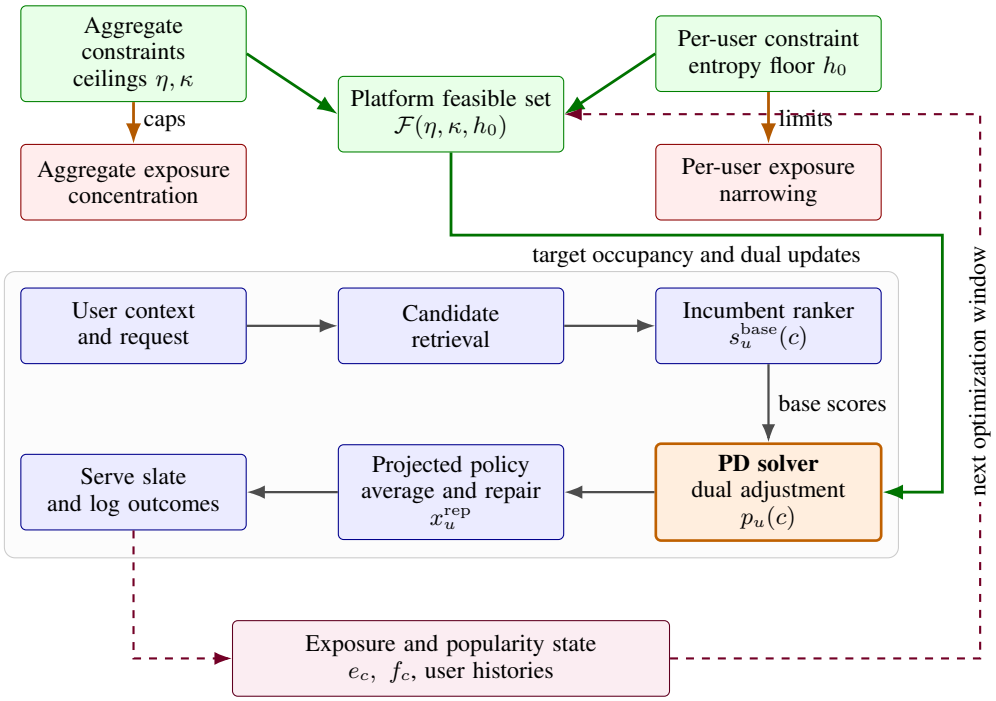

Figure 5: Request-level CWOP implementation. Aggregate ceilings and per-user entropy floors define feasible occupancy; request-aware primal-dual updates produce a randomized policy, repaired before serving. Outcomes update later optimization windows. Dual score adjustments alone are not a deterministic reranking implementation.

Per-type entropy, bounded-difference branch. Condition also on the type sequence. For n categorical observations from p on at most K points, the plug-in entropy bias satisfies

$$0 \le H(p) - \mathbb{E}H(\widehat{p}) = \mathbb{E}D(\widehat{p}\|p) \le \log(1 + (K-1)/n).$$

Indeed, $D(v\|p) \le \log \sum_i v_i^2/p_i$ and Jensen apply, with $\mathbb{E}\sum_i \widehat{p}_i^2/p_i = 1 + (|\operatorname{supp} p| - 1)/n$; omit zero-probability coordinates. This is the usual plug-in bias bound (Paninski, 2003). Moving one observation between bins changes entropy by at most $[n\log n - (n-1)\log(n-1)]/n \le (\log n + 1)/n$. One-sided bounded differences therefore give

$$H(p) - H(\widehat{p}) \le \log(1 + (K-1)/n) + (\log n + 1)\sqrt{\ell_U/(2n)}.$$

Both terms decrease for $n \ge 3$, so substitute n_u on the sufficient-count event. The trivial bound $\log K$ permits clipping. This branch costs at most $\delta/6$ over all JU pairs.

Per-type entropy, policy-dependent branch. For $p_i > 0$, direct expansion gives

$$H(p) - H(\widehat{p}) = \sum_i (\widehat{p}_i - p_i)\log p_i + D(\widehat{p}\|p).$$

The linear term is a mean of centered $\log p_I$, with variance $V(p)$ and range $A(p)$. Bernstein bounds it by $\sqrt{2V(p)\ell_U/n} + 2A(p)\ell_U/(3n)$. For any empirical type v, the multinomial theorem gives $\Pr(\widehat{p} = v) \le \exp[-nD(v\|p)]$. There are at most $(n+1)^{K-1}$ types: the first $K-1$ counts each lie in $\{0, \dots, n\}$ and determine the last. Consequently

$$\Pr\left(D(\widehat{p}\|p) > \frac{(K-1)\log(n+1) + \ell_U}{n}\right) \le e^{-\ell_U}.$$

Both linear and KL bounds decrease with $n \ge 3$, so replace n by n_u. The two event families cost $2\delta/6$ overall. For each type and history, select the smaller of the bounded-difference and policy-dependent branches before sampling. The selected branch has failure probability at most $2e^{-\ell_U}$, without paying for both branches. Clipping at $\log K$ is deterministic.

These bounds are uniform in the conditioned history; the tower property and union bound therefore apply under adaptive replanning. The six failure allocations sum to δ. The margins absorb the one-sided deviations. Global pooled exposure preserves the HHI and cap bounds by convexity. For each type, its pooled exposure uses its realized window-count weights; entropy concavity preserves its floor. □

C.2 General Slates and a Total-Variation Alternative

Policy-independent single-item certificate. When a margin is needed before a policy is available, use $\ell'_C = \log(5JC/\delta)$, $\ell'_H = \log(5J/\delta)$ and $\ell'_U = \log(5JU/\delta)$, with the corresponding count lower bounds $n'_u \geq 3$. For plans with $R_c \leq \kappa$, sufficient margins are

$$m'_C = \sqrt{2\kappa\ell'_C/N} + 2\ell'_C/(3N), \quad m'_H = \kappa\sqrt{2\ell'_H/N} + (1+\sqrt{\ell'_H})^2/N,$$

and the bounded-difference entropy margin $\min\{\log K_u, b'_u\}$. The proof uses $R_c(1-R_c) \leq \kappa$, Hoeffding for $R_{C_i} \in [0,\kappa]$, and five event families. Its base witness becomes feasible at 0.617 times the 20k reference window. This is a sufficient feasibility threshold; the utility and witness-mixing curves in Figure 3 describe its cost. Separate per-type entropy floors avoid replacing every floor by the largest entropy margin.

General slates. Normalized position-weighted slates contribute simplex-valued vectors rather than categorical vertices. A general certificate uses $\epsilon_R = \sqrt{\log(6JC/\delta)/(2N)}$ and

$$v_u = \min\left\{1, \frac{K_u}{2}\sqrt{\frac{\log(6J\sum_v K_v/\delta)}{Nq_u}}\right\}.$$

For $Nq_u \geq 8\log(3JU/\delta)$, coordinate Hoeffding bounds, the event $N_{j,u} \geq Nq_u/2$, and three union bounds give margins $(2\epsilon_R, \epsilon_R, \omega_{K_u}(v_u))$. Here $\omega_K(v) = v\log(K-1) + h_2(v)$ for $v \leq 1-1/K$, and $\log K$ otherwise, where h_2 is binary entropy (Audenaert, 2007). The HHI step uses $|\,\mathrm{HHI}(p) - \mathrm{HHI}(p')| \leq 2\|p-p'\|_\infty$. This coordinate certificate requires about 57.52 times the 20k reference window for the base witness. It applies to a broader class of exposure vectors than the categorical bounds above.

Alternatively, total variation is the maximum discrepancy over nonempty proper subsets. One-sided Hoeffding for their $2^{K_u}-2$ subset masses yields

$$\Pr(\mathrm{TV}(\widehat{S}_u, S_u) > v \mid N_u = n) \leq (2^{K_u}-2)e^{-2nv^2}.$$

This gives square-root rather than linear dependence on K_u, as in the classical finite-alphabet inequality (Weissman et al., 2003). The subset proof also covers simplex-valued slate contributions. The single-item theorem instead bounds categorical entropy directly, avoiding the additional continuity conversion.

C.3 Proof of Corollary 1

Proof. Mix an untightened optimizer with x^0 in proportion ϑ. Convexity gives residual at most $-\vartheta s_i \leq -\Delta_i$, preserving the exploration floor. Linearity gives loss $\vartheta(W^*_\epsilon - W_0)$. The tightened optimum can only improve this construction. The certificate does not apply if a required margin exceeds the witness slack. □

Together with Proposition 1, this separates target cost, sampling safety, exploration, and numerical error for the same window's score. Different controllers induce different future scores; their end-to-end utility difference is not this fixed-window budget or a regret bound.

