FEATURE DIVERSITY IS NOT MONOTONE IN CONTENT EXPLORATION: EVIDENCE FROM CO-ADAPTING USER–CREATOR RECOMMENDER SYSTEMS

Anonymous authors
Paper under double-blind review

ABSTRACT

Recommender systems influence both users, whose preferences can change with the content they see, and creators, who may adapt their content to the audiences they reach. We study how exploration affects recommendation diversity, measured by the variance of content features recommended to each user, while users remain influenced by their underlying interests and creators by their established specializations. We establish sufficient conditions under which intermediate exploration yields higher long-run recommendation diversity than both uniform recommendation, which splits each user's exposure equally across creators, and rates close to pure exploitation, which concentrates exposure on each user's highest-scoring creator. We also establish conditions for convergence to a unique equilibrium when exploration is sufficiently close to uniform. Simulations with synthetic vectors and representations learned from MovieLens show that intermediate exploration can improve recommendation diversity, with the benefit varying across input representations and prescribed user and creator response strengths.

1 INTRODUCTION

Recommender systems help users find relevant content in large catalogs by drawing on information such as their past interactions (Ziegler et al., 2005). Their influence extends beyond individual choices, however, because repeatedly recommending similar content can limit the variety users encounter, raising concerns about increasingly similar consumption and opinion clustering (Chaney et al., 2018; Schüepp et al., 2026). We study how to preserve the diversity of recommendations as users' preferences and creators' content change in response to the system.

Diverse recommendation lists can expose users to different topics and perspectives and broaden visibility for otherwise overlooked voices (Schelenz, 2021). Creating such lists involves balancing relevance with variety (Ziegler et al., 2005; Cheng et al., 2017), but recommendations also determine which audiences creators reach and can influence what they produce. Broad exposure today can therefore offer little variety tomorrow if creators come to produce similar content.

In controlled experiments, Adomavicius et al. (2013) show that displayed recommendation ratings influence users' subsequent evaluations. Through a qualitative analysis of Instagram influencers' discussions, Cotter (2019) documents how they interpret platform algorithms and develop strategies for visibility. Dean et al. (2025) advocate mathematical models that account for these user and creator responses when studying how algorithmic choices shape preferences and available content over time.

Theoretical work connects exploration to the content creators produce: with user preferences held fixed, Hron et al. (2023) show that more exploration in a softmax exposure rule can encourage broadly appealing content, while less exploration can support creator specialization. When both users and creators adapt, Lin et al. (2024) establish conditions for feature polarization, while Schüepp et al. (2026) also include social influence among users and design recommendations to mitigate opinion clusterization. Building on these results, we ask: how does exploration affect the long-run diversity of recommendations when both users and creators adapt?

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.

To study this question, we place users and creators in a shared feature space, following models of mutual influence (Lin et al., 2024; Schüepp et al., 2026). Users adjust toward the content they receive and creators toward the preferences of their audiences, while fixed reference vectors, called anchors, retain the influence of underlying interests and established specializations. The platform balances exploration and exploitation through an ε-greedy policy: a fraction $1 - \varepsilon$ of each user's recommendations goes to their highest-scoring creator, and the remaining fraction ε is shared equally across all creators. Following Lin et al. (2024), we measure recommendation diversity (RD) as the exposure-weighted variance of content features within each user's recommendations, averaged over users. This measures the variety of features users receive, which is distinct from the number of creators they encounter or the fairness of exposure.

Exploration changes both the content a user encounters at a given moment and the states that users and creators reach over time. To distinguish these effects, we express recommendation diversity in terms of how varied the creator vectors are and how the policy weights those vectors for each user. The resulting decomposition lets us identify when a small reduction from uniform recommendation increases long-run RD and compare that improvement with the low RD obtained near pure exploitation.

For exploration sufficiently close to uniform, we give conditions under which every initialization converges to the same equilibrium of user preferences and creator content. As exploration varies near uniform recommendation, we separate the effect of reallocating exposure at fixed states from the change in creator content. This decomposition yields conditions for an intermediate exploration rate to improve long-run RD over both uniform recommendation and rates close to pure exploitation.

We examine the size of this improvement using synthetic vectors and MovieLens representations, varying user and creator response strengths across five synthetic ensembles and five MovieLens fits. Synthetic inputs give intermediate RD maxima in all four response settings, while MovieLens gives them in three. When MovieLens users respond strongly and creators weakly, uniform recommendation instead gives the highest tested RD.

2 Related Work

Classic diversification methods balance relevance and variety within a recommendation list: Ziegler et al. (2005) study topic diversification and user satisfaction, while Cheng et al. (2017) learn to recommend accurate and diverse items jointly. Both approaches take a static view, constructing lists from observed preferences and an available catalog without modeling how recommendations subsequently change preferences or creator content. At the catalog level, diversification also connects to exposure fairness through the representation of less visible items and providers (Zhao et al., 2025).

In simulations, training on data influenced by earlier recommendations can homogenize consumption (Chaney et al., 2018). Models of preference dynamics examine a further effect: exposure can change the preferences that the system seeks to match (Dean & Morgenstern, 2022; Jiang et al., 2019). These feedback effects connect to performative prediction, where deploying a predictor changes the data it subsequently encounters (Perdomo et al., 2020). Brown et al. (2022) extend this perspective by making the population's response depend also on its current state.

On the supply side, Jagadeesan et al. (2023) study how competition for recommendations shapes creator specialization, while Hron et al. (2023) show how exploration changes the incentives for broadly appealing content. Such responses make available content partly an outcome of the policy itself. Eilat & Rosenfeld (2023) use this link to encourage diverse content through strategic incentives, whereas Yao et al. (2024) study how exploration balances user satisfaction against the quantity of content produced within fixed creator niches.

Exposure can also change who remains on the platform: Mladenov et al. (2020) study provider viability, and Kiyohara et al. (2025) model the coupled participation of viewers and providers. Our populations remain fixed; we instead let user preferences and creator features evolve together to study the variety of content each user receives.

Lin et al. (2024) establish conditions for feature polarization under mutual user–creator influence, showing that common diversity-promoting strategies can fail while selection among the top-k creators can preserve creator diversity. We follow their distinction between recommendation rele-

vance, creator diversity, and recommendation diversity because preserving differences across available creators alone does not ensure variety within each user's recommendations. Following their mutual-influence perspective, we introduce fixed reference vectors that preserve individual influences throughout the dynamics and are absent from their base updates.

Models of opinion dynamics combine social influence with attachment to individual views (Friedkin & Johnsen, 1990), with extensions to interactions between opinions on different topics (Parsegov et al., 2017) and recommendation-based control (Sprenger et al., 2024). Schüepp et al. (2026) apply this principle to users and creators, retaining their initial opinions as persistent inputs and using the users' social network to design recommendations that mitigate opinion clustering. Our model allows persistent anchors to differ from initial states and mediates interactions through an ε-greedy policy, without direct influence between users.

3 A Co-Adaptive Recommendation Framework

We consider $M \geq 1$ users and $N \geq 2$ creators in a shared feature space $\mathbb{R}^d$, with $d \geq 1$. We write $[K] = \{1, \ldots, K\}$ and index users by $j \in [M]$ and creators by $i \in [N]$. At round $t = 0, 1, \ldots$, $\mathbf{u}_j(t)$ represents user j's preferences and $\mathbf{c}_i(t)$ represents creator i's content. We model the mutual influence studied by Lin et al. (2024) using weighted-average updates with persistent individual influences (Friedkin & Johnsen, 1990; Schüepp et al., 2026). Users and creators update simultaneously as

$$\mathbf{u}_j(t+1) = (1 - \kappa_j - \eta_j)\mathbf{u}_j(t) + \kappa_j \mathbf{a}_j + \eta_j \sum_i p_{ij}^{\varepsilon}(t)\mathbf{c}_i(t), \tag{1}$$

$$\mathbf{c}_i(t+1) = (1 - \lambda_i - \xi_i)\mathbf{c}_i(t) + \lambda_i \mathbf{b}_i + \xi_i \frac{\sum_j p_{ij}^{\varepsilon}(t)\mathbf{u}_j(t)}{\sum_j p_{ij}^{\varepsilon}(t)}. \tag{2}$$

The recommendation fraction $p_{ij}^{\varepsilon}(t)$ is the share of user j's recommendations assigned to creator i, with exploration controlled by $\varepsilon \in (0, 1]$. Both updates use the states at time t. The initial states represent the information available to the recommender from prior interactions and lie in the unit ball $\mathbb{B} = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_2 \leq 1\}$.

Each user balances their previous state against persistent interests and the content they receive. Persistent interests are represented by the fixed preference anchor $\mathbf{a}_j \in \mathbb{B}$, which enters with weight $\kappa_j \in (0, 1]$, while the exposure-weighted mean of recommended content enters with response weight $\eta_j \in [0, 1 - \kappa_j]$. The remaining weight $1 - \kappa_j - \eta_j$ retains the previous state.

Creators likewise balance established specializations against the preferences of their audiences. The fixed anchor $\mathbf{b}_i \in \mathbb{B}$ enters with weight $\lambda_i \in (0, 1]$, and the audience's mean preference vector enters with response weight $\xi_i \in [0, 1 - \lambda_i]$; the remaining weight $1 - \lambda_i - \xi_i$ retains the previous content. Dividing by total exposure to creator i makes the audience term an average weighted by users' relative exposure. The audience vector therefore depends on the audience's composition, while ξ_i controls its weight independently of audience size. At $\xi_i = 0$, content moves toward the anchor independently of recommendations.

Because the update weights are nonnegative and sum to one, both sides remain in $\mathbb{B}$. The anchors continue to influence this evolution at every round, but may differ from the initial states, allowing us to compare different starting representations of the same persistent interests and specializations. Figure 1 shows how these anchored responses connect users and creators in a feedback loop.

The platform scores creator i for user j by $\mathbf{u}_j(t)^\top \mathbf{c}_i(t)$ (Koren et al., 2009). This dot-product score is also used in two-tower recommenders, which learn separate user and item representations in a shared space (Yang et al., 2020). Let $h_j(t)$ denote the index of the highest-scoring creator, breaking ties by the smallest index. The recommendation fractions $p_{ij}^{\varepsilon}(t)$ are set by an ε-greedy policy (Sutton & Barto, 2018, Section 2.2):

$$h_j(t) = \min \arg\max_{i \in [N]} \mathbf{u}_j(t)^\top \mathbf{c}_i(t), \qquad p_{ij}^{\varepsilon}(t) = (1 - \varepsilon)\mathbf{1}\{i = h_j(t)\} + \frac{\varepsilon}{N}. \tag{3}$$

This policy balances exploitation, which directs a fraction $1 - \varepsilon$ to the highest-scoring creator, with exploration, which shares the remaining fraction ε equally across all creators. Thus $\varepsilon = 1$ gives

uniform recommendation, while the policy approaches pure exploitation as ε decreases to zero. Each creator receives total exposure of at least $M\varepsilon/N > 0$, so the audience mean in (2) is well-defined. We call the users' highest-scoring choices their user–creator *assignment*. The updates use the exposure fractions directly, giving deterministic trajectories for comparing long-run outcomes across exploration rates. Appendix A discusses these modeling choices further.

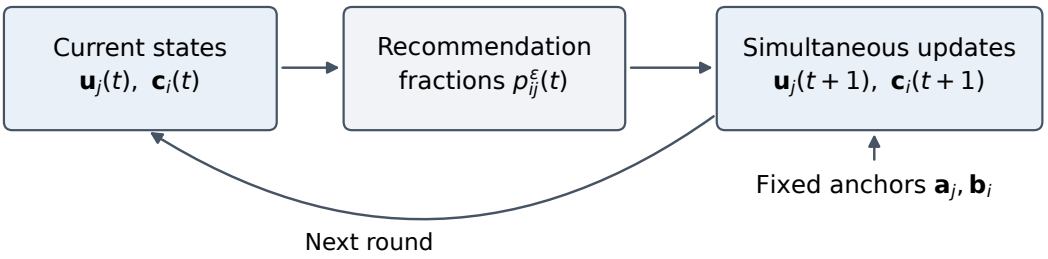


Figure 1: One round of the feedback loop. Current states determine recommendation fractions, which enter simultaneous updates alongside states at time t and fixed anchors. Users respond to content and creators to their audiences, producing the states that determine the next round.

We evaluate how recommendations match user preferences and how much variety they provide, using three metrics following Lin et al. (2024). Recommendation Relevance (RR) measures the match between users' current preferences and the content they receive; Creator Diversity (CD) measures differences across all creators; and Recommendation Diversity (RD) measures how much the content received by each user varies, averaged across users. At a given time, suppress t on the vectors, abbreviate $p_{ij}^{\varepsilon}(t)$ as p_{ij}, and write $\boldsymbol{\mu}_j = \sum_i p_{ij}\mathbf{c}_i$ for the mean content recommended to user j. Then

$$\mathrm{RR}(t) = \frac{1}{M}\sum_j \mathbf{u}_j^\top \boldsymbol{\mu}_j, \qquad \mathrm{CD}(t) = \frac{1}{N(N-1)}\sum_{i\neq k} \|\mathbf{c}_i - \mathbf{c}_k\|_2,$$
$$\mathrm{RD}(t) = \frac{1}{M}\sum_j \sum_i p_{ij}\|\mathbf{c}_i - \boldsymbol{\mu}_j\|_2^2. \tag{4}$$

The unit-ball assumption gives $\mathrm{RR} \in [-1, 1]$, $\mathrm{CD} \in [0, 2]$, and $\mathrm{RD} \in [0, 1]$. We focus on RD because our question concerns the variety received by each user after adaptation. CD describes the available content but does not account for what each user is shown. At the pure-exploitation boundary $\varepsilon = 0$, each user receives only one creator's content, so RD is zero even when distinct creator vectors give positive CD. At small positive exploration rates, RD can likewise be small despite substantial CD.