C.4 Proof of Lemma 1

Proof. Base feasibility implies $0 \leq R_c \leq \kappa$, hence $R_c(1-R_c) \leq \kappa$ and $A_R \leq \kappa$. For $Y = R_I \in [0,\kappa]$, $Y^2 \leq \kappa Y$ gives $\mathrm{Var}(Y) \leq \mathbb{E}Y(\kappa - \mathbb{E}Y) \leq \kappa^2/4$. Substitution in (8), together with $m_u \leq \min\{\log K_u, b_u\}$, proves the componentwise envelope. Mix a base optimum with x^0 using

$\bar{\vartheta}$. Convexity gives $g_i(Qx) \le -\bar{\vartheta} s_i \le -\bar{m}_i$. The mixture is base-feasible, so its own margins obey $m_i(x) \le \bar{m}_i$ and it passes the certificate. Its score is the claimed lower bound. The same argument applies to any base-feasible policy of score L. No lower bound from this witness is asserted when $\bar{\vartheta} > 1$. □

This is a bound on the best certified score, with a constructive attainable lower bound. Refinement need not find the globally best policy-dependent certified plan. An exploration floor alone does not determine which entropy branch is smaller: witness mixing can move all probabilities far above the floor.

C.5 Model Error and Simulation-Based Calibration

The request law and sampling model must match the plan. Arbitrary history-dependent scores include the experiment's reinforcing scores, but the theorem does not establish their convergence. If a separately validated request-law estimate satisfies $\mathrm{TV}(q, \widehat{q}) \le e_q$, a fixed policy has $\|R_q - R_{\widehat{q}}\|_\infty \le e_q$. Additional reserves e_q for the cap and $2e_q$ for HHI handle this mismatch; count bounds need valid lower bounds on true q_u. Confidence failures for those estimates must also be budgeted. Unknown-law learning is not evaluated here.

Finite Monte Carlo estimates are not exact safety guarantees. A certificate based on sampling a known policy needs independent validation, one-sided binomial confidence bounds, and allocation of validation and serving failures across windows and searched policies. Ordinary bootstrap margins do not automatically meet these requirements. The certified experiment uses analytic margins without outcome-based selection.

D Experimental Protocols and Additional Controller Results

These experiments trace the joint-control argument from numerical feasibility to realized exposure: optimization checks support Q1, feedback and sampling tests support Q2, and listening-derived candidates support Q3. The ablations identify which parts of CWOP produce joint feasibility, utility retention, and serving safety.

D.1 Relative Reporting Conventions

Experimental outcomes are expressed as dimensionless comparisons. For static synthetic scores, let W_g, W_{ref}, and W_0 denote greedy, conic-reference, and uniform-witness scores on the same instance. Tables use relative loss $L = (W_g - W)/(W_g - W_{\mathrm{ref}})$ and gap $G = (W_{\mathrm{ref}} - W)/(W_g - W_{\mathrm{ref}})$; safety curves use relative improvement $Q = (W - W_0)/(W_{\mathrm{ref}} - W_0)$. These quantities are invariant to shifting all scores by a common constant. The stress and same-state diagnostics also scale their gaps by the conic improvement over the witness.

Feedback utilities are divided by the mean no-margin conic utility for the same dataset and window protocol. Standard deviations use that same fixed denominator; they do not propagate uncertainty in the reference mean. Exceptions are explicitly captioned: sensitivity tables use uncontrolled utility at the same feedback strength, and margin ablations use the corresponding no-margin solver or frozen-policy utility. Public static Last.fm utilities use the conic value within each split. HHI and maximum-share outcomes are divided by their respective targets, entropy by $\log K$, and violation outcomes are percentages. Diagnostic urn HHI uses the uniform floor $1/K$. Runtime comparisons are ratios. Theoretical constants, illustrative mathematical examples, and public/synthetic design settings remain in their original units. Private-platform summaries instead use rounded ratios and generic signal labels (Appendix E).

D.2 Independent Optimum, Feasibility, and Error Accounting

The computational implementation separates the original controller, a vectorized eligible-edge implementation, and an independently formulated CVXPY conic model. The conic variables are the eligible single-item probabilities; row sums, exposure sums, quadratic HHI, and exponential-cone entropy constraints are assembled separately. Clarabel uses absolute/relative gap and feasibility tolerances 10^{-10} and a 300-iteration cap. SCS uses tolerance 2×10^{-7} and at most 200,000 iterations

Table 4: Base synthetic outcomes in relative units. L is defined in Table 1; $V = \max\{0, \mathrm{HHI}/\eta - 1, \max_c R_c/\kappa - 1, (h_0 - \min_u H(S_u))/h_0\}$. Entropy is divided by $\log K$. Numerical feasibility is checked with tolerance 10^{-10}.

Policy	L	HHI/η	Share/κ	$H/\log K$	V
Greedy	0.0000	16.8264	12.0153	0.0000	16
Softmax 0.30	0.7356	2.6706	3.6922	0.6157	2.7
Global-only PD	0.8315	0.9982	1.0003	0.0484	0.93
Local-only PD	0.5667	3.6562	4.6658	0.6502	3.7
PD raw	1.0017	0.9826	1.0006	0.6509	0.00058
PD repaired	1.0042	0.9815	1.0000	0.6530	0
Feasible softmax mixture	1.9542	0.7486	1.0000	0.9867	0
Conic reference	1.0000	0.9888	1.0000	0.6500	0

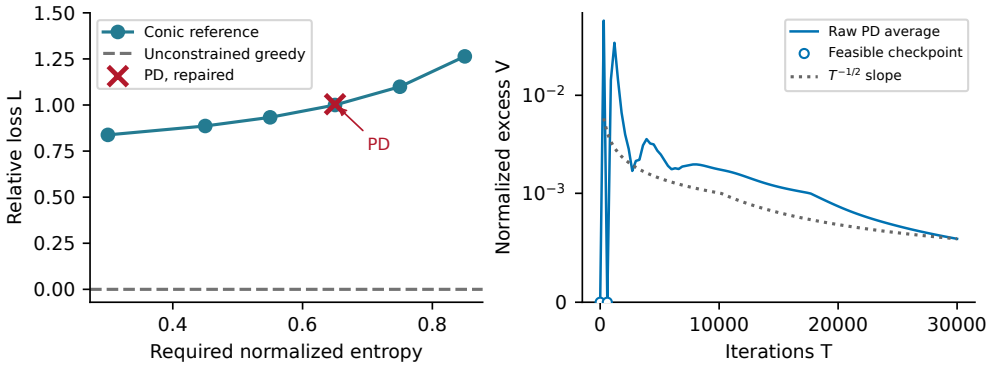

Figure 6: Static benchmark. Left: feasible frontier in relative-loss units L (Table 1), greedy reference, and repaired PD point. Right: the raw average's normalized maximum excess V (Table 4); the final serving policy is repaired. The red cross is drawn above the nearly coincident reference point, at its true coordinates. The vertical scale is linear below 10^{-3} and logarithmic above it, so the feasible checkpoint at iteration 600 is shown at zero without an artificial positive floor. Uniform initialization is feasible; subsequent utility ascent can increase violation, so convergence is not monotone. The dotted $T^{-1/2}$ line is a reference slope anchored at the last residual, not an error certificate.

on the base instance. Its feasible objective is consistent with Clarabel at the numerical tolerance used, but its tangent certificate is looser; the reported reference bounds use Clarabel.

For nonnegative returned constraint multipliers ξ, form the concave Lagrangian $L(x, \xi)$. At a strictly positive feasible point y, concavity gives the independently computed upper bound

$$\max_{x \in \mathcal{X}_\epsilon,\, g(Qx) \le 0} W(Qx) \le L(y, \xi) + \max_{x \in \mathcal{X}_\epsilon} \langle \nabla_x L(y, \xi), x - y \rangle.$$

The last maximum is rowwise linear optimization on a floored simplex. A repaired feasible policy supplies the lower bound. Computation uses a 10^{-9} mixture with uniformity to keep the tangent point strictly positive. These are floating-point numerical certificates, not interval-arithmetic proofs. Several solves report `optimal_inaccurate`; their statuses and checked upper–lower gaps are retained. The saved records retain the checked numerical gap for every reference solve.

The controller's raw average can be infeasible. We first compute the sufficient mixture weight from Proposition 1, then bisect along the segment toward a strict uniform-over-candidates witness, requiring every residual to be at most -10^{-10}. This reduces needless mixing without claiming an optimal projection. The oracle and every reported feasible baseline undergo the same residual checks. Table 4 reports the base comparison. The global-only policy retains low per-type entropy, while the local-only policy exceeds the concentration ceilings. The repaired joint policy meets both target families within the numerical tolerance. This illustrates the complementary roles of the two controls in Section 2.2; the raw joint policy also shows why numerical feasibility repair is part of the implementation.