4 Equilibria and Recommendation Diversity

The policy in (3) balances exploitation of the highest-scoring creator with exploration across all creators. Changing this balance also changes the audiences to which creators adapt and the content available to recommend. We first ask when the system converges to the same equilibrium from every initial state, making long-run policy comparisons independent of initialization. We then identify conditions under which intermediate exploration gives more recommendation diversity than either extreme. Proofs and the construction of candidate equilibria are in Appendix B.

4.1 Uniqueness of long-run outcomes

At an equilibrium, user preferences and creator content no longer change. Establishing convergence requires accounting for changes in each user's highest-scoring creator as these states evolve under the policy in (3).

Under uniform recommendation, every user receives content from all creators in equal proportions, so the updates no longer depend on the scores. With nonzero anchor weights, the system converges to a unique equilibrium from every initial state, including when scores are tied (Lemma B.1). We

collect all user and creator states in $\mathbf{z} = (\mathbf{u}_1, \ldots, \mathbf{u}_M, \mathbf{c}_1, \ldots, \mathbf{c}_N)$ and write this uniform equilibrium as $\mathbf{z}^*_{\text{unif}}$. Here $*$ denotes an equilibrium and "unif" denotes $\varepsilon = 1$. Although scores do not affect uniform exposure, they determine where exploitation weight goes when ε decreases below one. Writing $h_j(\mathbf{z})$ for user j's choice at state $\mathbf{z}$, we set $g_j^{\text{unif}} := h_j(\mathbf{z}^*_{\text{unif}})$ and collect these choices in $G_{\text{unif}} = (g_1^{\text{unif}}, \ldots, g_M^{\text{unif}})$. Its smallest score gap is

$$\gamma := \min_j \min_{k \neq g_j^{\text{unif}}} (\mathbf{u}^*_{j,\text{unif}})^\top \big(\mathbf{c}^*_{g_j^{\text{unif}},\text{unif}} - \mathbf{c}^*_{k,\text{unif}}\big). \tag{5}$$

When $\gamma > 0$, each user has a unique highest-scoring creator at the uniform equilibrium, and sufficiently small changes in the states preserve these choices. The tie-breaking rule in (3) determines a choice at an exact tie, but does not ensure the same choice at nearby states. The following theorem uses the gap to establish convergence when exploration is sufficiently close to uniform. Let $\underline{\alpha} = \min\{\min_j \kappa_j, \min_i \lambda_i\} > 0$ be the smallest anchor weight, and let $\eta_{\max} = \max_j \eta_j$ and $\xi_{\max} = \max_i \xi_i$ be the largest user and creator response weights.

Theorem 4.1 (Global convergence under sufficient exploration)**.** *Suppose $\gamma > 0$. For any fixed $\varepsilon \in (0, 1]$ satisfying*

$$\max\left\{\eta_{\max}(1-\varepsilon),\, \xi_{\max}\frac{N(1-\varepsilon)}{N(1-\varepsilon)+\varepsilon}\right\} < \frac{\underline{\alpha}\gamma}{16}, \tag{6}$$

the system (1)–(3) *has a unique equilibrium, and every trajectory starting from an admissible initial state converges geometrically to it. After finitely many updates, each user's highest-scoring creator remains g_j^{unif}; at equilibrium, every user's highest-scoring creator is unique.*

Condition (6) compares the perturbation caused by nonuniform exposure with the anchoring strength and score gap that protect the uniform assignment. As ε approaches one, the perturbation terms on the left vanish while the right-hand side stays positive, so the condition always holds sufficiently close to uniform exploration when $\gamma > 0$.

This convergence occurs because every trajectory eventually enters and remains in a neighborhood of the uniform equilibrium in which the score gap prevents further changes of assignment. Although user and creator states may continue to evolve, their recommendation fractions remain fixed, and Lemma B.1 reduces the error to a common equilibrium by a factor at most $1 - \underline{\alpha}$ per update. Appendix B.2 gives the full argument.

4.2 When intermediate exploration improves diversity

Exploration can change recommendation diversity both by reallocating exposure across existing content and by changing that content through adaptation. To separate these effects, let $\text{RD}(\mathbf{z}, \varepsilon)$ denote the metric in (4) under policy (3), and first hold the state $\mathbf{z}$ fixed while varying ε. With $\overline{\mathbf{c}} = N^{-1}\sum_i \mathbf{c}_i$ denoting the mean creator vector, we compare the spread of all creators around this mean, V_{all}, with the mean squared distance of each user's highest-scoring creator from the same reference, V_{top}:

$$V_{\text{all}}(\mathbf{z}) := \frac{1}{N}\sum_i \|\mathbf{c}_i - \overline{\mathbf{c}}\|_2^2, \qquad V_{\text{top}}(\mathbf{z}) := \frac{1}{M}\sum_j \|\mathbf{c}_{h_j(\mathbf{z})} - \overline{\mathbf{c}}\|_2^2. \tag{7}$$

Each user's recommendations mix their highest-scoring creator with uniform exposure to all creators, giving the personalized mean $\boldsymbol{\mu}_j = (1-\varepsilon)\mathbf{c}_{h_j(\mathbf{z})} + \varepsilon\overline{\mathbf{c}}$. Using this mixture in (4) yields the exact identity

$$\text{RD}(\mathbf{z}, \varepsilon) = \varepsilon V_{\text{all}}(\mathbf{z}) + \varepsilon(1-\varepsilon)V_{\text{top}}(\mathbf{z}). \tag{8}$$

The second term comes from mixing each user's highest-scoring creator with the broader creator population. At fixed states, RD has a strict interior maximum exactly when $V_{\text{top}} > V_{\text{all}}$: selected creators have a larger mean squared distance from the creator mean than the full creator population. When creators adapt, exploration also changes the vectors being mixed, so we next examine both effects at equilibrium.

Under Theorem 4.1, write $\mathbf{z}^*(\varepsilon)$ for the unique equilibrium near uniform exploration, so $\mathbf{z}^*(1) = \mathbf{z}^*_{\text{unif}}$, and define

$$\text{RD}^*(\varepsilon) := \text{RD}(\mathbf{z}^*(\varepsilon), \varepsilon). \tag{9}$$

The next proposition separates the derivative of equilibrium RD into the effect of changing recommendation fractions and the effect of changing creator content as the states adapt. The derivative at 1^- is taken from below: if it is negative, reducing ε slightly below one increases long-run RD. In the statement, $\underline{\alpha}$ is the smallest anchor weight and γ is the minimum score gap at the uniform equilibrium, defined in (5).

Proposition 4.2 (Two effects on diversity near uniform recommendation)**.** *If $\underline{\alpha} > 0$ and $\gamma > 0$, then*

$$\left.\frac{d}{d\varepsilon}\mathrm{RD}^*(\varepsilon)\right|_{\varepsilon=1^-} = \underbrace{V_{\mathrm{all}}(\mathbf{z}^*_{\mathrm{unif}}) - V_{\mathrm{top}}(\mathbf{z}^*_{\mathrm{unif}})}_{\textit{direct allocation effect}} + \underbrace{\left.\frac{d}{d\varepsilon}V_{\mathrm{all}}(\mathbf{z}^*(\varepsilon))\right|_{\varepsilon=1^-}}_{\textit{creator-adaptation effect}}. \tag{10}$$

If the derivative in (10) *is negative, every exploration rate sufficiently close to and below one gives greater long-run RD than uniform recommendation from every initialization.*

The allocation term in (10) describes how RD changes when exposure is redistributed across fixed user and creator vectors. The adaptation term accounts for the accompanying change in the spread of creator content, including feedback from user preferences as the joint equilibrium changes with exploration. If this term is negative, increasing ε toward one makes creator vectors less dispersed; whether reducing exploration improves RD depends on its sum with the allocation term. Appendix B.3 derives both contributions for heterogeneous response weights.

We now compare this improvement near uniform recommendation with the low-exploration regime. As exploration tends to zero, RD tends to zero uniformly over all admissible states, whether or not trajectories converge. Combining this bound with a negative derivative in (10) gives the following comparison with both limits.

Theorem 4.3 (Intermediate exploration improves on both limits)**.** *Suppose $\underline{\alpha} > 0$, $\gamma > 0$, and*

$$\left.\frac{d}{d\varepsilon}\mathrm{RD}^*(\varepsilon)\right|_{\varepsilon=1^-} < 0.$$

Then there exist $0 < \varepsilon_- < \varepsilon_+ < 1$ such that the system at ε_+ has a unique globally attracting equilibrium and

$$\mathrm{RD}^*(\varepsilon_+) > \mathrm{RD}^*(1), \qquad \mathrm{RD}(\mathbf{z}, \varepsilon) < \mathrm{RD}^*(\varepsilon_+) \quad (\mathbf{z} \in \mathbb{B}^{M+N},\ 0 < \varepsilon \le \varepsilon_-). \tag{11}$$

The rate ε_+ gives higher equilibrium RD than $\varepsilon = 1$, and higher RD than the ε-greedy policy at every rate $0 < \varepsilon \le \varepsilon_-$, whatever the state of that low-exploration system. To understand the second comparison, the policy in (3) assigns total weight $\varepsilon(N-1)/N$ to creators other than $h_j(\mathbf{z})$. Their squared distances from $\mathbf{c}_{h_j(\mathbf{z})}$ are at most four in the unit ball. The variance around the recommended mean is no larger than the mean squared distance from this selected creator, giving

$$0 \le \mathrm{RD}(\mathbf{z}, \varepsilon) \le 4\varepsilon\frac{N-1}{N}. \tag{12}$$

Thus RD tends to zero as ε decreases to zero, uniformly over all admissible states. At $\varepsilon = 0$, each user receives only their highest-scoring creator and RD is zero. The dynamic comparison uses $\varepsilon > 0$ so every creator has an audience and its update is defined. The theorem establishes an intermediate improvement without identifying the best rate. Appendix B.4 gives the proof and a linear system for computing equilibria with fixed highest-scoring choices.

A homogeneous case helps explain the role of creator adaptation. Under uniform recommendation, creators with common weights λ, ξ all respond to the same audience, so their equilibrium differences satisfy

$$\mathbf{c}^*_{i,\mathrm{unif}} - \mathbf{c}^*_{k,\mathrm{unif}} = \frac{\lambda}{\lambda + \xi}(\mathbf{b}_i - \mathbf{b}_k).$$

Under uniform recommendation, differences between creators therefore come from their anchors and shrink as response strengthens relative to anchoring. Changing exploration affects a different part of this balance: when only some users select a creator, reducing the uniform component increases their relative weight in its audience. Corollary B.3 shows that, when users share a common ratio η_j/κ_j and creators share a common ratio ξ_i/λ_i, the creator-adaptation derivative with respect to ε at 1^- is nonpositive. With a positive score gap, the contribution is strictly negative if creators respond to their audiences and at least two users select different creators. The common-ratio assumption provides this sign guarantee; heterogeneous settings, including our experiments, use the general derivative formula.

5 SYNTHETIC AND MOVIELENS-INFORMED EXPERIMENTS

We compare long-run RD across exploration rates using synthetic vectors and representations that we fit to MovieLens 1M ratings (Harper & Konstan, 2015) with a matrix-factorization model (Koren et al., 2009). Each population contains $M = 256$ users and $N = 80$ creators in dimension $d = 16$, with initial states and anchors in the unit ball. For synthetic inputs, these vectors have independently sampled isotropic directions and radii uniform on $[0, 1]$. In the MovieLens setting, movies serve as creator proxies: representations learned from ratings supply the geometry, while the subsequent adaptation is simulated. Fits to recent ratings provide the initial states, and fits to the longer history at the same cutoff provide the anchors. Because the two fits can have different coordinate orientations, we align the recent vectors with the historical ones using a common orthogonal transformation that preserves their norms and dot products; neither set is normalized to unit length (Appendix C).

The four regimes in Table 1 vary how strongly users and creators respond relative to their anchors. Coefficients differ across agents but satisfy $\kappa_j + \eta_j = \lambda_i + \xi_i = 0.6$; they are prescribed simulation settings rather than estimates of MovieLens behavior. We test 13 exploration rates from $\varepsilon = 0.01$ to 1 on ten populations, formed from independently sampled subsets of users and movies for MovieLens and independent vector draws for synthetic inputs. Reusing each population's initial states, anchors, and coefficients across rates makes these comparisons paired. The primary results use one MovieLens fit and one synthetic ensemble, with robustness checks across five fitting seeds and five ensembles (Figure 7). Long-run metrics are taken at numerically verified fixed points or averaged over one period when the outcome is periodic (Appendix C.3).

Table 1: Response regimes. Each pair gives the distribution medians of the anchor and response weights, respectively. Individual coefficients vary; their sum is 0.6, leaving weight 0.4 on the previous state.