The original synthetic utility generator uses creator effects $N(0, 0.12^2)$, shared-head bonuses linearly spaced from 1.35 to 0.90, group bonuses 0.85, and independent user–creator noise $N(0, 0.10^2)$. Request weights are independent Uniform$(0.65, 1.35)$ draws, normalized to sum to one. The base graph gives every type six shared head creators, five group creators, and seven sampled alternatives. The larger graph has 48 types, 72 creators, 30 candidates, eight head creators and eight groups. Its targets are $(2/72, 2.88/72, 0.65 \log 30)$. All sampled utilities, masks, request probabilities, and computed policies are saved.

Base graph seeds are 20260722–20260726; the larger graph uses 20260752. PD uses $\epsilon = 10^{-7}$, primal rate $0.75/\sqrt{t}$, and dual rate $1.25/\sqrt{t}$; the global-only ablation uses $5/\sqrt{t}$. Uniform witnesses certify dual ℓ_1 bounds below $0.42B$ on the base graphs with $B = 100$. The larger graph's bound is below $0.76B$ with $B = 200$. Exploration-gap upper bounds are below 6.4×10^{-6} times each instance's greedy-to-conic cost; both floored and unfloored oracle solutions are retained. These are distinct from numerical optimization error.

The feasible softmax envelope searches temperatures $\{0.01, 0.05, 0.1, 0.2, 0.3, 0.5, 1, 2\}$ and applies exact segment feasibility checks. It optimistically chooses the best feasible utility on each instance, rather than claiming a held-out tuning result. Global-only and local-only are diagnostic ablations, not competitive joint-constraint methods. Figure 6's frontier solves the constrained problem independently at entropy floors $\{0.30, 0.45, 0.55, 0.65, 0.75, 0.85\}$; it does not connect infeasible controller iterates.

D.3 Repeated Feedback and Comparator Definitions

The response model and frozen-window feedback score are specified in Section 5.2. We give the split, tuning and comparator details below. The means $\bar{f}$ and $\bar{A}_u$ are over the global catalog and each type's candidates, respectively. Impressions update the exposure counts used by the log-share and FairCo adaptations; sampled clicks update popularity and affinity.

Synthetic training graphs use seeds 20260901 and 20260902; tests use 20260910–20260914. Last.fm tuning uses SVD splits 20260916 and 20260917, disjoint from test splits 20260920–20260922. Every method receives 20 windows of 20,000 requests on synthetic graphs and 100,000 on Last.fm graphs. A deterministic seed by graph and window governs request counts, creator counts and responses. Different policies induce different outcomes. Tuning selects highest mean intrinsic utility among settings with at most 5% individual-window violations across both training trajectories. If none qualify, minimize violation fraction, then mean normalized excess, then maximize utility. The normalized excess in a window is $\max\{0, \mathrm{HHI}/\eta - 1, \max_c R_c/\kappa - 1, (h_0 - \min_u H(S_u))/h_0\}$. A violation means a raw residual exceeding 10^{-6}, with entropy in nats. Each dataset has 46 training trajectories. Four selected baselines and two conic policies give 30 synthetic and 18 Last.fm test trajectories. All outcomes and final policies are retained. The FW grid uses 10,000 steps on every training and test run. The comparator definitions are:

- *Temperature*: policy entropy regularization, with temperature searched over $\{0.15, 0.3, 0.6, 1.2, 2.4, 4.8, 9.6, 19.2\}$.
- *Log-share*: subtract λ times both the global and local log-exposure ratios, using one pseudocount per eligible creator, then softmax at temperature 0.1. Search $\lambda \in \{0.05, 0.15, 0.3, 0.6, 1.2\}$.
- *FairCo-style batch*: singleton creator groups with known merit $m_c = \sum_u q_u p_{u,c} \mathbf{1}\{c \in \mathcal{C}_u\}$. Add $\lambda[\max_k E_k/m_k - E_c/m_c]$ to the current score, with cumulative exposures E, then choose the top item (floor 10^{-7}). Search $\{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}\}$. This adapts the exposure-disparity score in FairCo (Morik et al., 2020), but freezes updates for an entire window, uses a biased score and known merits, and does not reproduce its online estimator. Its native merit-proportionality target differs from our HHI/entropy intersection, so this comparison evaluates the adaptation on our specified targets.
- *Two-sided batch FW*: 10,000 conditional-gradient steps, initialized at uniformity, maximize $W_s(x) - AC\,\mathrm{HHI}(R) + A\sum_u q_u H_{10^{-3}}(x_u)$, where $H_\zeta(x) = -\sum_c (x_c + \zeta)\log(x_c + \zeta)$ on eligible edges. The vertex selects the largest adjusted score per type, with step $2/(t+2)$. Search $A \in \{0.1, 0.3, 1, 3, 10\}$. This uses the Frank–Wolfe ranking principle of Offr (Usunier et al., 2022), but is a batch, known-q, new-objective adaptation rather than the original online algorithm.

Table 5: Matched-window policies. Concentration ratios are computed using each graph's targets before averaging pooled outcomes across trajectories. Exposure and clicked-item entropies are divided by $\log K$. Parameters are selected on training instances.

Data	Method	Parameter	HHI/η	Share/κ	Exposure H	Clicked H
Synthetic	Temperature	2.4	0.702	0.865	0.995	0.940
Synthetic	Log-share	0.6	0.605	0.760	0.974	0.917
Synthetic	FairCo-style batch	0.0001	0.866	1.127	0.573	0.557
Synthetic	Two-sided batch FW	0.3	0.548	0.616	0.845	0.803
Synthetic	Conic reference, no margin	—	0.998	1.003	0.646	0.625
Synthetic	Conic reference + certificate	—	0.907	0.892	0.762	0.733
Last.fm	Temperature	2.4	0.768	0.876	0.993	0.786
Last.fm	Log-share	0.6	0.704	0.813	0.968	0.712
Last.fm	FairCo-style batch	$1e-06$	13.146	11.824	0.000	0.000
Last.fm	Two-sided batch FW	0.3	0.679	0.625	0.844	0.540
Last.fm	Conic reference, no margin	—	0.998	0.960	0.644	0.259
Last.fm	Conic reference + certificate	—	0.963	0.871	0.845	0.449

Table 6: Construction ablation under the matched protocol. Utility and loss use the dataset-specific no-margin conic reference from Table 2. Window and frozen-policy failures are percentages. These compare separate feedback trajectories, not the fixed-history score bound.

Data	Construction	Utility ratio	Loss	Window fail.	Frozen fail.
Synthetic	Witness segment	0.904	9.6%	0.0%	0.0000%
Synthetic	Reoptimized	0.959	4.1%	0.0%	0.0025%
Last.fm	Witness segment	0.754	24.6%	0.0%	0.0000%
Last.fm	Reoptimized	0.829	17.1%	0.0%	0.0000%

- *Conic reference*: independently solve the same window score objective and original targets using the conic oracle, repair numerical residuals, and deploy it. It supplies the most accurate constrained solutions on these graphs.
- *Conic reference + certificate*: checked reoptimization from Section 4, with $J = 20$ and $\delta = 0.05$.

The per-window objective is the reinforced score, whereas reported utility is evaluated using p. All controllers have the same information but different computational budgets; equal iteration counts do not imply equal computational cost. Feedback runtimes are omitted; static optimizer timings are reported separately. FORM provides related unknown-reward learning guarantees beyond this known-prior protocol. Table 5 reports the selected parameters and clicked-item entropy.

Constraint-specific outcome accounting. Table 3 uses the same saved matched-window trajectories. Global failure means empirical HHI or maximum share exceeds its own target by more than 10^{-6}; local failure means any type's entropy falls below its floor by more than 10^{-6} nats. Concentration columns maximize each window's ratio to its graph-specific target, and the entropy column minimizes $H/\log K$ over all types and windows. These worst-window quantities differ from the pooled averages in Table 5. Regenerating request, creator, and click draws from all 48 saved policy sequences and their original seeds reproduces all 960 window metrics to floating-point precision. The local verification script records input checksums and constraint-specific counts for every baseline.

Certified construction ablation. Table 6 compares the witness segment with reoptimization. Both are evaluated at the matched window sizes and retain the same probability certificate; their utility differences arise along separate feedback trajectories. Reoptimization retains more utility on both datasets while preserving certification, supporting the usefulness of optimizing within the jointly safe set.

Additional 10k-window diagnostics. Separate synthetic ablations use $20 \times 10{,}000$ requests to compare PD and the conic reference and vary feedback strength. PD uses 10,000 updates

Table 7: Feedback sensitivity at 10k requests per window. Utility is divided by uncontrolled utility at the same χ; concentration uses the target as denominator. Means are over five graphs. Parameters are selected only at $\chi = 0.15$.