Regime	Users (κ_j, η_j)	Creators (λ_i, ξ_i)	Interpretation
Weak response	$(0.51, 0.09)$	$(0.51, 0.09)$	Both follow their anchors closely.
Creator response	$(0.57, 0.03)$	$(0.06, 0.54)$	Creators respond strongly; users weakly.
User response	$(0.09, 0.51)$	$(0.57, 0.03)$	Users respond strongly; creators weakly.
Coupled response	$(0.21, 0.39)$	$(0.21, 0.39)$	Both respond strongly.

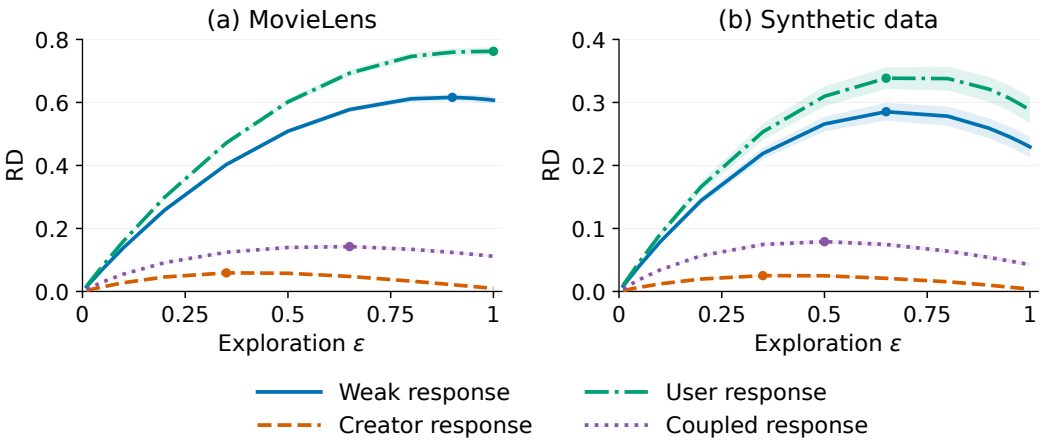


Figure 2: Long-run recommendation diversity (RD) versus exploration for MovieLens (left) and synthetic vectors (right), across four response regimes. Lines connect means over ten populations at 13 rates ($0.01 \leq \varepsilon \leq 1$), with pointwise 95% Student-t bands and dots marking the grid maxima. Periodic outcomes enter these means through their period-average RD, and the two panels use different vertical scales.

We use *RD gain* to mean the difference between population-mean RD at a given exploration rate and its value under uniform recommendation. Figure 2 shows intermediate RD maxima in three MovieLens regimes and all four synthetic regimes. With MovieLens and strong creator response, mean RD reaches 0.0590 at $\varepsilon = 0.35$, compared with 0.0102 under uniform recommendation. Under coupled response it reaches 0.1424 at $\varepsilon = 0.65$, compared with 0.1118 at uniform. The increase over uniform RD is smaller under weak response, 0.0089, while user response gives its highest tested RD, 0.7623, at uniform recommendation. Synthetic vectors give an intermediate improvement in every regime, including user response: RD reaches 0.3385 at $\varepsilon = 0.65$, compared with 0.2878 at uniform recommendation. The size of the improvement therefore depends on both the response strengths and the input vectors. Table 2 reports all maxima, selected grid rates, and paired gain intervals adjusted for all nonuniform comparisons in the two primary experiments. The intervals in the figures are pointwise, not adjusted for selecting the maximum.

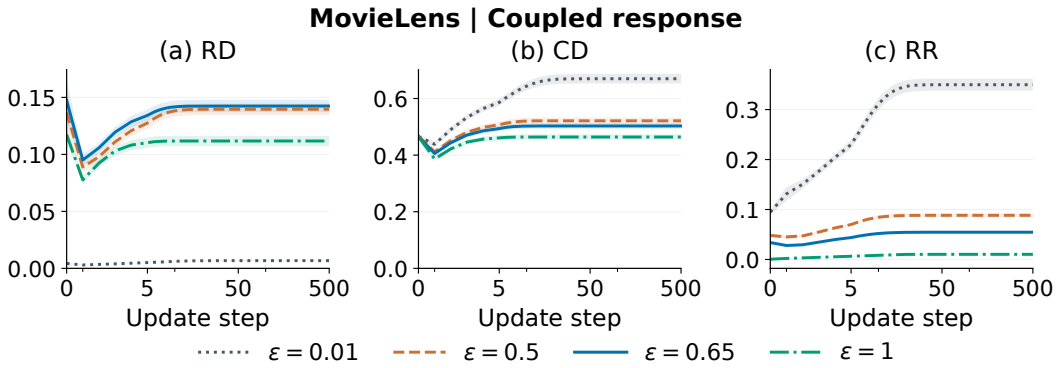


Figure 3: Evolution of recommendation diversity (RD), creator diversity (CD), and recommendation relevance (RR) through 500 updates in the MovieLens simulations under coupled response. Lines show means over ten populations, with pointwise 95% Student-t bands. The time axis is linear through step 5 and logarithmic thereafter.

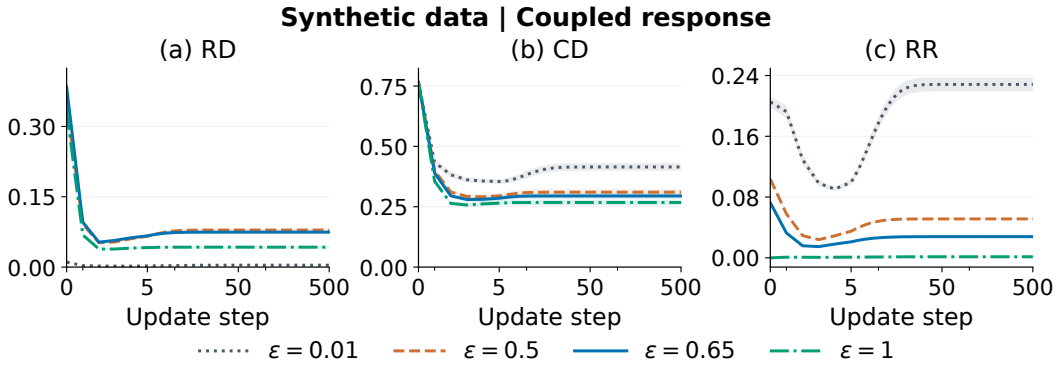


Figure 4: Evolution of recommendation diversity (RD), creator diversity (CD), and recommendation relevance (RR) through 500 updates in the synthetic simulations under coupled response. Lines show means over ten populations, with pointwise 95% Student-t bands. The time axis is linear through step 5 and logarithmic thereafter, while vertical scales differ from Figure 3.

Figures 3 and 4 follow RD, CD, and RR through 500 updates under coupled response. At the end of both experiments, RD at the intermediate rates exceeds RD under uniform recommendation. Near-pure exploitation, however, ends with the largest CD and RR among the four illustrated rates but the smallest RD in both protocols. The comparison illustrates why varied available content does not by itself guarantee variety within each user's recommendations. Appendix C.5 extends the metric comparison to every regime (Figure 6) and complements it with two-coordinate views of the initial and final states (Figure 8).

To assess how changing states affects the RD gain, we also hold both user and creator vectors fixed at their equilibrium under uniform recommendation and vary only ε. Figure 5 compares the largest dynamic gain on the tested grid with the largest fixed-state gain over the entire interval $[0, 1]$, using the same uniform baseline. With MovieLens, the creator-response gain is 0.0487 for the dynamics versus 0.0014 for the fixed states; the coupled-response gains are 0.0307 and 0.0057. The dynamic gain exceeds the largest fixed-state gain in six of the eight settings. Synthetic user response gives the opposite comparison, 0.0507 versus 0.0580, while MovieLens user response gives no gain in either case. Fixed states also give an intermediate maximum in all seven settings where the dynamic mean has one. Reallocation alone can therefore generate an intermediate maximum in these settings; joint adaptation changes the gain. Appendix C.4 gives the construction and the calculation of the fixed-state maximum.

Across the five MovieLens fits and five synthetic ensembles, 5,193 of 5,200 trajectories reach numerically verified fixed points and seven synthetic trajectories reach short periodic cycles. The primary experiments contain one such cycle, under creator response at $\varepsilon = 0.02$. Appendix C.3 describes the stopping and verification rules. The mean uniform-equilibrium derivative in Table 3 is negative in the seven settings with an intermediate grid maximum and positive for MovieLens user response. These derivatives describe local changes near uniform exposure, not the height or location of the grid maxima. The sufficient condition in Theorem 4.1 guarantees convergence only in narrow neighborhoods of $\varepsilon = 1$ for these experiments. All nonuniform grid rates lie outside those neighborhoods, so their reported convergence is supported by the numerical checks above; Appendix C.3 gives the neighborhood widths.

6 Conclusion

When users and creators adapt to one another, increasing exploration changes both how exposure is allocated and the content available for recommendation. In the anchored model studied here, intermediate exploration can yield higher long-run RD than both uniform recommendation and rates close to pure exploitation. Our analysis gives sufficient conditions for this improvement and for convergence from every initialization to a unique equilibrium when exploration is sufficiently close to uniform.

The simulations show how this benefit depends on feature geometry and response strengths. Intermediate exploration improves RD in both protocols when creators respond strongly, including when users also respond strongly. The contrast appears when users respond strongly and creators weakly: synthetic vectors still give an intermediate maximum, whereas MovieLens vectors give their highest tested RD at uniform exposure. These comparisons characterize feature diversity within the simulated response model; connecting the results to observed user welfare or creator behavior remains an empirical question.

Our policy adapts its recommendations to the current states while keeping the exploration rate constant. A natural next step is to let this rate vary over time and study whether such policies can achieve greater long-run recommendation diversity than the best constant rate, together with conditions under which the resulting dynamics converge.

Reproducibility statement

The supplementary source package includes the experiment configuration, training and simulation code, tests, fitted representations, population states, individual outcomes, and scripts that regenerate every figure and numeric table. MovieLens ratings are downloaded separately with checksum verification. Appendix C specifies data preparation, optimization, randomization, stopping rules, and statistical summaries; Appendix B.1 and the following proof sections provide the mathematical arguments.

AI use statement

We used generative AI tools to support literature search and discovery, draft and revise sections of the manuscript, improve readability, and assist with LaTeX formatting and references. We also

used these tools to discuss modeling choices and hypotheses, help formulate mathematical claims, develop and write proofs, and examine the consistency of derivations and their interpretation.

AI assistance supported the design and refinement of experiments, the implementation and revision of data-processing, training, simulation, and evaluation code, the preparation of figures and tables, and the interpretation of numerical results. It also supported the development of code for generating synthetic simulation inputs, including heterogeneous adaptation coefficients, initial states, and anchors, and for computing simulated trajectories. The MovieLens ratings used to learn initial representations and anchors were obtained from the existing dataset.

AI-assisted code was evaluated using automated tests and numerical consistency checks, and bibliographic references were checked against original sources. These checks do not replace author review. The authors take responsibility for the final manuscript, including its mathematical claims, proofs, experimental methodology, code, results, and interpretations.

REFERENCES

Gediminas Adomavicius, Jesse C. Bockstedt, Shawn P. Curley, and Jingjing Zhang. Do recommender systems manipulate consumer preferences? a study of anchoring effects. *Information Systems Research*, 24(4):956–975, 2013. doi:10.1287/isre.2013.0497.

Gavin Brown, Shlomi Hod, and Iden Kalemaj. Performative prediction in a stateful world. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 6045–6061. PMLR, 2022. URL https://proceedings.mlr.press/v151/brown22a.html.

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. Association for Computing Machinery, 2018. doi:10.1145/3240323.3240370.

Peizhe Cheng, Shuaiqiang Wang, Jun Ma, Jiankai Sun, and Hui Xiong. Learning to recommend accurate and diverse items. In *Proceedings of the 26th International Conference on World Wide Web*, pp. 183–192, 2017. doi:10.1145/3038912.3052585.

Kelley Cotter. Playing the visibility game: How digital influencers and algorithms negotiate influence on Instagram. *New Media & Society*, 21(4):895–913, 2019. doi:10.1177/1461444818815684.

Sarah Dean and Jamie Morgenstern. Preference dynamics under personalized recommendations. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pp. 795–816. Association for Computing Machinery, 2022. doi:10.1145/3490486.3538346.

Sarah Dean, Evan Dong, Meena Jagadeesan, and Liu Leqi. Accounting for AI and users shaping one another: The role of mathematical models. *Transactions on Machine Learning Research*, 2025. URL https://openreview.net/forum?id=UkP4DhrJt1.

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. PMLR, 2023. URL https://proceedings.mlr.press/v202/eilat23a.html.

Noah E. Friedkin and Eugene C. Johnsen. Social influence and opinions. *The Journal of Mathematical Sociology*, 15(3–4):193–206, 1990. doi:10.1080/0022250X.1990.9990069.

F. Maxwell Harper and Joseph A. Konstan. The MovieLens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems*, 5(4):19:1–19:19, 2015. doi:10.1145/2827872.

Jiri Hron, Karl Krauth, Michael I. Jordan, Niki Kilbertus, and Sarah Dean. Modeling content creator incentives on algorithm-curated platforms. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=l6CpxixmUg.

Meena Jagadeesan, Nikhil Garg, and Jacob Steinhardt. Supply-side equilibria in recommender systems. In *Advances in Neural Information Processing Systems*, volume 36, pp. 14597–14608, 2023. doi:10.52202/075280-0642.

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. doi:10.1145/3306618.3314288.

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.

Yehuda Koren. Collaborative filtering with temporal dynamics. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 447–456, 2009. doi:10.1145/1557019.1557072.

Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009. doi:10.1109/MC.2009.263.

Tao Lin, Kun Jin, Andrew Estornell, Xiaoying Zhang, Yiling Chen, and Yang Liu. User-creator feature polarization in recommender systems with dual influence. In *Advances in Neural Information Processing Systems*, volume 37, pp. 125663–125701, 2024. doi:10.52202/079017-3993.

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. PMLR, 2020. URL https://proceedings.mlr.press/v119/mladenov20a.html.

Sergey E. Parsegov, Anton V. Proskurnikov, Roberto Tempo, and Noah E. Friedkin. Novel multidimensional models of opinion dynamics in social networks. *IEEE Transactions on Automatic Control*, 62(5):2270–2285, 2017. doi:10.1109/TAC.2016.2613905.

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. PMLR, 2020. URL https://proceedings.mlr.press/v119/perdomo20a.html.

Anton V. Proskurnikov and Roberto Tempo. A tutorial on modeling and analysis of dynamic social networks. part I. *Annual Reviews in Control*, 43:65–79, 2017. doi:10.1016/j.arcontrol.2017.03.002.

Laura Schelenz. Diversity-aware recommendations for social justice? exploring user diversity and fairness in recommender systems. In *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, pp. 404–410, 2021. doi:10.1145/3450614.3463293.

Peter H. Schönemann. A generalized solution of the orthogonal Procrustes problem. *Psychometrika*, 31(1):1–10, 1966. doi:10.1007/BF02289451.

Lukas Schüepp, Carmen Amo Alonso, Florian Dörfler, and Giulia De Pasquale. Socially-aware recommender systems mitigate opinion clusterization, 2026. URL https://arxiv.org/abs/2601.02412v1. arXiv:2601.02412v1.

Ben Sprenger, Giulia De Pasquale, Raffaele Soloperto, John Lygeros, and Florian Dörfler. Control strategies for recommendation systems in social networks. *IEEE Control Systems Letters*, 8:634–639, 2024. doi:10.1109/LCSYS.2024.3400701.

Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 2 edition, 2018. ISBN 9780262039246. URL https://mitpress.mit.edu/9780262039246/reinforcement-learning/.

Ji Yang, Xinyang Yi, Derek Zhiyuan Cheng, Lichan Hong, Yang Li, Simon Xiaoming Wang, Taibai Xu, and Ed H. Chi. Mixed negative sampling for learning two-tower neural networks in recommendations. In *Companion Proceedings of the Web Conference 2020*, pp. 441–447. Association for Computing Machinery, 2020. doi:10.1145/3366424.3386195.

Fan Yao, Yiming Liao, Jingzhou Liu, Shaoliang Nie, Qifan Wang, Haifeng Xu, and Hongning Wang. Unveiling user satisfaction and creator productivity trade-offs in recommendation platforms. In *Advances in Neural Information Processing Systems*, volume 37, pp. 86958–86984. Curran Associates, Inc., 2024. doi:10.52202/079017-2759.

Yuying Zhao, Yu Wang, Yunchao Liu, Xueqi Cheng, Charu C. Aggarwal, and Tyler Derr. Fairness and diversity in recommender systems: A survey. *ACM Transactions on Intelligent Systems and Technology*, 16(1):2:1–2:28, 2025. doi:10.1145/3664928.

Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, and Georg Lausen. Improving recommendation lists through topic diversification. In *Proceedings of the 14th International Conference on World Wide Web*, pp. 22–32. Association for Computing Machinery, 2005. doi:10.1145/1060745.1060754.

A Modeling Details and Metric Relationships

We adopt the mutual-influence perspective of Lin et al. (2024). Their updates use sampled recommendations, normalize states to the unit sphere, and allow responses away from negatively scored content. Here all responses are assimilative: we use convex averages in the unit ball with fixed reference vectors, following the structure of Friedkin & Johnsen (1990); Schüepp et al. (2026). Averaging captures a gradual response toward recommended content or audience preferences and gives an affine update map when recommendation fractions are fixed, which supports the equilibrium analysis in Appendix B.1.

The anchors represent persistent user interests and creator specialization or production constraints, following the idea of persistent exogenous influence in social-influence models (Friedkin & Johnsen, 1990) and their extension to users and creators (Schüepp et al., 2026). Our anchors are fixed through time but may differ from initial states, as in opinion models with persistent inputs independent of initialization (Proskurnikov & Tempo, 2017). A related distinction appears in temporal recommendation models: Koren (2009, Section 4.3) separates learned user factors into stationary, drifting, and day-specific components. Although anchors enter from the first update, their contribution accumulates as the direct weight of the initial state declines. Persistent characteristics can therefore take time to emerge in the current representation, where they are balanced against the response to exposure. This balance need not select a single outcome: different assignments can support stable equilibria with the same anchors when each gives every user a strictly highest-scoring creator at its candidate equilibrium.

With an assignment held fixed, the equilibrium balances each agent's anchor against its response to recommendations. The response has weight $\eta_j/(\kappa_j + \eta_j)$ for user j and $\xi_i/(\lambda_i + \xi_i)$ for creator i, with the remaining weight assigned to the corresponding anchor. Multiplying an agent's anchor and response coefficients by the same positive factor leaves these proportions, and hence its equilibrium equation, unchanged, provided the new coefficients still satisfy the model's constraints. Such rescaling preserves the equilibria, but changes the size of each update and can affect convergence speed and which equilibrium is reached from a given initial state. This distinction explains why the experiments fix $\kappa_j + \eta_j = \lambda_i + \xi_i = 0.6$ while varying the balance between anchoring and response.

By making creator adaptation depend on audience composition, the model captures a response to whom the content reaches. Audience size does not change the response weight, and revenue or production decisions are not modeled separately.

Each round uses the exposure fractions directly. If recommendations were sampled instead, the average audience would be a random ratio whose expectation can differ from the audience average in (2). The model is deterministic, with the platform scoring current states directly and no separate retraining step or direct influence between users. We approach pure exploitation as ε decreases to zero; keeping $\varepsilon > 0$ ensures that every creator has an audience, so its update is defined.

A.1 Relationships between the outcome metrics

The three primary metrics are defined in Section 3; exposure Gini is a secondary diagnostic defined below. Under the unit-ball assumption, $\mathrm{RR} \in [-1, 1]$, $\mathrm{CD} \in [0, 2]$, and $\mathrm{RD} \in [0, 1]$. RR averages user–creator dot products, so it can be negative; zero RR can also result from cancellation of positive and negative dot products. CD is zero exactly when all creators coincide, while RD is zero when each user's exposed creator vectors coincide. Since every creator has positive exposure at $\varepsilon > 0$, RD and CD vanish together in this domain; at $\varepsilon = 0$, RD is zero regardless of creator differences. The RD upper bound follows from the variance identity in (24).

At uniform exposure, the pairwise variance identity gives, at every time,

$$\mathrm{RD}(t)\big|_{\varepsilon=1} = \frac{1}{N}\sum_i \|\mathbf{c}_i(t) - \overline{\mathbf{c}}(t)\|_2^2 = \frac{1}{2N^2}\sum_{i,k} \|\mathbf{c}_i(t) - \mathbf{c}_k(t)\|_2^2. \tag{13}$$

Uniform RD therefore depends on squared pairwise distances, whereas CD averages the distances themselves. Because these summaries weight separation differently, CD does not in general determine RD.

Let $e_i = \sum_j p^{\varepsilon}_{ij}$ be creator i's total exposure and $n_i = |\{j : h_j = i\}|$ the number of users selecting it for exploitation. We use the Gini index $\sum_{i,k} |e_i - e_k|/(2N \sum_i e_i)$. The policy gives $e_i = (1-\varepsilon)n_i + M\varepsilon/N$ and $\sum_i e_i = M$, hence

$$\mathrm{Gini}(t) = (1-\varepsilon)\frac{\sum_{i,k} |n_i - n_k|}{2NM} \le (1-\varepsilon)\frac{N-1}{N}. \tag{14}$$

For a fixed exploration rate, the upper bound is attained when all users select the same creator for exploitation. The remaining creators then receive exposure only through uniform exploration. Although the explicit factor $1-\varepsilon$ decreases with exploration, the assignment counts also change as preferences and content evolve, so equilibrium Gini need not decrease. A zero value means that creators receive equal total exposure, but allows users to see different mixtures and creators to provide similar content.

B Proofs and Equilibrium Computations

B.1 Contraction and local stability

We use F^{ε} for the update map defined by (1)–(3), and $H(\mathbf{z}) = (h_1(\mathbf{z}), \dots, h_M(\mathbf{z}))$ for the assignment selected by the current scores. Thus $\mathbf{z}(t+1) = F^{\varepsilon}(\mathbf{z}(t))$. The distance used below is the largest Euclidean distance between corresponding agent states. Its associated norm and the smallest anchor weight are

$$\|\mathbf{z}\|_{\max} := \max\left\{ \max_j \|\mathbf{u}_j\|_2, \max_i \|\mathbf{c}_i\|_2 \right\}, \qquad \underline{\alpha} := \min\left\{ \min_j \kappa_j, \min_i \lambda_i \right\}. \tag{15}$$

All neighborhoods below are relative to the admissible domain $\mathbb{B}^{M+N}$. For an assignment $G = (g_1, \dots, g_M) \in [N]^M$, the auxiliary affine map F^{ε}_G holds the highest-scoring choices fixed at G and uses

$$p^{G,\varepsilon}_{ij} := (1-\varepsilon)\mathbf{1}\{i = g_j\} + \frac{\varepsilon}{N}. \tag{16}$$

The map F^{ε} recomputes choices from the current state using the policy in (3) at every round: $F^{\varepsilon}(\mathbf{z}) = F^{\varepsilon}_{H(\mathbf{z})}(\mathbf{z})$. The following lemma gives a unique fixed point $\mathbf{z}^*_G(\varepsilon)$ for each auxiliary map F^{ε}_G. This point is an equilibrium of the full system whenever its scores reproduce assignment G, because both maps then use the same recommendation fractions. Several assignments may satisfy this consistency condition, allowing several equilibria for the same model. Holding G fixed while varying ε traces a family of candidate equilibria, which we call an *equilibrium branch*. To determine whether a candidate reproduces its assigned choices with strict inequalities, we use its smallest assigned score gap:

$$\Delta_G(\varepsilon) := \min_j \min_{k \ne g_j} (\mathbf{u}^*_{j,G}(\varepsilon))^{\top} \big(\mathbf{c}^*_{g_j,G}(\varepsilon) - \mathbf{c}^*_{k,G}(\varepsilon) \big). \tag{17}$$

A positive gap means that the candidate assignment is reproduced by its own equilibrium scores. At $\varepsilon = 1$, all fixed-assignment maps coincide, so $\gamma = \Delta_{G_{\mathrm{unif}}}(1)$. Near one, the equilibrium denoted by $\mathbf{z}^*(\varepsilon)$ in the main text is $\mathbf{z}^*_{G_{\mathrm{unif}}}(\varepsilon)$.

The next lemma applies the standard contraction argument for anchored averaging to this fixed-exposure system; compare the fixed-partition dynamics in Schüepp et al. (2026, Theorem 1). We include the proof to make the convergence rate explicit.

Lemma B.1 (Contraction under a fixed assignment)**.** *If $\underline{\alpha} > 0$, then for every G and $\varepsilon \in (0,1]$, F^{ε}_G has a unique equilibrium $\mathbf{z}^*_G(\varepsilon) \in \mathbb{B}^{M+N}$. For any admissible states $\mathbf{z}, \widetilde{\mathbf{z}}$ and every integer $t \ge 0$,*

$$\|(F^{\varepsilon}_G)^t(\mathbf{z}) - (F^{\varepsilon}_G)^t(\widetilde{\mathbf{z}})\|_{\max} \le (1-\underline{\alpha})^t \|\mathbf{z} - \widetilde{\mathbf{z}}\|_{\max}. \tag{18}$$

Proof of Lemma B.1. We fix G and $\varepsilon > 0$. The exposure fractions in (16) sum to one over creators for each user. Each creator has positive exposure, so its audience weights $w_{ij} = p^{G,\varepsilon}_{ij} / \sum_{\ell} p^{G,\varepsilon}_{i\ell}$ are well-defined and sum to one over users. The updates are convex combinations of vectors in $\mathbb{B}$, hence preserve $\mathbb{B}^{M+N}$.