χ	Method	Utility ratio	HHI/η	Share/κ	$H/\log 18$
0.00	Uncontrolled	1.000	2.356	3.165	0.697
0.00	Temperature	0.745	0.656	0.785	0.997
0.00	Log-share	0.753	0.565	0.648	0.983
0.00	PD, no margin	0.972	0.945	1.004	0.658
0.15	Uncontrolled	1.000	16.074	11.747	0.082
0.15	Temperature	0.700	0.702	0.866	0.994
0.15	Log-share	0.721	0.605	0.759	0.974
0.15	PD, no margin	0.866	0.997	0.966	0.733
0.30	Uncontrolled	1.000	16.588	11.937	0.073
0.30	Temperature	0.716	0.768	0.964	0.988
0.30	Log-share	0.760	0.692	0.951	0.953
0.30	PD, no margin	0.852	0.996	0.951	0.757

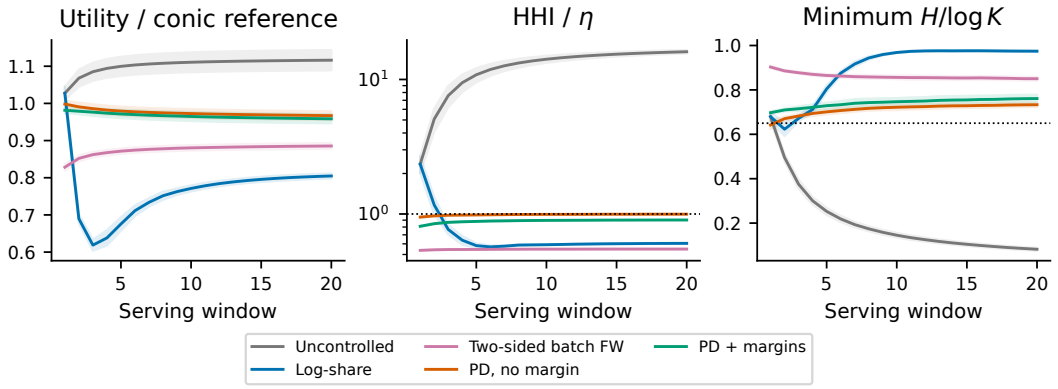

Figure 7: Controllers inside the feedback process. Curves show cumulative served distributions and relative intrinsic utility, normalized to the final mean no-margin conic utility in this 10k-window protocol. HHI is divided by its target, and entropy by $\log K$. Shaded bands are one SD across five held-out graphs. Dotted lines mark original HHI and entropy targets. Pooling can hide individual-window violations, which are reported separately in Table 2.

per window. An uncontrolled comparison uses softmax temperature 0.1. Empirical margins $(0.005, 0.01, 0.07 \log 18)$ are fixed before serving. These experiments isolate solver error and margin sensitivity; the matched-window certified comparisons above all use Theorem 1. Their role is to distinguish protection of exposure targets from the choice of solver and from the strength of the feedback score.

Feasibility without margins is a property of the planned distribution. A window is counted as violating when its largest raw residual exceeds 10^{-6} (entropy residuals are in nats). The oracle can therefore have optimal planned utility and still violate empirical targets. The pooled distribution uses actual request counts, not the planned q. Clicked-item entropy is reported in Table 8; it is not constrained by CWOP. Reported spreads are sample standard deviations across five graphs. Paired utility intervals use Student's t_4 multiplier on within-seed differences; they quantify this small synthetic design. A coordinate union-bound certificate needs about 57.52 times the 20k reference window, versus 0.617 for the policy-independent categorical bound on the same base witness, $J = 20$, and $\delta = 0.05$.

D.4 Additional 10k-Window Diagnostics and Solver Comparison

Where log-share fails. At the diagnostic window size of 10k, the log-share controller violates at least one target in 48% of held-out windows: 38% violate HHI, 48% the cap, and 34% en-

Table 8: Clicked-item entropy divided by $\log 18$ and percentage of violating trajectories (any failure in 20 windows). Runtime is omitted because execution conditions differed.

Method	N	Min clicked $H/\log 18$	Violating runs
Uncontrolled	10k	0.071	100%
Temperature	10k	0.941	0%
Log-share	10k	0.918	100%
FairCo-style batch	10k	0.602	100%
Two-sided batch FW	10k	0.803	0%
PD, no margin	10k	0.682	100%
Conic reference, no margin	10k	0.625	100%
PD + margins	10k	0.714	0%
Conic reference + margins	10k	0.688	0%

tropy (overlapping rates). Cap failures occur within windows 1–10, HHI within 1–8, and entropy within 2–8. Later redistribution yields diverse pooled exposure but cannot undo earlier failures. This supports measuring the two exposure targets in each serving window, as required by CWOP's safety guarantee, rather than relying on pooled diversity alone. The authors retain every violating seed/window/constraint tuple in the computational records.

Same-state optimization diagnostic. The projected and reference policies induce different feedback trajectories. Their pooled entropy difference (0.733 versus 0.643) alone cannot diagnose convergence. We therefore rerun all five projected trajectories and solve the reference problem on each identical frozen score. Dividing each score gap by its same-state conic improvement over the uniform witness gives mean 0.102, maximum 0.135, and minimum 0.015. The $10{,}000$-step projected optimizer is materially less accurate here than in the static benchmark. The extra reference solve is diagnostic only and does not affect serving or state updates. This comparison attributes the gap to numerical optimization at the same history, so it is not mistaken for an unavoidable cost of the exposure targets.

Uncertainty and clicked-item diversity. Windows within a trajectory share evolving states. The empirical-margin experiment has zero violations across five independent generated trajectories, not 100 independent trials. Under an i.i.d.-trajectory interpretation, the exact one-sided 95% binomial upper bound for the probability of any violation during a 20-window trajectory is $1 - 0.05^{1/5} = 0.451$, a weak bound. We do not substitute a 3% per-window bound. The analytic theorem supplies a different guarantee conditional on its model. Minimum normalized clicked-item entropy is 0.682 without PD margins and 0.714 with the empirical margins, versus corresponding exposure entropies 0.733 and 0.761. The conic reference with margins has clicked entropy 0.688 and exposure entropy 0.713. The exposure–consumption distinction is therefore also present in the measured outcomes: the established joint guarantee applies to what is served.

Margin sweep and cost budget. Both optimizers use empirical reserves $t(0.005, 0.01, 0.07\log 18)$ for $t \in \{0, 0.5, 1, 1.5\}$ at $N = 10$k. PD violation percentages are $(74, 3, 0, 0)$ and conic reference percentages $(100, 58, 0, 0)$. At $t = 1$, Corollary 1 gives $\vartheta = 0.4091$ on all five graphs. Thus the sufficient score-cost budget is 40.91% of the greedy-to-witness score range in each fixed window. This budget is distinct from the observed cross-trajectory intrinsic-utility changes.

Table 9 reports this empirical-margin ablation. Analytically certified comparisons use the single protocol in Table 6. The reduction in violations for both solvers supports the sampling margin as a separate component of joint exposure control. PD's lower violation rate at the intermediate margin does not establish a solver advantage: its less accurate solutions can lie further inside the targets.

D.5 Public Artist-Listening Protocol and Optimization Tuning

We use only the numeric listening triples in the official HetRec Last.fm archive, which permits noncommercial use (Cantador et al., 2011). The archive checksum is retained in the computational records. Select the 256 artists with most distinct listeners (artist identifier breaks ties). Among

Table 9: Empirical-margin ablation at 10k requests. Multiplier t scales $(0.005, 0.01, 0.07 \log 18)$. Utility is relative to the same solver without margins; ϑ is the sufficient witness mixing fraction in Corollary 1. These settings have no analytic sampling certificate.

Method	t	Utility ratio	Window fail.	ϑ
PD	0.0	1.000	74%	0.000
PD	0.5	0.996	3%	0.205
PD	1.0	0.991	0%	0.409
PD	1.5	0.988	0%	0.614
Conic reference	0.0	1.000	100%	0.000
Conic reference	0.5	0.991	58%	0.205
Conic reference	1.0	0.979	0%	0.409
Conic reference	1.5	0.965	0%	0.614

Table 10: Public artist-listening test splits. Utilities and PD gaps are relative to the conic reference within each split. Time is PD/conic solver time, excluding preference fitting.

Split	Greedy/ref.	Conic/ref.	PD/ref.	Gap/ref.	Time ratio
Split 1	1.811	1.000	0.988	1.24%	2.28
Split 2	1.790	1.000	0.989	1.11%	2.29
Split 3	1.821	1.000	0.989	1.06%	2.43

users with at least 20 positive entries in that submatrix, sample 128 with seed 20260918. Use all 1,892 users for fitting. For each split, withhold $\lfloor 0.2 n_u \rfloor$ observed positives for users with at least five positives, transform training counts by $\log(1 + \text{count})$, and compute a rank-16 truncated SVD. Clamp reconstructed values below zero to zero and divide by the largest training log count. Unobserved entries are zeros in this surrogate objective, not assumed genuine negative preferences. The three test split seeds (Splits 1–3) are 20260920–20260922.