For two states $\mathbf{z}, \widetilde{\mathbf{z}}$ with the same anchors, the anchor terms cancel, and subtracting the updates gives

$$F^{\varepsilon}_{G,u_j}(\mathbf{z}) - F^{\varepsilon}_{G,u_j}(\widetilde{\mathbf{z}}) = (1-\kappa_j-\eta_j)(\mathbf{u}_j - \widetilde{\mathbf{u}}_j) + \eta_j \sum_i p^{G,\varepsilon}_{ij}(\mathbf{c}_i - \widetilde{\mathbf{c}}_i),$$

$$F^{\varepsilon}_{G,c_i}(\mathbf{z}) - F^{\varepsilon}_{G,c_i}(\widetilde{\mathbf{z}}) = (1-\lambda_i-\xi_i)(\mathbf{c}_i - \widetilde{\mathbf{c}}_i) + \xi_i \sum_j w_{ij}(\mathbf{u}_j - \widetilde{\mathbf{u}}_j).$$

Taking norms gives

$$\begin{aligned}
\|F^{\varepsilon}_{G,u_j}(\mathbf{z}) - F^{\varepsilon}_{G,u_j}(\widetilde{\mathbf{z}})\|_2 &\le (1-\kappa_j-\eta_j)\|\mathbf{u}_j - \widetilde{\mathbf{u}}_j\|_2 + \eta_j \sum_i p^{G,\varepsilon}_{ij}\|\mathbf{c}_i - \widetilde{\mathbf{c}}_i\|_2 \\
&\le (1-\kappa_j)\|\mathbf{z} - \widetilde{\mathbf{z}}\|_{\max}, \\
\|F^{\varepsilon}_{G,c_i}(\mathbf{z}) - F^{\varepsilon}_{G,c_i}(\widetilde{\mathbf{z}})\|_2 &\le (1-\lambda_i-\xi_i)\|\mathbf{c}_i - \widetilde{\mathbf{c}}_i\|_2 + \xi_i \sum_j w_{ij}\|\mathbf{u}_j - \widetilde{\mathbf{u}}_j\|_2 \\
&\le (1-\lambda_i)\|\mathbf{z} - \widetilde{\mathbf{z}}\|_{\max}.
\end{aligned}$$

Taking the maximum gives contraction factor $1-\underline{\alpha} < 1$. The closed product ball is complete, so the contraction mapping theorem gives a unique fixed point there. Iteration proves (18). □

We use the following score estimate repeatedly. For two admissible states $\mathbf{z}, \widetilde{\mathbf{z}}$, each pairwise score gap obeys

$$\begin{aligned}
&\big|\mathbf{u}_j^\top(\mathbf{c}_i - \mathbf{c}_k) - \widetilde{\mathbf{u}}_j^\top(\widetilde{\mathbf{c}}_i - \widetilde{\mathbf{c}}_k)\big| \\
&\quad\le \|\mathbf{u}_j - \widetilde{\mathbf{u}}_j\|_2\|\mathbf{c}_i - \mathbf{c}_k\|_2 + \|\widetilde{\mathbf{u}}_j\|_2\big(\|\mathbf{c}_i - \widetilde{\mathbf{c}}_i\|_2 + \|\mathbf{c}_k - \widetilde{\mathbf{c}}_k\|_2\big) \\
&\quad\le 4\|\mathbf{z} - \widetilde{\mathbf{z}}\|_{\max}.
\end{aligned} \tag{19}$$

If each selected creator has a strictly higher score than its competitors at an equilibrium, sufficiently small changes in the states preserve those choices. The fixed-assignment contraction then brings the system back to that equilibrium, as the following proposition establishes.

Proposition B.2 (Strict margins give local stability)**.** *Suppose $\underline{\alpha} > 0$ and $\varepsilon \in (0,1]$. If $\Delta_G(\varepsilon) > 0$, then $\mathbf{z}^*_G(\varepsilon)$ is an equilibrium of F^ε. Every admissible initialization satisfying*

$$\|\mathbf{z}(0) - \mathbf{z}^*_G(\varepsilon)\|_{\max} < \frac{\Delta_G(\varepsilon)}{4} \tag{20}$$

keeps assignment G and converges at rate at most $1-\underline{\alpha}$. Conversely, every equilibrium $\mathbf{z}^*$ of F^ε equals $\mathbf{z}^*_{H(\mathbf{z}^*)}(\varepsilon)$.

Proof of Proposition B.2. The inequality $\Delta_G(\varepsilon) > 0$ gives $H(\mathbf{z}^*_G(\varepsilon)) = G$, so the auxiliary fixed point also satisfies $F^\varepsilon(\mathbf{z}^*_G) = \mathbf{z}^*_G$. If (20) holds, (19) keeps every assigned score gap positive throughout the closed ball centered at $\mathbf{z}^*_G(\varepsilon)$ with radius $\|\mathbf{z}(0) - \mathbf{z}^*_G(\varepsilon)\|_{\max}$. In that ball the state-dependent map equals F^ε_G. Contraction maps the ball into itself and gives

$$\|\mathbf{z}(t) - \mathbf{z}^*_G(\varepsilon)\|_{\max} \le (1-\underline{\alpha})^t\|\mathbf{z}(0) - \mathbf{z}^*_G(\varepsilon)\|_{\max}.$$

This proves stability and absence of switching. Conversely, if $F^\varepsilon(\mathbf{z}^*) = \mathbf{z}^*$, then $F^\varepsilon_{H(\mathbf{z}^*)}(\mathbf{z}^*) = \mathbf{z}^*$. Uniqueness of that auxiliary fixed point gives $\mathbf{z}^* = \mathbf{z}^*_{H(\mathbf{z}^*)}(\varepsilon)$, including equilibria with ties. □

B.2 Global convergence under sufficient exploration

Proof of Theorem 4.1. The uniform map F^1 is independent of G and contracts with factor $1-\underline{\alpha} < 1$. We bound its difference from every F^ε_G on the entire state space. For user j, the change in the recommendation term is

$$\eta_j(1-\varepsilon)(\mathbf{c}_{g_j} - \overline{\mathbf{c}}),$$

whose norm is at most $2\eta_{\max}(1-\varepsilon)$. For creator i let $\mathcal{P}_i(G) = \{j : g_j = i\}$ and $\overline{\mathbf{u}} = M^{-1}\sum_j \mathbf{u}_j$. Its audience mean differs from the uniform audience mean by

$$\frac{(1-\varepsilon)\sum_{j\in\mathcal{P}_i(G)}(\mathbf{u}_j - \overline{\mathbf{u}})}{(1-\varepsilon)|\mathcal{P}_i(G)| + M\varepsilon/N}. \tag{21}$$

The unit-ball assumption gives $\|\overline{\mathbf{u}}\|_2 \le M^{-1}\sum_j \|\mathbf{u}_j\|_2 \le 1$, and hence $\|\mathbf{u}_j - \overline{\mathbf{u}}\|_2 \le \|\mathbf{u}_j\|_2 + \|\overline{\mathbf{u}}\|_2 \le 2$. Using also $|\mathcal{P}_i(G)| \le M$, its norm is at most

$$\frac{2(1-\varepsilon)|\mathcal{P}_i(G)|}{(1-\varepsilon)|\mathcal{P}_i(G)| + M\varepsilon/N} \le \frac{2N(1-\varepsilon)}{N(1-\varepsilon)+\varepsilon}.$$

An empty pool gives zero difference and is covered by this bound. Therefore

$$\|F^{\varepsilon}_G(\mathbf{z}) - F^1(\mathbf{z})\|_{\max} \le \omega_\varepsilon := 2\max\left\{\eta_{\max}(1-\varepsilon),\, \xi_{\max}\frac{N(1-\varepsilon)}{N(1-\varepsilon)+\varepsilon}\right\} \tag{22}$$

for every G and admissible $\mathbf{z}$.

In particular this holds for the assignment selected by the state-dependent map. Writing $\mathrm{err}_t = \|\mathbf{z}(t) - \mathbf{z}^*_{\mathrm{unif}}\|_{\max}$ gives

$$\begin{aligned} \mathrm{err}_{t+1} &\le (1-\underline{\alpha})\mathrm{err}_t + \omega_\varepsilon, \\ \mathrm{err}_t &\le 2(1-\underline{\alpha})^t + \frac{\omega_\varepsilon}{\underline{\alpha}}, \end{aligned} \tag{23}$$

using $\mathrm{err}_0 \le 2$. By (19), all assigned gaps are at least $\gamma/2$ in the closed ball of radius $\gamma/8$ around $\mathbf{z}^*_{\mathrm{unif}}$. Hence its assignment is always G_{unif}.

Condition (6) is precisely $\omega_\varepsilon < \underline{\alpha}\gamma/8$. Equation (23) then puts every trajectory in this ball after finitely many steps. The ball is forward invariant because $(1-\underline{\alpha})\gamma/8 + \omega_\varepsilon < \gamma/8$. Thereafter the state-dependent map equals $F^{\varepsilon}_{G_{\mathrm{unif}}}$, so its trajectory converges geometrically to that map's unique fixed point. The limit remains in the closed ball and has unique highest-scoring choices, so it is also an equilibrium of the state-dependent map. Every initialization reaches this same limit, and any other equilibrium would give a constant trajectory with a different limit, which is impossible. Finally, $\omega_\varepsilon \to 0$ as $\varepsilon \uparrow 1$, so the strict condition holds in a neighborhood of full exploration. □

B.3 RD DECOMPOSITION AND THE DERIVATIVE AT UNIFORM EXPOSURE

For any probability vector p and creator vectors, the variance identity is

$$\sum_i p_i\|\mathbf{c}_i - \boldsymbol{\mu}\|_2^2 = \sum_i p_i\|\mathbf{c}_i\|_2^2 - \|\boldsymbol{\mu}\|_2^2, \qquad \boldsymbol{\mu} = \sum_i p_i\mathbf{c}_i. \tag{24}$$

This also gives $0 \le \mathrm{RD} \le 1$ in the unit ball. Let $\mathbf{c}^\circ_i = \mathbf{c}_i - \overline{\mathbf{c}}$. For the uniform-plus-top mixture, the centered second moment is $\varepsilon V_{\mathrm{all}} + (1-\varepsilon)\|\mathbf{c}^\circ_{h_j(\mathbf{z})}\|_2^2$, and the centered mean is $(1-\varepsilon)\mathbf{c}^\circ_{h_j(\mathbf{z})}$. Subtracting its squared norm and averaging over users proves (8).

At equilibrium, it is useful to express response as a fraction of the combined weight of anchoring and response. For a fixed assignment, define

$$r_j := \frac{\eta_j}{\kappa_j + \eta_j}, \qquad s_i := \frac{\xi_i}{\lambda_i + \xi_i}. \tag{25}$$

Positive anchoring ensures $0 \le r_j, s_i < 1$. Rearranging the fixed-point equations from (1)–(2) gives the following system, where $\mathbf{m}_i(\varepsilon)$ denotes creator i's audience mean and assignment and equilibrium superscripts are suppressed:

$$\mathbf{u}_j(\varepsilon) = (1-r_j)\mathbf{a}_j + r_j[(1-\varepsilon)\mathbf{c}_{g_j}(\varepsilon) + \varepsilon\overline{\mathbf{c}}(\varepsilon)], \tag{26}$$

$$\mathbf{c}_i(\varepsilon) = (1-s_i)\mathbf{b}_i + s_i\mathbf{m}_i(\varepsilon), \tag{27}$$

$$\mathbf{m}_i(\varepsilon) = \frac{(1-\varepsilon)\sum_{j\in\mathcal{P}_i(G)}\mathbf{u}_j(\varepsilon) + \varepsilon M\overline{\mathbf{u}}(\varepsilon)/N}{(1-\varepsilon)|\mathcal{P}_i(G)| + \varepsilon M/N}. \tag{28}$$

For $\varepsilon \in (0,1]$ every denominator is positive. The fixed-assignment update map depends smoothly on ε, and its linear part has norm at most $1-\underline{\alpha} < 1$ in the block maximum norm of (15). Consequently its fixed-point linear system is invertible and $\mathbf{z}^*_G(\varepsilon)$ is smooth on $(0,1]$, with derivatives from the left at one. A reduced system is given in Appendix B.4. When $\gamma > 0$, Theorem 4.1 identifies this smooth branch with the unique equilibrium of the state-dependent system for all ε sufficiently close to one, with $G = G_{\mathrm{unif}}$.

Proof of Proposition 4.2. Differentiating (8) along that branch at $\varepsilon = 1$ from the left gives (10). The term containing the derivative of V_{top} vanishes because its coefficient is $\varepsilon(1-\varepsilon)$. To compute the adaptation term with heterogeneous coefficients, we track how changing exposure affects users and creators at the uniform equilibrium. Define

$$\boldsymbol{\psi}_i := \frac{N}{M} \sum_{j \in \mathcal{P}_i(G_{\text{unif}})} (\mathbf{u}^*_{j,\text{unif}} - \overline{\mathbf{u}}^*_{\text{unif}}), \qquad \mathbf{f}_u := \frac{1}{M} \sum_j r_j (\mathbf{c}^*_{g_j^{\text{unif}},\text{unif}} - \overline{\mathbf{c}}^*_{\text{unif}}),$$
$$\mathbf{f}_c := \frac{1}{N} \sum_i s_i \boldsymbol{\psi}_i, \qquad \overline{r} := \frac{1}{M} \sum_j r_j, \qquad \overline{s} := \frac{1}{N} \sum_i s_i. \tag{29}$$

For creator i, the group $\mathcal{P}_i(G_{\text{unif}})$ contains the users whose highest-scoring creator at the uniform equilibrium is i. The sum $\boldsymbol{\psi}_i$ captures their aggregate deviation from the population mean, scaled by N/M, and therefore reflects both their preferences and their number; an empty group contributes zero. The vectors $\mathbf{f}_u$ and $\mathbf{f}_c$ collect the corresponding response-weighted differences on the two sides of the model. All derivatives below are with respect to ε and evaluated at 1^- unless stated otherwise. Differentiating the audience quotient and the user equation gives

$$\left.\frac{d\mathbf{m}_i}{d\varepsilon}\right|_{1^-} = \left.\frac{d\overline{\mathbf{u}}}{d\varepsilon}\right|_{1^-} - \boldsymbol{\psi}_i, \qquad \left.\frac{d\overline{\mathbf{u}}}{d\varepsilon}\right|_{1^-} = -\mathbf{f}_u + \overline{r}\, \left.\frac{d\overline{\mathbf{c}}}{d\varepsilon}\right|_{1^-}.$$

The creator equation gives $\left.\frac{d\overline{\mathbf{c}}}{d\varepsilon}\right|_{1^-} = \overline{s}\, \left.\frac{d\overline{\mathbf{u}}}{d\varepsilon}\right|_{1^-} - \mathbf{f}_c$. Solving these two mean equations gives

$$\left.\frac{d\overline{\mathbf{u}}}{d\varepsilon}\right|_{1^-} = -\frac{\mathbf{f}_u + \overline{r}\mathbf{f}_c}{1 - \overline{r}\,\overline{s}}, \qquad \left.\frac{d\mathbf{c}_i}{d\varepsilon}\right|_{1^-} = s_i \left(\left.\frac{d\overline{\mathbf{u}}}{d\varepsilon}\right|_{1^-} - \boldsymbol{\psi}_i \right). \tag{30}$$

The denominator is positive. Since the centered creator vectors sum to zero, differentiation of their variance simplifies to

$$\left.\frac{d}{d\varepsilon} V_{\text{all}}(\mathbf{z}^*(\varepsilon))\right|_{1^-} = \frac{2}{N} \sum_i (\mathbf{c}^*_{i,\text{unif}} - \overline{\mathbf{c}}^*_{\text{unif}})^\top \left.\frac{d\mathbf{c}_i}{d\varepsilon}\right|_{1^-}. \tag{31}$$

If $\left.\frac{d}{d\varepsilon}\text{RD}^*(\varepsilon)\right|_{1^-} < 0$, differentiability gives $\text{RD}^*(\varepsilon) > \text{RD}^*(1)$ for every $\varepsilon < 1$ sufficiently close to one. Theorem 4.1 makes these equilibria globally attracting, proving the initialization-independent conclusion. □

Corollary B.3 (Equal response fractions within each side)**.** *Suppose $\underline{\alpha} > 0$, $\gamma > 0$, and all users share $r = \eta_j/(\kappa_j + \eta_j)$ while all creators share $s = \xi_i/(\lambda_i + \xi_i)$. These fractions measure the weight given to adaptation relative to the total weight of adaptation and anchoring. Let $\overline{\mathbf{a}}, \overline{\mathbf{b}}$ be the user and creator anchor means. Define the dispersion of all creator anchors, the mean squared distance of selected anchors from the same creator-anchor mean, and the alignment between user anchors and their selected creator anchors by*

$$V_b := \frac{1}{N} \sum_i \|\mathbf{b}_i - \overline{\mathbf{b}}\|_2^2, \qquad T_b := \frac{1}{M} \sum_j \|\mathbf{b}_{g_j^{\text{unif}}} - \overline{\mathbf{b}}\|_2^2,$$
$$C_{ab} := \frac{1}{M} \sum_j (\mathbf{a}_j - \overline{\mathbf{a}})^\top (\mathbf{b}_{g_j^{\text{unif}}} - \overline{\mathbf{b}}). \tag{32}$$

Then

$$\left.\frac{d}{d\varepsilon}\text{RD}^*(\varepsilon)\right|_{\varepsilon=1^-} = (1-s)\big[(1-s)(V_b - T_b) - 2s(1-r)C_{ab}\big], \tag{33}$$

where $C_{ab} \geq 0$, with strict inequality whenever two users have different, unique highest-scoring creators at the uniform equilibrium.

The allocation contribution is $(1-s)^2(V_b - T_b)$, while adaptation contributes $-2s(1-s)(1-r)C_{ab} \leq 0$. Thus adaptation can make the total slope negative even when reallocation alone would not. This expression uses anchor statistics and response fractions; the assumptions require equal fractions within each side, but allow different total update weights.

Proof of Corollary B.3. At uniform exposure, $\mathbf{m}^*_{i,\mathrm{unif}} = \overline{\mathbf{u}}^*_{\mathrm{unif}}$ for every creator, so common fractions r, s imply

$$\mathbf{c}^*_{i,\mathrm{unif}} - \overline{\mathbf{c}}^*_{\mathrm{unif}} = (1-s)(\mathbf{b}_i - \overline{\mathbf{b}}), \qquad \mathbf{u}^*_{j,\mathrm{unif}} - \overline{\mathbf{u}}^*_{\mathrm{unif}} = (1-r)(\mathbf{a}_j - \overline{\mathbf{a}}).$$

Therefore $V_{\mathrm{all}}(\mathbf{z}^*_{\mathrm{unif}}) = (1-s)^2 V_b$ and $V_{\mathrm{top}}(\mathbf{z}^*_{\mathrm{unif}}) = (1-s)^2 T_b$. Substituting (30) into (31), the common mean-derivative term cancels and leaves

$$\frac{d}{d\varepsilon} V_{\mathrm{all}}(\mathbf{z}^*(\varepsilon))\bigg|_{1^-} = -\frac{2s(1-s)(1-r)}{M} \sum_j (\mathbf{b}_{g_j^{\mathrm{unif}}} - \overline{\mathbf{b}})^\top (\mathbf{a}_j - \overline{\mathbf{a}}) = -2s(1-s)(1-r)C_{ab}.$$

This proves (33). For the sign assertion, user j scores g_j^{unif} at least as highly as g_ℓ^{unif}, while user ℓ has the reverse ordering. Adding these two inequalities gives

$$(\mathbf{u}^*_{j,\mathrm{unif}} - \mathbf{u}^*_{\ell,\mathrm{unif}})^\top (\mathbf{c}^*_{g_j^{\mathrm{unif}},\mathrm{unif}} - \mathbf{c}^*_{g_\ell^{\mathrm{unif}},\mathrm{unif}}) \geq 0.$$

After dividing by $(1-r)(1-s) > 0$, summing over ordered pairs, and using the definition of C_{ab}, this becomes

$$C_{ab} = \frac{1}{2M^2} \sum_{j,\ell} (\mathbf{a}_j - \mathbf{a}_\ell)^\top (\mathbf{b}_{g_j^{\mathrm{unif}}} - \mathbf{b}_{g_\ell^{\mathrm{unif}}}) \geq 0. \tag{34}$$

If some pair has distinct unique maximizers, both corresponding optimality inequalities are strict, making this sum strictly positive. □

B.4 Intermediate advantage and equilibrium computation

Proof of Theorem 4.3. Proposition 4.2 and Theorem 4.1 provide $\varepsilon_+ \in (0,1)$ with a unique globally attracting equilibrium and $\mathrm{RD}^*(\varepsilon_+) > \mathrm{RD}^*(1)$. For any state and user, write $\boldsymbol{\mu}_j^\varepsilon = \sum_i p_{ij}^\varepsilon(\mathbf{z}) \mathbf{c}_i$ for the mean recommended content. This mean minimizes exposure-weighted squared distance, hence

$$\sum_i p_{ij}^\varepsilon(\mathbf{z}) \|\mathbf{c}_i - \boldsymbol{\mu}_j^\varepsilon\|_2^2 \leq \sum_i p_{ij}^\varepsilon(\mathbf{z}) \|\mathbf{c}_i - \mathbf{c}_{h_j(\mathbf{z})}\|_2^2 \leq 4\varepsilon \frac{N-1}{N}. \tag{35}$$

The selected term is zero, and each other term has distance at most two and exposure fraction ε/N. Averaging proves (12). Because this bound tends to zero as $\varepsilon \downarrow 0$ and $\mathrm{RD}^*(\varepsilon_+) > 0$, choose $\varepsilon_- \in (0, \varepsilon_+)$ with $4\varepsilon_-(N-1)/N < \mathrm{RD}^*(\varepsilon_+)$. The same strict comparison then holds for all states and all $\varepsilon \in (0, \varepsilon_-]$. □

For a fixed assignment G, we can compute its candidate equilibrium by first solving for creator states and then recovering user states. This provides the independent equilibrium calculation used in the numerical checks in Appendix C.3. The assignment matrix $P_G(\varepsilon) \in \mathbb{R}^{M \times N}$ has entries $[P_G(\varepsilon)]_{ji} = (1-\varepsilon)\mathbf{1}\{i = g_j\} + \varepsilon/N$, and the corresponding normalized audience matrix and diagonal response matrices are

$$W_G(\varepsilon) := \mathrm{diag}(P_G(\varepsilon)^\top \mathbf{1}_M)^{-1} P_G(\varepsilon)^\top, \quad R := \mathrm{diag}(r_j), \quad S := \mathrm{diag}(s_i). \tag{36}$$

Here $\mathbf{1}_M$ is the vector of M ones. We use A, B for the matrices with rows $(1-r_j)\mathbf{a}_j^\top$ and $(1-s_i)\mathbf{b}_i^\top$, respectively. Collect equilibrium user and creator vectors as rows of U_G^* and C_G^*. These matrices satisfy

$$\begin{aligned} U_G^*(\varepsilon) &= \mathsf{A} + RP_G(\varepsilon)C_G^*(\varepsilon), & L_G(\varepsilon)C_G^*(\varepsilon) &= Y_G(\varepsilon), \\ L_G(\varepsilon) &:= I_N - SW_G(\varepsilon)RP_G(\varepsilon), & Y_G(\varepsilon) &:= \mathsf{B} + SW_G(\varepsilon)\mathsf{A}. \end{aligned} \tag{37}$$

Here I_N is the $N \times N$ identity matrix. Both P_G and W_G have nonnegative entries and rows summing to one for $\varepsilon \in (0,1]$. Thus

$$\|SW_G(\varepsilon)RP_G(\varepsilon)\|_\infty \leq \big(\max_i s_i\big)\big(\max_j r_j\big) < 1. \tag{38}$$

Here $\|\cdot\|_\infty$ is the maximum absolute row-sum matrix norm. The Neumann series therefore makes L_G invertible. We solve the same $N \times N$ system in (37) for each of the d feature coordinates, then recover user states from the first equation. The matrices L_G and Y_G vary smoothly for $\varepsilon \in (0,1]$, so

the equilibrium branch is smooth as well. Its first derivative is obtained using the same coefficient matrix:

$$L_G C'_G = Y'_G - L'_G C_G. \tag{39}$$

Here $C_G = C^*_G$, and primes denote derivatives with respect to ε. For $G = G_{\mathrm{unif}}$, evaluate the derivative from the left at $\varepsilon = 1$ and substitute it into (31) to obtain the adaptation contribution reported in Table 3. The candidate is an equilibrium of the policy in (3) when its scores reproduce G; a positive margin Δ_G guarantees this consistency and local stability by Proposition B.2.

C Experimental Details and Additional Results

C.1 MovieLens representations

MovieLens 1M contains 1,000,209 ratings from 6,040 users on 3,706 rated movies (Harper & Konstan, 2015, Table II). The downloader obtains the archive from `https://files.grouplens.org/datasets/movielens/ml-1m.zip` and checks its SHA-256 against the value stored in the source package. We use a cutoff of 1 July 2001 (UTC) and a recent window of 180 days, starting on 2 January 2001. Historical fitting uses ratings before the cutoff, including the recent window; recent fitting uses only the window. Before fitting, each pre-cutoff rating is assigned to validation with probability 0.1, using seed 20260925. The historical training and validation sets contain 857,376 and 94,695 ratings, respectively; their recent subsets contain 42,170 and 4,510 ratings. The 48,138 ratings after the cutoff are excluded from fitting and model selection.

We use a biased matrix-factorization rating model (Koren et al., 2009), adding a score scale τ and constraining both user and movie vectors to the unit ball:

$$\widehat{y}_{ji} = \overline{y}_{\mathrm{train}} + \beta_j + \zeta_i + \tau\,(\mathbf{v}_j^{\mathrm{user}})^\top \mathbf{v}_i^{\mathrm{movie}}. \tag{40}$$

Here $\overline{y}_{\mathrm{train}}$ is the mean training rating, β_j, ζ_i are user and movie biases, and τ is a positive score scale. The vectors $\mathbf{v}_j^{\mathrm{user}}, \mathbf{v}_i^{\mathrm{movie}} \in \mathbb{B} \subset \mathbb{R}^{16}$ are the fitted representations, with norms at most one, before they are assigned their roles in the dynamics. For a training set $\mathcal{D}_{\mathrm{train}}$, we minimize

$$\frac{1}{2} \sum_{(j,i,y_{ji}) \in \mathcal{D}_{\mathrm{train}}} (y_{ji} - \widehat{y}_{ji})^2 + \frac{\rho_{\mathrm{vec}}}{2} \left(\sum_j \|\mathbf{v}_j^{\mathrm{user}}\|_2^2 + \sum_i \|\mathbf{v}_i^{\mathrm{movie}}\|_2^2 \right) + \frac{\rho_{\mathrm{bias}}}{2} \left(\sum_j \beta_j^2 + \sum_i \zeta_i^2 \right). \tag{41}$$

The vector and bias penalties are apportioned across each agent's training ratings in stochastic gradient updates. Factors start with independent $\mathcal{N}(0, 0.08^2)$ entries and biases at zero. Each rating update uses the pre-update factors and residual, then projects a vector onto the unit ball only if its norm exceeds one. The learning rate at epoch e is $0.03/[1 + 0.035(e-1)]$. Training stops after ten epochs without validation improvement or at 100 epochs, retaining the checkpoint with lowest validation root mean squared error (RMSE).