Each selected user's constructed retrieval set consists of its top 32 predicted artists plus 32 seeded alternatives; request probabilities are uniform. Held-out Recall@20, excluding training positives, is 0.539–0.583, versus 0.286–0.327 for training popularity. This diagnoses prediction on recorded positives only; it does not calibrate missing feedback or causal utility. Targets are 1.5 times the uniform-over-candidates witness's HHI and maximum share, plus $0.65 \log 64$ entropy. The graph, not a universal uniform catalog, therefore determines feasibility; these data-dependent targets are declared before optimizer comparison.

On the disjoint split 20260919, test primal coefficients $\{0.25, 0.75, 2.25, 6.75\}$ and dual coefficients $\{5, 20, 80, 240\}$ for 5,000 iterations, choosing maximum repaired feasible utility. The selected coefficients $(2.25, 240)$ are then frozen for all test splits, each run for 30,000 iterations. Uniform witnesses certify $B = 200$. Table 10 reports both optimizer times for the three test splits. The conic comparison separates the accuracy of the PD implementation from the feasibility of joint exposure control on learned preference scores. Serving safety on the same candidate construction is tested in Q3.

D.6 Policy-Dependent Safety and Frozen-Policy Tests

These checks support the serving component of the thesis: a joint target is useful only when its certificate is computationally attainable and its utility cost is measured at the intended window size.

Larger optimization instance. A separately seeded stress test has 512 request types, 1,024 creators, and 64 candidates per type (32,768 variables). It uses eight shared-head creators, 32 groups, and the synthetic utility generator above. Targets are 1.5 times the uniform-over-candidates witness's HHI and maximum share, with entropy floor $0.65 \log 64$. Relative to the conic score improvement over the uniform witness, the conic upper–lower gap is 1.1×10^{-7}. At 10,000 iterations and a witness-certified dual box $B = 800$, repaired PD's gap is 0.350 on this scale, with witness mixture 0.2494. The policy-dependent witness threshold is 2.284 times a 100k reference window. The non-negligible PD gap identifies an implementation-accuracy constraint, while the witness threshold

Table 11: Base synthetic safety cost. Window size is relative to $N_0 = 20$k. $Q = (W - W_0)/(W_{\mathrm{ref}} - W_0)$ measures score gain above the uniform witness, with no-margin conic gain equal to one. ϑ uses policy-independent per-type margins. Policy-dependent values are feasible constructions, not certified optima.

N/N_0	ϑ	Common Q	Per-type Q	Policy-dep. Q
0.617	0.999	0.041	0.485	0.750
1.000	0.807	0.556	0.670	0.848
2.500	0.546	0.781	0.824	0.925
5.000	0.409	0.858	0.882	0.953

measures the sampling cost of protecting more request types. Together they locate two distinct costs of joint control on this single-item graph.

Constructing and checking policies. The implementation evaluates every term of Theorem 1 from the known request probabilities and the proposed policy. The witness is uniform over each request type's candidates. A 64-interval search along the segment from a conic solution to the witness locates a certified point, then 30 bisection steps refine that interval. Every final policy is checked directly (floating-point tolerance 10^{-10}). This search does not assume global monotonicity of the policy-dependent certificate. If no point passes, the procedure reports no certificate. No serving outcomes enter this search.

For both static curves and feedback experiments, up to three additional conic solves use the preceding certified policy's margins and a separate entropy floor for every type. Each candidate is checked again and, when necessary, mixed with the witness. The highest-utility certified construction is retained. The result is a feasible lower bound on attainable certified utility. The fixed common-cap implementation uses $\max_c m_c$ during these extra conic solves, while final validation checks each creator separately. Whenever the uniform envelope fits the witness slack, the implementation also considers the mixture attaining the computable lower bound of Lemma 1. The segment-only construction is an ablation. One optional refinement solve fails on Last.fm Split 2, window 3; the procedure retains its previously certified plan. Every deployed plan passes the final check. The policy-dependent entropy branch is selected in 100% of synthetic type-windows and 0% of Last.fm type-windows. An exploration lower bound alone does not force the actual probabilities to lie near that bound.

The base policy-independent witness is feasible at 0.617 times $N_0 = 20$k requests. With per-type entropy margins its coefficient is 0.999 there, versus 0.807, 0.546, and 0.409 at $N/N_0 = 1, 2.5, 5$. The policy-dependent threshold is $0.254N_0$ on that graph and $0.232N_0$–$0.267N_0$ across the five feedback test graphs. For Last.fm it is 0.337 of the 100k reference window on each split, slightly above the policy-independent ratio 0.332. Thus the policy-dependent bound is not uniformly tighter. On the first static Last.fm split, certification retains relative score gain $Q = 0.506$ at 50k and $Q = 0.720$ at 100k, with $Q = 1$ for the conic reference and $Q = 0$ for the uniform witness. Missing curve points occur when the chosen witness and construction do not yield a certificate. All reported curve solves have an empty failure log. Missing values are not interpolated and do not establish that every policy is infeasible at that size.

Feedback and response generation. The synthetic safety runs use seeds 20260910–20260914, $J = 20$, and $N = 20{,}000$. Last.fm uses test splits 20260920–20260922, $J = 20$, and $N = 100{,}000$. Response probabilities on Last.fm equal the learned proxies clipped to $[0, 1]$. The underlying observations are listening counts; the Bernoulli responses are generated for this experiment. Per-window counts are drawn by first sampling a multinomial type count and then a multinomial creator count within each type. Bernoulli responses are generated by binomial thinning of those counts. This is distributionally equivalent to independent request-level serving under the specified frozen policy. The no-margin and certified conic policies have separate evolving feedback histories. The matched comparison saves all 960 final policies, graphs, request laws, and per-window results, including 160 policies with analytic certificates.

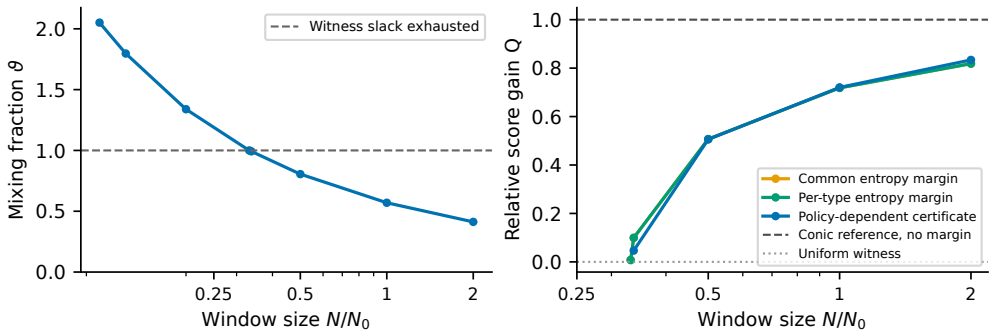

Figure 8: Last.fm Split 1: witness coefficient and relative static score gain Q (Appendix D.1), against N/N_0 with $N_0 = 100$k. These use a fixed learned score, distinct from the evolving feedback experiment.

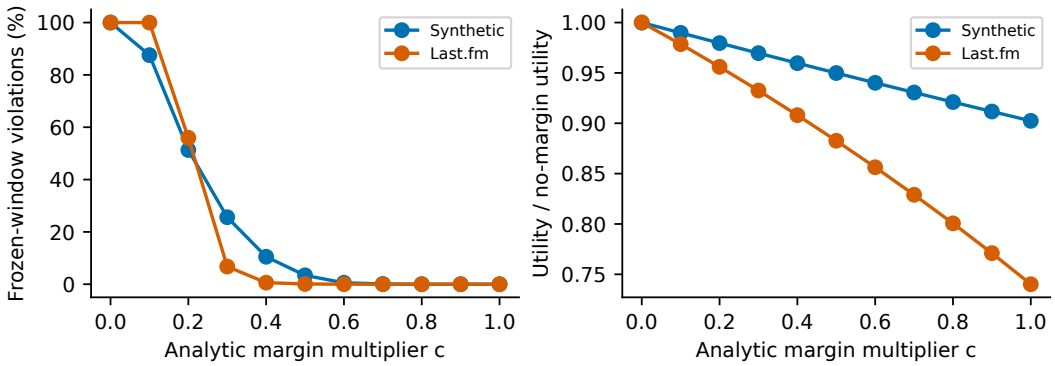

Figure 9: Safety and relative utility on fixed first-window scores. Each utility curve is divided by its own no-margin value. For each multiplier c, the witness-segment policy satisfies targets tightened by c times its own analytic margins. Each point uses 2,000 independent resamples; $c = 1$ passes the full certificate. Synthetic windows have 20k requests and Last.fm windows 100k.