For the historical fit, validation selects $\tau \in \{1, 2\}$ and $\rho_{\mathrm{vec}} \in \{5, 20, 100\}$ with $\rho_{\mathrm{bias}} = 5$. The selected values are $\tau = 2$ and $\rho_{\mathrm{vec}} = 20$. The recent fit retains $\tau = 2$ and selects $\rho_{\mathrm{vec}} \in \{1, 5, 20\}$, again choosing 20. The primary seed is 61431; historical and recent checkpoints occur at epochs 93 and 33, with validation RMSE 0.8416 and 0.8807. Seeds 61432–61435 reuse the selected hyperparameters and select their own checkpoints. Diversity outcomes are not part of the fitting objective or the hyperparameter selection rule.

To obtain representations supported by both time ranges, eligible users have at least 16 training ratings and movies at least 20, separately in the recent window and in the preceding history. This leaves 444 users and 730 movies. We draw 256 users and 80 movies without replacement within each population, independently between users and movies, using seeds 2026091701–2026091710. Population ID sets are shared across the five fits. All construction and fitting steps precede simulation of the dynamics.

The two fits have arbitrary coordinate orientations. Let U_{rec} and C_{rec} contain the eligible recent vectors as rows, and let $U_{\mathrm{hist}}, C_{\mathrm{hist}}$ contain their historical counterparts. We find one orthogonal matrix Q minimizing

$$\frac{\|U_{\mathrm{rec}} Q - U_{\mathrm{hist}}\|_F^2}{444} + \frac{\|C_{\mathrm{rec}} Q - C_{\mathrm{hist}}\|_F^2}{730}, \qquad Q^\top Q = I. \tag{42}$$

We solve this orthogonal Procrustes problem (Schönemann, 1966) using the singular value decomposition of the combined cross-product matrix. In the objective, the squared Frobenius norms sum squared coordinate differences, while the denominators give users and movies equal total weight. The aligned recent vectors then supply $\mathbf{u}_j(0), \mathbf{c}_i(0)$, and the historical vectors supply $\mathbf{a}_j, \mathbf{b}_i$. Because the same orthogonal transformation is applied to users and movies, it preserves their recent norms and dot products; apart from this alignment, norms remain as learned. The historical factors serve as proxies for persistent interests and specializations in this experiment, although ratings alone do not identify those model quantities or the subsequent adaptation coefficients. The simulation policy uses only vector dot products, without the fitted biases or rating scale.

C.2 Synthetic inputs and paired comparisons

For every synthetic initial state and anchor, we independently draw a vector from $\mathcal{N}(\mathbf{0}, I_{16})$, divide it by its norm, and multiply it by an independent radius drawn uniformly from $[0, 1]$. This gives isotropic directions and a uniform distribution of radii, rather than a uniform distribution with respect to volume in the ball. Initial states and anchors use separate random streams. Five ensembles use seeds 2026092501–2026092505, each with ten population seeds 2026091701–2026091710. The first ensemble supplies the primary curves.

For both protocols, write $r_j = \eta_j/(\kappa_j + \eta_j)$ and $s_i = \xi_i/(\lambda_i + \xi_i)$ for the response fractions. Their distribution medians (r_0, s_0) are $(0.15, 0.15)$, $(0.05, 0.90)$, $(0.85, 0.05)$, and $(0.65, 0.65)$ in the order of Table 1. With independent standard Gaussian Z_j, Z'_i, set

$$\begin{aligned} r_j &= \mathrm{logistic}(\mathrm{logit}(r_0) + 0.35Z_j), & s_i &= \mathrm{logistic}(\mathrm{logit}(s_0) + 0.35Z'_i), \\ \kappa_j &= 0.6(1 - r_j), \quad \eta_j = 0.6r_j, & \lambda_i &= 0.6(1 - s_i), \quad \xi_i = 0.6s_i. \end{aligned} \tag{43}$$

Here $\mathrm{logistic}(x) = 1/(1+e^{-x})$ and $\mathrm{logit}(p) = \log[p/(1-p)]$, so the draws keep response fractions strictly between zero and one. The previous-state weight is 0.4 for every agent. The Gaussian coefficient draws are shared across protocols, fits, and ensembles for each population seed. Within each population, every regime and exploration rate uses the same initial states and anchors. The exploration grid is

$$\{0.01, 0.02, 0.05, 0.10, 0.20, 0.35, 0.50, 0.65, 0.80, 0.90, 0.95, 0.98, 1.00\}.$$

All tested rates satisfy $\varepsilon > 0$, so every creator's audience mean is defined. At pure exploitation, RD evaluated from the policy is exactly zero, as shown in Section 4; no trajectory at $\varepsilon = 0$ is included in the figures.

Primary curves average ten populations, with pointwise Student-t intervals using nine degrees of freedom. MovieLens intervals describe population sampling conditional on a fit; the populations can share agents. Synthetic intervals are conditional on the ensemble construction. We report the five fits or ensembles separately rather than pooling all 50 populations into one interval. Table 2 uses paired population differences from $\varepsilon = 1$. To account for reporting the best tested rate, its intervals use a Bonferroni correction over $2 \times 4 \times 12 = 96$ nonuniform comparisons, with approximate simultaneous 95% coverage under the Student-t approximation. The dynamic maxima in Table 2 refer to this discrete grid. Only the separate fixed-state comparison in Appendix C.4 maximizes RD over the continuous interval $[0, 1]$.

C.3 Numerical convergence and the uniform derivative

Updates are synchronous and use the exposure fractions directly. A fixed-point candidate is detected when the largest Euclidean change of any user or creator vector is below 10^{-9} and, for $\varepsilon < 1$, the selected creators have remained unchanged for 30 rounds. The initial budget is 10,000 rounds. In parallel, every 20 rounds we test for periods from 2 to 12 by comparing five consecutive repetitions to tolerance 10^{-9}. Cases not verified under these rules are rerun for up to 100,000 rounds, with tolerance 10^{-12} and candidate periods up to 128. The records retain each run's status and period.

For a terminal state $\widehat{\mathbf{z}}$ with assignment G, define the residual $\delta = \|F^{\varepsilon}_G(\widehat{\mathbf{z}}) - \widehat{\mathbf{z}}\|_{\max}$ and let $\Delta(\widehat{\mathbf{z}})$ be its smallest top-versus-second score gap. With $\underline{\alpha}$ from Lemma B.1, we check

$$\Delta(\widehat{\mathbf{z}}) > \frac{4\delta}{\underline{\alpha}}. \tag{44}$$

Table 2: Primary recommendation-diversity (RD) results. The Uniform and Maximum columns report mean RD over ten populations at $\varepsilon = 1$ and at the best tested rate, identified in the ε column. Gain is the difference between these means before rounding, with paired intervals adjusted by Bonferroni correction over 96 comparisons for approximate simultaneous 95% coverage. For MovieLens user response, the maximum is the uniform baseline itself.

Inputs	Response	ε	Uniform	Maximum	Gain	Adjusted interval
MovieLens	Weak	0.90	0.6071	0.6161	0.0089	[0.0062, 0.0116]
MovieLens	Creator	0.35	0.0102	0.0590	0.0487	[0.0457, 0.0517]
MovieLens	User	1.00	0.7623	0.7623	0.0000	—
MovieLens	Coupled	0.65	0.1118	0.1424	0.0307	[0.0251, 0.0362]
Synthetic	Weak	0.65	0.2294	0.2853	0.0558	[0.0480, 0.0636]
Synthetic	Creator	0.35	0.0039	0.0251	0.0212	[0.0174, 0.0251]
Synthetic	User	0.65	0.2878	0.3385	0.0507	[0.0361, 0.0653]
Synthetic	Coupled	0.50	0.0425	0.0790	0.0366	[0.0327, 0.0405]

The ball of radius $\delta/\underline{\alpha}$ about $\widehat{\mathbf{z}}$, intersected with the state domain, is invariant under the fixed-assignment contraction: for any point in this ball,

$$\|F_G^{\varepsilon}(\mathbf{z}) - \widehat{\mathbf{z}}\|_{\max} \le (1-\underline{\alpha})\frac{\delta}{\underline{\alpha}} + \delta = \frac{\delta}{\underline{\alpha}}.$$

A score gap changes by at most $4\delta/\underline{\alpha}$ there, so (44) keeps the assignment unchanged and makes the state-dependent map coincide with the contraction. At $\varepsilon = 1$, contraction holds without a score-gap condition. We also solve the fixed-assignment equilibrium independently and verify its assignment. All 5,193 fixed-point runs pass these floating-point checks; the largest residual is 8.6×10^{-10} and the largest distance to the independent equilibrium solution is 6.7×10^{-9} (upper-rounded values).

For a candidate period L, we continue the updates through complete periods to refine its closing residual. With a fixed sequence of assignments, the L-step map contracts by at most $(1-\underline{\alpha})^L$, giving a radius bound equal to the closing residual divided by $1-(1-\underline{\alpha})^L$. We check that every phase has score gap exceeding four times this bound. All seven periodic cases pass this check. Their periods are between two and eight, all under synthetic creator response at $\varepsilon \le 0.1$. The primary case has period four at $\varepsilon = 0.02$; its RD ranges from 0.0026025 to 0.0026069. Each cycle contributes the mean of its metric values over one period. These checks are numerical, without rigorous rounding-error bounds, and establish local behavior from the simulated starts rather than the global guarantee of Theorem 4.1.

There are 80 primary uniform equilibria: ten populations times four regimes for each of the two input protocols. All have positive score gaps. Table 3 decomposes their RD derivative using (10). The allocation contribution $V_{\mathrm{all}}(\mathbf{z}^*_{\mathrm{unif}}) - V_{\mathrm{top}}(\mathbf{z}^*_{\mathrm{unif}})$ describes changing exposure while holding the uniform-equilibrium states fixed; the adaptation contribution is the derivative of $V_{\mathrm{all}}(\mathbf{z}^*(\varepsilon))$ along the equilibrium branch. The MovieLens user-response mean slope is positive, although three of its ten populations have negative slopes. All other protocol–regime combinations have negative slopes in every population. Negative total slopes predict an increase in RD when ε is reduced slightly from one. These local predictions do not determine the height or location of a maximum farther along the grid.

Richardson finite differences on the same, assignment-consistent branch agree with the analytic derivatives to within 1.9×10^{-8}. Across these 80 configurations, the sufficient neighborhoods in Theorem 4.1 have widths between approximately 3.7×10^{-11} and 7.2×10^{-7} below one. Even the largest nonuniform grid value, 0.98, is outside these neighborhoods. Dense and optimized implementations of the update agree within 3.0×10^{-15} across the saved outcomes.

Table 3: Derivative of equilibrium recommendation diversity (RD) with respect to exploration at $\varepsilon = 1$ from below, decomposed into the allocation contribution with states held fixed and the adaptation contribution from changing creator states along the joint equilibrium. Entries are means over ten populations, except Negative, which counts populations whose total derivative is negative and therefore predicts higher RD when exploration is reduced slightly from one.

Inputs	Response	Allocation	Adaptation	Total	Negative
MovieLens	Weak	-0.0774	-0.0772	-0.1546	10/10
MovieLens	Creator	-0.0111	-0.0793	-0.0904	10/10
MovieLens	User	0.0177	-0.0040	0.0137	3/10
MovieLens	Coupled	-0.0634	-0.0571	-0.1205	10/10
Synthetic	Weak	-0.3028	-0.0412	-0.3441	10/10
Synthetic	Creator	-0.0100	-0.0396	-0.0496	10/10
Synthetic	User	-0.3990	-0.0024	-0.4014	10/10
Synthetic	Coupled	-0.0875	-0.0304	-0.1180	10/10

C.4 Comparison with states fixed at uniform equilibrium

For each population and response regime, we hold the user and creator vectors at the states reached under uniform recommendation and vary ε only in the policy. Because these vectors remain fixed, each user's highest-scoring creator also remains unchanged, so (8) gives the entire fixed-state curve,

$$\mathrm{RD}(\mathbf{z}^*_{\mathrm{unif}}, \varepsilon) = \varepsilon V_{\mathrm{all}}(\mathbf{z}^*_{\mathrm{unif}}) + \varepsilon(1-\varepsilon) V_{\mathrm{top}}(\mathbf{z}^*_{\mathrm{unif}}). \tag{45}$$

The fixed-state and dynamic comparisons share the same RD at $\varepsilon = 1$, so their gains use a common baseline within each population. This comparison was added after the main simulations, using their saved uniform outcomes without refitting or selecting new populations.

We average each curve over the same ten populations before finding its maximum. For the dynamics, we use the largest mean RD on the 13 tested rates, as in Table 2. For fixed states, we maximize the mean quadratic in (45) analytically over $0 \le \varepsilon \le 1$. Thus the fixed-state comparison includes every rate in the interval, including rates outside the simulation grid. An interior maximum exists when the population mean of $V_{\mathrm{top}}(\mathbf{z}^*_{\mathrm{unif}})$ exceeds the population mean of $V_{\mathrm{all}}(\mathbf{z}^*_{\mathrm{unif}})$; otherwise uniform recommendation maximizes the fixed-state mean RD. Writing these population means as $\overline{V}_{\mathrm{top}}$ and $\overline{V}_{\mathrm{all}}$, the interior maximizing rate is $(\overline{V}_{\mathrm{all}} + \overline{V}_{\mathrm{top}})/(2\overline{V}_{\mathrm{top}})$.

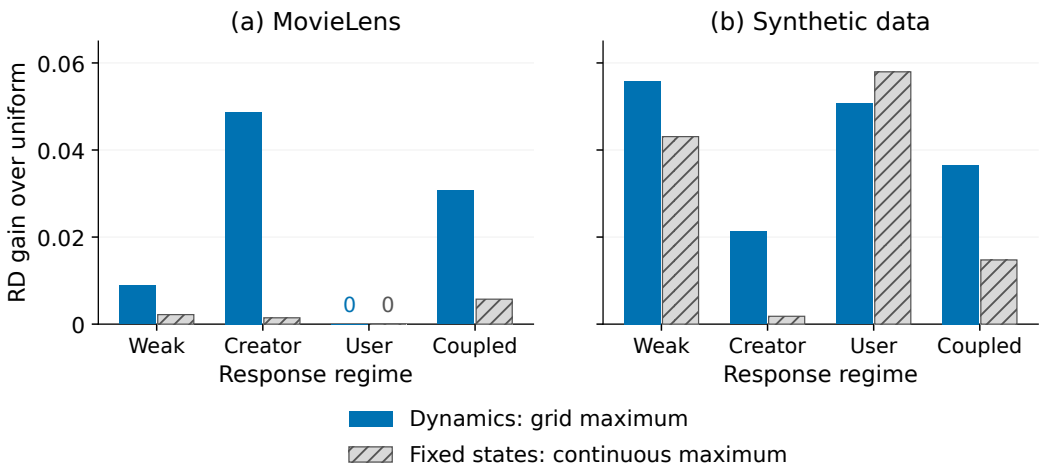


Figure 5: Recommendation-diversity (RD) gain over the uniform baseline, comparing the largest mean RD on the dynamic simulation grid (blue) with the maximum over all $\varepsilon \in [0, 1]$ when user and creator states remain fixed at each population's uniform equilibrium (hatched). Each pair uses the same ten populations and subtracts the same uniform baseline. The bars are descriptive maxima of mean curves, shown without confidence intervals; both gains are zero for MovieLens user response.

Figure 5 shows a fixed-state intermediate maximum in the same seven settings that have a dynamic intermediate maximum. In the creator-response regime, the dynamic and fixed-state gains are respectively 0.0487 and 0.0014 for MovieLens, and 0.0212 and 0.0018 for synthetic inputs. Under coupled response they are 0.0307 and 0.0057 for MovieLens, and 0.0366 and 0.0147 for synthetic inputs. Under synthetic weak response, the fixed-state gain is 0.0431, compared with 0.0558 for the dynamics; the fixed-state maximum gain is larger under synthetic user response (0.0580 versus 0.0507). The comparison includes changes in both user and creator states in the dynamic runs, and shows how their joint evolution can affect the gain beyond that obtained by changing recommendation fractions at fixed states.

C.5 Additional metrics, robustness, and state snapshots

Figure 6 shows RD, CD, RR, and exposure Gini across all regimes. The Gini coefficient, defined in Appendix A.1, is included as a secondary diagnostic of how exposure is divided across creators, whereas RD measures variation in the content received by each user. At $\varepsilon = 1$ every creator receives equal total exposure and Gini is zero, even though RD differs substantially across response regimes. The separate CD and RD curves also distinguish variation in available content from variation within each user's recommendations.

Figure 7 reports all five MovieLens fits and all five synthetic ensembles. The qualitative pattern persists in every replicate: synthetic RD has an intermediate grid maximum in each regime, while MovieLens user response is maximized at uniform recommendation. For MovieLens, the absolute RD gain ranges from approximately 0.0486 to 0.0501 under creator response and from 0.0303 to 0.0312 under coupled response. Synthetic coupled-response gains range from 0.0348 to 0.0366. Each line averages its own ten populations, with coefficient noise paired across replicates and population IDs also paired across MovieLens fits.

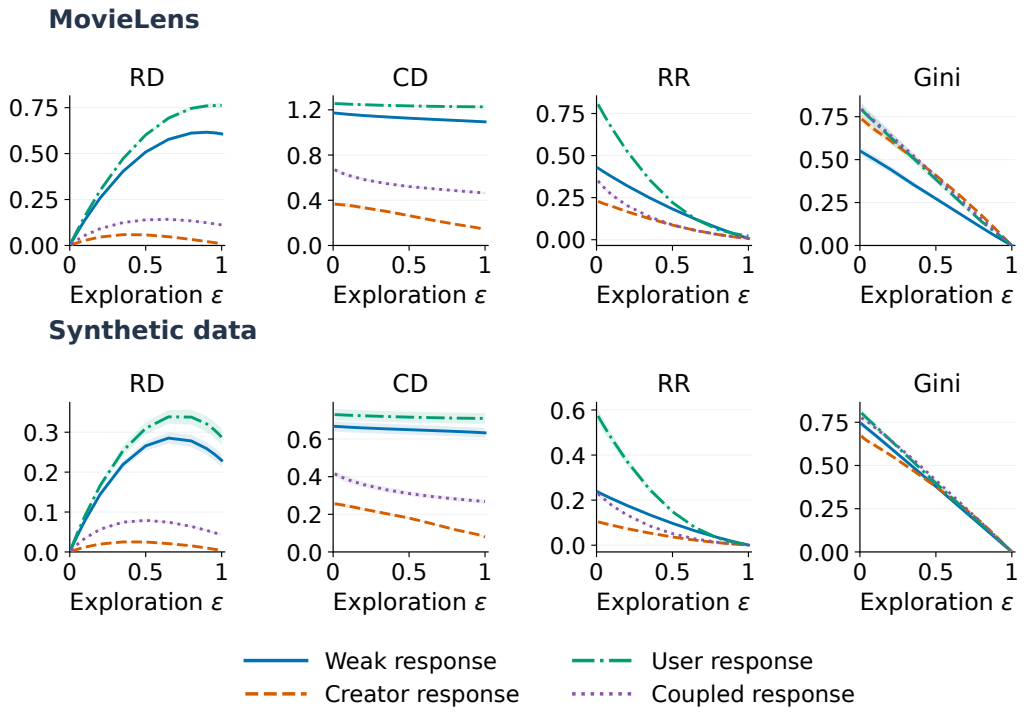


Figure 6: Long-run recommendation diversity (RD), creator diversity (CD), recommendation relevance (RR), and exposure Gini for MovieLens (top) and synthetic inputs (bottom), with response regimes distinguished by color and line style. Lines connect means over ten populations at 13 rates from 0.01 to 1, with pointwise 95% Student-t bands; periodic outcomes enter the means through their period averages. Each panel uses its own vertical scale.

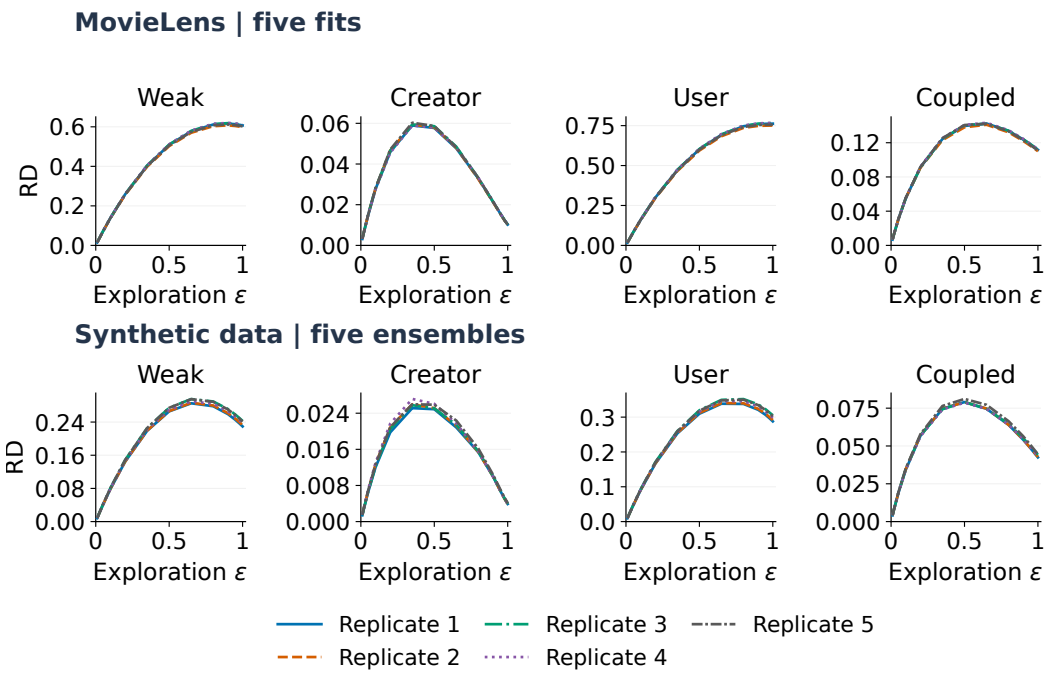


Figure 7: Recommendation diversity (RD) across five MovieLens fits (top) and five synthetic ensembles (bottom), separated by response regime. Each curve connects means over ten populations at 13 rates from 0.01 to 1, using period averages for periodic outcomes. Replicates remain separate without a pooled uncertainty band, and panels use different RD scales.

The temporal plots in Figures 3 and 4 recompute every update through step 500 at $\varepsilon \in \{0.01, 0.50, 0.65, 1\}$; they do not freeze a trajectory when the convergence detector triggers. All 80 paths underlying the temporal plots reach numerically fixed states by step 500. Their terminal metrics agree with the corresponding long-run records within 6.6×10^{-10}. Figure 8 complements these averages with the first population, seed 2026091701, under coupled response. Each column shows its common initial state and the states at step 500 for the four exploration rates, placing users and creators in the same plane. The plots use the first two coordinates of each 16-dimensional representation; MovieLens uses the alignment in (42). Axis limits are shared, with no additional projection fitting. They illustrate the states; all metrics are computed in all 16 dimensions.

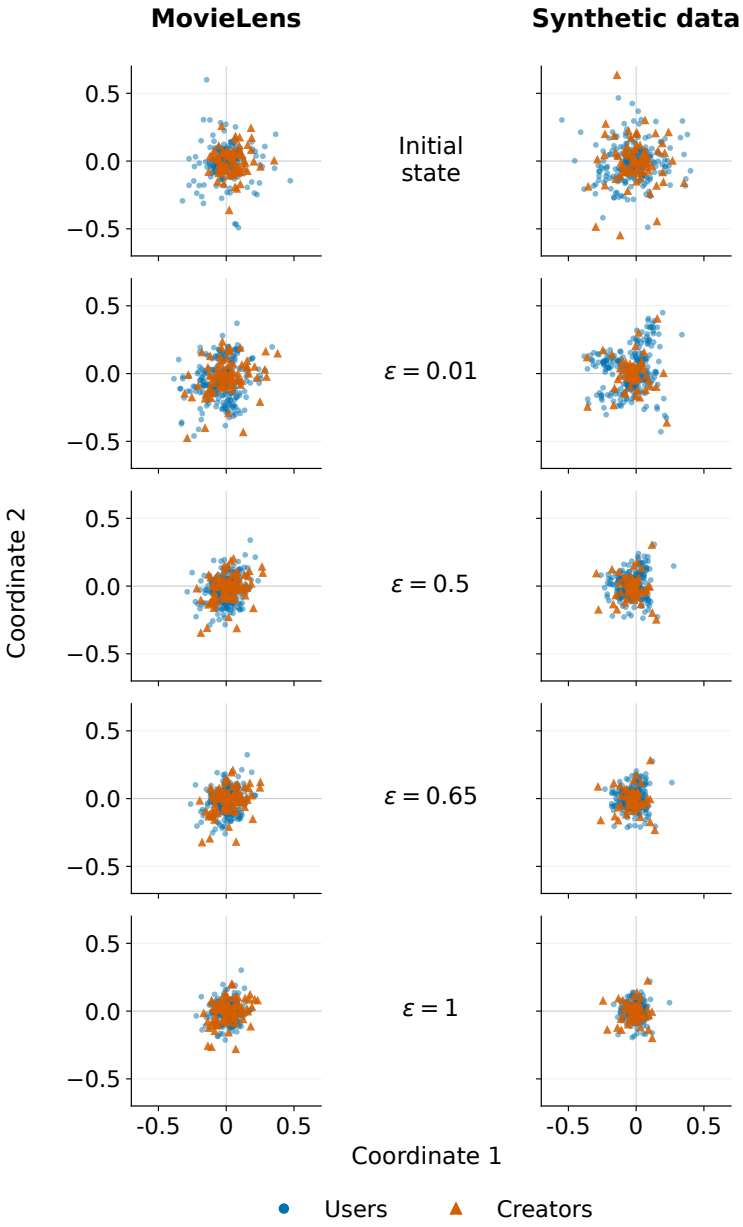


Figure 8: Two-coordinate views of 16-dimensional user and creator states under coupled response, with columns showing the first MovieLens and synthetic populations. Within each column, rows show the shared initial state and the states after 500 updates at four exploration rates. All panels share axis limits for coordinates 1 and 2, while reported metrics use all 16 coordinates.