Repeated conditional sampling. For each data source we use the first test graph/split and freeze each of its 20 saved policies, separately for the no-margin and certified trajectories. Every policy is resampled 2,000 times with independent multinomial draws at its original window size. This gives 40,000 replications per method and data source. Every no-margin replication violates at least one target. The certified synthetic plan at window 10 has a 0.05% cap-failure rate; all other certified tests pass. The one-sided 95% Clopper–Pearson upper limit is 0.002370 for that plan. For each zero-failure plan it is $1 - 0.05^{1/2000} = 0.001497$. These pointwise intervals quantify the conditional risk reduction achieved by certified policies. Simultaneous empirical inference over all saved plans would require multiplicity adjustment; Theorem 1 instead supplies the adaptive trajectory guarantee under its sampling assumptions.

Margin sensitivity. On the first window's fixed score, take its conic optimum and search the witness segment for policies satisfying $g_i(Qx) + c\Delta_i(x) \le -10^{-10}$, where $c \in \{0, 0.1, \dots, 1\}$ and $\Delta(x)$ is the final policy's analytic margin vector. The search uses 64 grid intervals and 35 bisection steps, checking every returned policy directly. Figure 9 reports 2,000 independent resamples at each point. At $c = 1$, the complete probability certificate holds. Smaller multipliers diagnose conservatism without the full theorem's failure budget. The first zero-failure grid point is $c = 0.8$ on the synthetic graph and $c = 0.6$ on Last.fm. At $c = 0.5$, failure rates are 3.4% and 0.1%, respectively; at $c = 0.6$, they are 0.55% and 0%. These tests show model- and policy-specific conservatism, not a universal factor of two or proof of zero failure probability.

Table 12: Matched-set concentration ratios, rounded to one decimal place. Intervals are 95% creator-resampling sensitivity intervals, with endpoints rounded outward to avoid narrowing the original intervals. No absolute HHI or creator counts are released.

Comparison	HHI ratio	Sensitivity interval
Delivered exposure / audience growth	$1.5\times$	$[0.4, 12.0]$
Delivered exposure / audience stock	$8.3\times$	$[3.3, 15.8]$

D.7 Implementation, Runtime, and Reproduction Scope

The compact implementation stores UK eligible probabilities for equal-size candidate sets and sorts each row for Euclidean simplex projection. It uses the same current-state residuals, pre-update average, and x-coordinate gradients as Algorithm 1. Finite-difference checks and an equivalence test compare it with the original loop implementation. Eligible-edge storage is $O(UK)$, excluding the conic solver and saved traces; this is not a peak-memory benchmark. Static optimizer runtimes accompany their tables; feedback runtimes are omitted. All evaluated implementations allocate a single item per request. The general slate formulation specifies the mathematical extension, rather than an additional implemented benchmark.

The authors retain fixed seeds, complete numerical outputs and certificates, environment arrays, tests, and a frozen dependency list. These records are separate from the anonymized source package. The production audit requires governed access: the manuscript supplies coarsened relative summaries and reproduction requirements, while private row-level data and identifying metadata are withheld. The public Last.fm archive is downloaded from its official source rather than bundled for redistribution.

E Anonymized Platform Audit and Constraint Accounting

This audit motivates the concentration-control component of the thesis and quantifies how a cap redistributes an observed exposure distribution. It supplies descriptive context for CWOP; the evidence for simultaneously controlling creator concentration and per-user narrowing comes from the controlled feedback tests.

E.1 Scope and Relative Concentration

The confidential extract records delivered content and aggregate creator-audience measurements from a content platform. We use generic analytical labels throughout: *delivered exposures* measure serving, *qualified views* and *engaged reads* measure attention, and *positive reactions* and *saved items* measure engagement. These labels describe signal roles rather than reproduce source-field names. The public description omits platform identity, absolute dataset sizes, collection dates, schema details, and exact sampling intervals. Only rounded relative summaries are reported. Rounding is a disclosure precaution, not a formal privacy guarantee.

Figure 10 divides each signal's creator HHI by the HHI of delivered exposure on the same complete extract. Both attention indicators are more concentrated than exposure, whereas the two engagement indicators are less concentrated. These signals represent different actions and supports, so the result describes their realized distributions rather than interchangeable measures of demand.

For comparisons with audience measures, we intersect exposed creators with creators observed in a pair of audience snapshots. This matched subset accounts for a small fraction of delivered exposure and is not representative of the complete serving distribution. *Audience stock* is the later recorded audience size; *audience growth* is the nonnegative part of the difference between snapshots. The same creator set is used for every ratio in Table 12. Intervals resample creators with replacement using a fixed bootstrap protocol. The growth comparison includes parity. The stock comparison spans different accumulation horizons and cannot isolate an allocation effect.

The matched Lorenz curves cross in the lower tail. Accordingly, we claim index-specific concentration differences, not Lorenz dominance. The fine-grained curve, exact crossing, and underlying

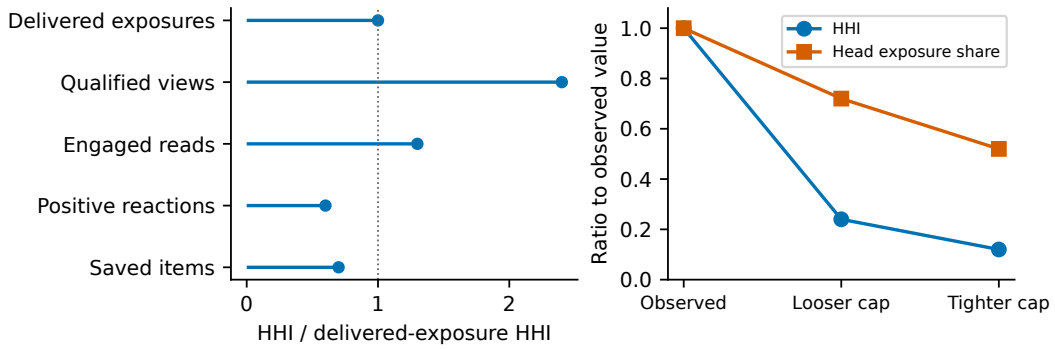

Figure 10: Rounded relative platform summaries. Left: signal HHI divided by delivered-exposure HHI on the complete extract. Right: concentration after aggregate cap accounting, divided by the observed value on the matched set. The two panels use different creator universes and denominators. Head share uses the same fixed leading-creator group definition across accounting conditions.

Table 13: Aggregate cap accounting. Every outcome is divided by its observed value on the same matched creator set. Caps and the fixed head-group size are withheld; labels indicate their ordering. Ratios are rounded for disclosure control.

Condition	HHI ratio	Effective-count ratio	Head-share ratio
Observed	1.00	1.0	1.00
Looser cap	0.24	4.1	0.72
Tighter cap	0.12	8.4	0.52

vectors are withheld to reduce source identification risk. The plotted summaries below retain the direction of the concentration comparison without releasing that distributional detail.

E.2 Audience Dynamics and Cap Accounting

A separate panel associates audience stock with the subsequent recorded growth measure, retaining zero growth and pooling sparse categories. A Poisson GLM with interval and category effects and creator-clustered standard errors estimates the stock–growth elasticity. Relative to unit elasticity, the estimate is about 0.7, with an outward-rounded 95% interval $[0.6, 0.9]$. This recorded growth measure differs from the reconstructed snapshot difference above. Its aggregation window also need not equal the spacing between snapshots. Creator quality, eligibility, and prior recommendations confound the estimate, which therefore does not identify the exposure-reinforcement exponent β. The panel describes persistent audience rankings over its observation horizon.

Cap accounting. Proportional water filling applies $\tilde{\rho}_c = \min\{\kappa, a\rho_c\}$ with a chosen so shares sum to one. Table 13 divides the resulting HHI, effective creator count, and leading-group share by their respective observed values on the matched set. The tighter cap produces a greater relative reduction in concentration. This is deterministic redistribution of aggregate exposure; candidate eligibility, relevance, and user responses would also affect a deployed policy. The calculation illustrates the aggregate lever's intended effect. The separate entropy floor is needed to specify per-user variety, which aggregate cap accounting does not evaluate.

Verification and access. The underlying audit retains parsing checks, snapshot alignment, resampling settings, and provenance in controlled records. Absolute scales and source-specific metadata are excluded from the submission package, so its relative summaries do not permit independent raw-log reproduction. Authorized access to the original records and pipeline is required for that purpose. This observational evidence motivates the problem; the joint CWOP guarantees and intervention results come from the formal model and the separately specified controlled experiments.

F Diagnostic Simulations and Robustness

These simulations isolate the two motivating mechanisms: reinforced global exposure can concentrate visibility, while within-user reinforcement can narrow individual menus. They test the scoped models of Appendix A and motivate CWOP's separate concentration and entropy constraints. Direct evaluation of CWOP's joint control is given in Section 5.2.

F.1 Supply Trajectory and Penalty Threshold

The supply simulation tests whether the regime ordering and corrective penalty remain visible when popularity and exposure are distinct states. The single-state identity $\beta_{\text{eff}} = \beta - \lambda/\tau$ is algebraic, so the simulation couples slow popularity f_c and fast exposure e_c through conversion q_{conv}. At the linear boundary $\beta = 1$, $q_{\text{conv}} = 1$, random non-uniformity persists at approximately 1.95 times the uniform HHI floor across 50 seeds, with no monopoly, matching the symmetric-Dirichlet prediction of Theorem 2(ii), $\mathbb{E}[\text{HHI}]/(1/K) = 2K/(K+1) \approx 2$. This quantitative match confirms the linear boundary as persistent non-uniformity rather than oligopoly.

Simulation protocol. The slow stock drives attachment through f_c^β, the fast flow carries the log-share penalty, and selection uses $w_c \propto f_c^\beta e_c^{-\lambda/\tau}$. Each served impression becomes durable popularity with probability q_{conv}. We use $K = 500$, $\tau = 1$, a Fenwick-tree sampler, and fixed seeds; the diagnostic contrasts use 30 to 50 seeds, while every cell in the horizon-penalty grid uses 60. We report HHI, top-1 share, and the collapse fraction, defined as the fraction of seeds whose top-1 share reaches 0.5.

At $\beta = 1.35$ and $\lambda = 0$, slower conversion reduces finite-horizon collapse, while static quality produces no collapse. Appendix Table 14 and Appendix F.4 give these contrasts, conversion accounting, and cutoff sensitivity.

Horizon-dependent transition. At $q_{\text{conv}} = 0.2$, $\beta = 1.35$, and no penalty, the top-1 ≥ 0.5 collapse fraction grows from 0.033 to 0.467 to 0.883 across 50k, 200k, and 1M steps. At 1M, it falls to 0.217, 0.033, and 0 at penalty ratios 0.3, 0.4, and 0.5 (Appendix Table 16). The ordering supports exposure-dependent intervention as a response to reinforcement in this model. The coupled sweep does not establish convergence to the single-state threshold.

F.2 Per-User Exposure Entropy

The per-user simulation tests exposure narrowing under the same two-state dynamics. At $\beta = 1.35$, normalized entropy increases from 0.247 without shaping to 0.874 at the single-state reference $\lambda^\star$; the low-entropy fraction falls from 0.967 to zero (Appendix Table 15 and Figure 11). This supports monitoring individual exposure variety alongside aggregate concentration. Entropy remains below exact balance, and the simulated outcome is served exposure rather than consumption.

F.3 Simulation Details

The supply-side two-state simulator uses a Fenwick tree sampler with $K = 500$, $\tau = 1$, $q_{\text{conv}} = 0.2$, $\beta = 1.35$, horizons $T \in \{50\text{k}, 200\text{k}, 1\text{M}\}$, and penalty ratios $\lambda/\lambda^\star \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.65, 0.8, 1\}$. Each of the 27 cells uses 60 deterministic seeds. We save HHI, Gini, top-1 share, and collapse indicators at top-1 thresholds 0.3, 0.5, and 0.7; Table 16 uses the 0.5 cutoff and Wilson 95% intervals. Relative to q_{conv}, the aggregate state ratio $\sum_c f_c / \sum_c e_c$ is approximately 1.04, 1.01, and 1.00 at the three horizons, respectively, with across-seed SD 0.0022 in the same relative units at 1M. This is an aggregate conversion accounting check, not a test of creator-level convergence. The separate linear-boundary supply run uses $\beta = 1.0$, $K = 500$, $T = 200{,}000$, and 50 seeds.

The per-user process uses $M = 50$ options and horizons $T \in \{5\text{k}, 20\text{k}, 100\text{k}\}$, matching the supply sweep's allocations per option. The super-linear grid uses the same nine penalty ratios and 60 seeds per cell. The linear-boundary comparison uses 100 seeds at $T = 100{,}000$ for $\lambda \in \{0, 0.05\}$.

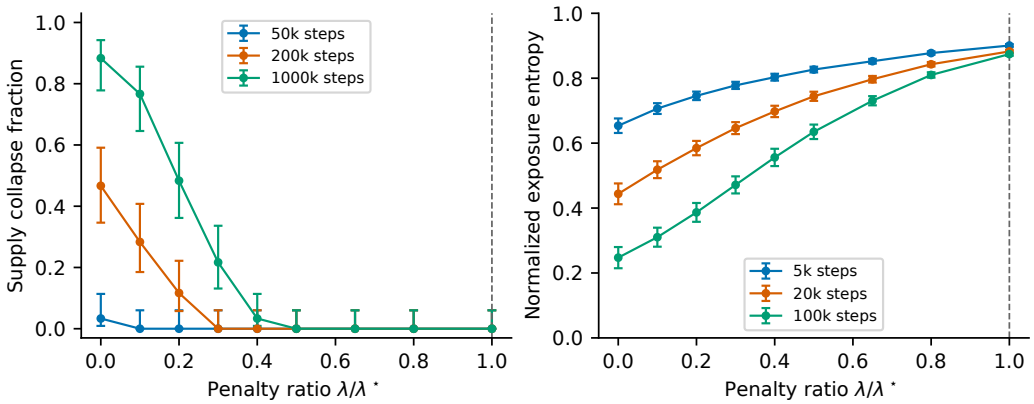

Figure 11: Controlled supply- and demand-side penalty sweeps. Left: supply collapse fraction. Right: normalized per-user entropy. The dashed line marks the single-state threshold $\lambda^\star$.

Table 14: Two-state supply simulation at $\beta = 1.35$, $\lambda = 0$, and $T = 200{,}000$, with 30 seeds per condition. Collapse denotes top-1 share at least 0.5. Seeds 0–29 differ from the 60-seed sweep starting at 2026072100 in Table 16; the repeated parameter setting therefore has different sample means. HHI is reported in units of $1/K$ (rounded); top-1 and collapse are fractions.

Mode	q_{conv}	HHI/$(1/K)$	**Top-1**	**Collapse**
Proportional	1.0	312	0.752	0.867
Proportional	0.2	160	0.488	0.400
Static quality	N/A	5.95	0.0603	0.000

Paired supply and demand view. Figure 11 places the supply-side collapse and demand-side entropy sweeps on the same normalized penalty axis at $\beta = 1.35$, $\tau = 1$, and $q_{\text{conv}} = 0.2$. Each plotted cell aggregates 60 fixed-seed runs. Horizons match allocations per option across sides ($T/K = T/M$); supply error bars are Wilson 95% intervals, and demand error bars are 95% intervals for the mean. Both responses move monotonically in the predicted direction. The dashed analytic threshold is the single-state reference for this controlled comparison.

F.4 Two-State Diagnostic Contrasts and Robustness

Separating popularity from exposure. Table 14 checks that the coupled simulator is not merely reproducing the single-state identity. At $\beta = 1.35$ and $\lambda = 0$, reducing conversion from $q_{\text{conv}} = 1$ to $q_{\text{conv}} = 0.2$ lowers finite-horizon collapse from 0.867 to 0.400. A fixed-quality baseline has HHI about 5.95 times the uniform floor and zero collapse, so static heterogeneity alone does not generate the simulated horizon drift.

Accounting and cutoff robustness. For $q_{\text{conv}} = 0.2$, the aggregate stock-to-exposure ratio approaches the specified conversion rate across horizons (approximately 1.04, 1.01, and 1.00 times q_{conv}). At $T = 1$M and $\lambda/\lambda^\star \in \{0, 0.3, 0.5\}$, collapse fractions are $(0.983, 0.600, 0.050)$ under a 0.3 cutoff and $(0.667, 0.050, 0)$ under a 0.7 cutoff. Both checks preserve the penalty ordering reported below rather than making it an artifact of conversion accounting or the body cutoff.

G Ethical Considerations and Compliance

The production audit analyzes supplied records and reports aggregate statistics. The platform's identity is withheld. No direct identifiers were used; user and creator identifiers were pseudonymized before analysis. Any future online experiment would require the applicable review and approval procedures for production A/B testing. It would be limited to content exposure and ranking mechanisms and would not alter content moderation policies, user privacy settings, or creators' control over con-

Table 15: Per-user entropy simulation at $q_{\text{conv}} = 0.2$, $M = 50$, $T = 100{,}000$, and $\tau = 1$. Low entropy denotes $H(\sigma_u)/\log M \leq 0.5$.

β	**Penalty**	**Norm. entropy**	**Low-entropy frac.**
1.00	$\lambda = 0$	0.898	0.000
1.00	$\lambda = 0.05$	0.928	0.000
1.35	$0.00\lambda^\star$	0.247	0.967
1.35	$0.50\lambda^\star$	0.635	0.100
1.35	$1.00\lambda^\star$	0.874	0.000

Table 16: Two-state sweep: HHI is divided by the uniform floor $1/K$; top-1 and collapse are fractions. Intervals are Wilson 95% intervals over 60 seeds. The 200k zero-penalty cell uses different seeds from Table 14.

T	$\lambda/\lambda^\star$	HHI/$(1/K)$	Top-1 share	Collapse	95% CI
50k	0.00	39.58	0.216	0.033	$[0.009, 0.114]$
200k	0.00	157.01	0.497	0.467	$[0.346, 0.591]$
1000k	0.00	309.34	0.746	0.883	$[0.778, 0.942]$
1000k	0.20	163.09	0.498	0.483	$[0.362, 0.607]$
1000k	0.30	101.41	0.381	0.217	$[0.131, 0.336]$
1000k	0.40	46.28	0.228	0.033	$[0.009, 0.114]$
1000k	0.50	19.01	0.126	0.000	$[0.000, 0.060]$
1000k	1.00	2.26	0.015	0.000	$[0.000, 0.060]$

tent publication. The public Last.fm experiment uses only numeric artist-listening relations from the HetRec 2011 noncommercial release, not social links, names, or tags (Cantador et al., 2011). The reported formulation does not use sensitive attributes; any deployment would alter exposure and would require fairness and ecosystem-impact monitoring.

H Discussion

Joint concentration ceilings and per-user entropy floors allow CWOP to control exposure-based information oligopoly and information cocoons within one allocation framework. The results make the associated utility cost explicit and show why finite-window serving requires safety margins in addition to feasible optimization.

I Extended Related Work

This section places CWOP within research on recommendation dynamics, fair allocation, and probabilistic constraints. These literatures contribute different parts of the problem: how recommendations change an ecosystem, which exposure distributions are desirable, and how reliably a randomized policy delivers those distributions.

Recommendation feedback and performativity. Chaney et al. (2018) show through simulation that training on algorithmically confounded recommendation data can homogenize behavior without improving utility. Perdomo et al. (2020) formalize the broader problem of performative prediction, in which deployment changes the distribution on which a predictor is evaluated, and analyze conditions for stable retraining. These results motivate treating recommendation as a sequential intervention rather than a prediction problem on a fixed dataset. CWOP addresses a different question within that sequence: conditional on the history available at the start of a window, can a selected exposure policy meet its targets during serving? Our certificate permits history-dependent policy updates between windows while requiring a known request law and conditionally independent sampling within each window. It establishes exposure safety under these assumptions, without requiring convergence to a performatively stable predictor.

Strategic creators and participation dynamics. Exposure changes creators' incentives as well as the data observed by a ranker. Eilat & Rosenfeld (2023) design recommendations that anticipate strategic content creation and encourage content diversity. Jagadeesan et al. (2023) characterize supply-side equilibria and producer specialization. Yao et al. (2023) derive welfare guarantees for top-K recommendation under competing creators, showing that strategic feedback can have bounded welfare costs under a specified behavioral model. Participation introduces another mechanism: Mladenov et al. (2020) formulate constrained matching to sustain viable providers, while Kiyohara et al. (2025) study policy design when both sides' participation evolves. CWOP instead takes the current candidate structure and utility scores as inputs. Its safety margins protect served exposure as those scores change across windows. Predicting creator entry, content adaptation, or long-run ecosystem welfare would additionally require the behavioral and population models studied in this literature.

Exposure fairness and two-sided allocation. Singh & Joachims (2018) optimize ranking utility subject to exposure-based fairness constraints, and Singh & Joachims (2019) learn stochastic ranking policies with such constraints. Biega et al. (2018) amortize individual fairness over a sequence of rankings, aligning accumulated attention with accumulated relevance. These works establish exposure allocation as a constrained decision problem. CWOP adopts that perspective but specifies aggregate HHI and maximum-share ceilings together with per-request-type entropy floors. These are structural diversity targets chosen by the platform; they do not by themselves define merit-based or demographic fairness. Do et al. (2021) use Lorenz efficiency to improve the utilities of worse-off users and producers through concave welfare maximization. Our constrained utility objective and their distributional welfare objective encode different normative choices, so meeting CWOP's targets need not imply a Lorenz-efficient outcome.

Diversity, calibration, and the object being controlled. Wang & Joachims (2021) show that user fairness, item fairness, and ranking diversity are distinct objectives that can conflict. Steck (2018) calibrate recommendations to the proportions of a user's past interests. Calibration preserves a reference distribution, whereas an entropy floor requires a minimum spread without selecting one particular distribution. CWOP measures this spread over creator identities in served exposure. Several creators can cover the same topic, and users can consume only a subset of what they receive. Thus creator exposure, topical diversity, and consumption diversity require separate measurements. This distinction determines both the interpretation of our entropy constraint and the scope of the experiments.

Fairness in expectation and in realized rankings. Randomization already plays a central role in fair ranking. Singh et al. (2021) allocate ranking opportunities according to uncertainty about merit. Gorantla et al. (2023) sample from distributions supported on rankings that satisfy group-representation constraints in every realization. CWOP's serving problem concerns a third object:

nonlinear statistics of all exposures in a finite window, including the entropy within each request type. An expected exposure plan can satisfy these constraints while a sampled window exceeds them. Our margins bound that window-level risk. This complements ex-post ranking constructions: a sampler with dependent allocations or exact quotas could have different, potentially smaller fluctuations, but would require a guarantee for that sampling mechanism.

Online allocation and learning-based controllers. FairCo combines control of amortized exposure disparities with debiased relevance estimation in dynamic learning-to-rank (Morik et al., 2020). FORM balances item and user fairness through constrained optimization and low-regret online learning (Chen et al., 2024). Offr uses online Frank–Wolfe updates to optimize smooth concave objectives over rankings, exploiting their structure for efficient allocation (Usunier et al., 2022). These methods address learning or computational challenges beyond our implemented single-item setting. Our known-score adaptations isolate utility and constraint behavior under common serving conditions; their results apply to those adaptations. CWOP's certificate is independent of the optimizer that supplies a candidate policy. Accordingly, the conic reference and PD are alternative numerical implementations, and the independent reference measures the gap between target cost and solver error.

Chance constraints and safe sequential learning. Reserving slack to control random violations is a standard idea in stochastic optimization. Nemirovski & Shapiro (2006) construct tractable convex approximations to chance-constrained programs. In constrained reinforcement learning, Liu et al. (2021) obtain high-probability zero or bounded constraint violation using safe-policy assumptions and optimistic or pessimistic learning. CWOP conditions on the current request distribution and selected policy instead of estimating unknown transitions. Its certificate specializes constraint tightening to the joint behavior of share caps, HHI, and empirical entropy, accounting for random request-type counts and adaptive choices across windows. The fixed failure budget covers the prescribed collection of windows. Extending that guarantee to an unknown request law, misspecified response estimates, or unrestricted within-window adaptation requires additional uncertainty bounds.

Convex optimization and the cost of constraints. The request-policy formulation uses standard convex analysis: linear expected utility, convex concentration ceilings, and superlevel sets of concave entropy. Boyd & Vandenberghe (2004) provide the foundations for duality, Slater conditions, and barrier methods. Nedić & Ozdaglar (2009) analyze primal averaging and infeasibility bounds for dual subgradient methods. These results support our numerical convergence analysis and multiplier certificates. The feasible witness also supplies a concrete way to repair an approximate policy and reserve additional slack. Our cost decomposition separates the loss imposed by the original exposure targets from numerical error and the further loss needed for random serving. The uniform-envelope argument connects this mixture construction to policy-dependent certification, giving a utility lower bound even though a policy-dependent certified set need not be convex.

Concentration, entropy, and reinforcement foundations. Variance-sensitive and bounded-difference inequalities are classical tools for functions of independent random variables (Boucheron et al., 2013). Entropy, relative entropy, and the method of types are developed systematically by Cover & Thomas (2006); finite-sample entropy estimation and bias are studied by Paninski (2003). We combine these tools using the variance of a proposed exposure policy and the decomposition of plug-in entropy error into a linear term and relative entropy. Conditioning at the start of each window preserves the required sampling assumptions. The local cap-margin necessity result uses classical central limit reasoning (Durrett, 2019) and establishes the $N^{-1/2}$ order for a nondegenerate cap boundary, rather than tightness of every margin constant. Separately, Pemantle (2007) surveys reinforced processes underlying the motivating urn examples. Those examples describe path dependence; the serving theorem applies without identifying a reinforcement exponent.

Identification and empirical interpretation. Schnabel et al. (2016) develop methods for learning and evaluation under selection-biased recommendation data. Their perspective explains why descriptive production concentration cannot identify the effect of deploying a new exposure policy. Our audit motivates the target statistics, while controlled feedback experiments and Last.fm-derived preference proxies test optimization and sampling safety. Taken together, the contribution is a quantified connection between a specified exposure plan, its utility cost, and finite-window delivery.