Recommendation Quality and the Concentration of Consumption: Experimental Evidence from Netflix*

Guy Aridor Winston Chou Nathan Kallus

Antoine Scheid Allen Tran Kevin Zielnicki

August 21, 2026

Abstract

We study an experiment with 8.5 million users on Netflix's recommender system to measure how improvements in recommendation technology affect the set of products that get consumed. Improvements increase total consumption and users' reliance on recommendations while diffusing recommendations and consumption away from the most popular titles ("superstars") toward a larger number of moderately popular titles ("middle-tail"), with minimal effects on the most niche titles ("long-tail"). Our results challenge the notion that recommender systems polarize consumption – raising the consumption shares of the head and tail at the expense of the middle – and suggest that the returns to investing in middle-tail products grow as algorithms improve and platforms scale.

*Contact: Aridor: Netflix and Northwestern Kellogg, garidor@netflix.com. Chou: Netflix, wchou@netflix.com. Kallus: Netflix and Cornell University and Cornell Tech, nkallus@netflix.com. Scheid: Netflix, ascheid@netflix.com. Tran: Netflix, atran@netflix.com, Zielnicki: Netflix, kzielnicki@netflix.com. All authors were employed at Netflix during the duration of this project. We thank Rafael Jiménez-Durán and Malika Korganbekova for helpful feedback and suggestions. We thank audiences at AI Algorithms and Markets Symposium (LUISS), Cornell Johnson, Conference on AI and Human Decisions (CMU), and Research Roundtable on Platform Dynamics (Northwestern) for helpful comments. All errors are our own.

1 Introduction

Recommender systems (RecSys), ubiquitous in the digital economy, use historical interaction data to connect users with relevant products (Resnick and Varian, 1997). Their efficacy depends on extracting meaningful signals about consumer preferences from limited data. As this technology has improved, it has reshaped which products are consumed, with a large literature debating whether it concentrates consumption among the most popular products or spreads it across a long tail of niche products (Brynjolfsson et al., 2003; Salganik et al., 2006; Fleder and Hosanagar, 2009; Calvano et al., 2025), with the products in the middle losing share to the extremes (Bar-Isaac et al., 2012). However, the state-of-the-art algorithms have improved dramatically over the past several decades – from item-based collaborative filtering algorithms (Sarwar et al., 2001), to matrix factorization techniques (Bennett and Lanning, 2007; Koren et al., 2009), to deep learning-based recommenders (He et al., 2017), and, more recently, to sequence and foundation models (Quadrana et al., 2018; Zhao et al., 2024; Hsiao et al., 2025). These improvements continuously expand the frontier of goods that a RecSys can effectively target, yet the implications of these improvements for the concentration of consumption remain unknown.

This paper studies how improvements to recommendations reshape viewership using a unique experiment on Netflix's RecSys involving more than 8.5 million subscribers. The experiment, a "long-term holdback," compares a frozen algorithm with an evolving production treatment incorporating all algorithmic innovations deployed during a 60-day experimental period. We find that ongoing improvements to the RecSys increase overall engagement on the platform and users' reliance on the recommendations for plays, primarily by shifting the recommendations – and subsequently plays – away from the most popular titles ("superstars") and toward moderately popular titles ("middle-tail"), with minimal effects on the long-tail. As recommendation technology matures, then, the plausible set of targetable titles continues to shift toward the middle-tail.

We argue that these results are consistent with a stylized model in which algorithmic improvements allow the RecSys to better infer consumer preferences from a fixed amount of title-level interaction data. Since less popular titles generate fewer interactions, the model captures this data constraint as a tendency for the RecSys to default toward popular titles when preference signals for less popular titles are weak.[1] Improvements in recommendation technology attenuate this bias by allowing the RecSys to identify relevant titles with less data. The gains are largest for middle-tail titles: unlike superstars, they are not already reliably matched, and unlike long-tail titles, they generate enough data and command a large enough audience for better inference to matter. Improvements therefore shift recommendations toward these better-matched titles, raising overall engagement.

This paper contributes to a long-standing literature on how digitization has reshaped the types of goods that get consumed (Goldfarb and Tucker, 2019), and in particular to conflicting findings on whether recommendation systems concentrate or disperse consumption. Several early studies document that online consumption shifted away from popular titles (Brynjolfsson et al., 2003; Zentner et al., 2013), with a dramatic increase in the number of new products brought to market and substantial welfare gains (Waldfogel, 2017; Aguiar and Waldfogel, 2018; Reimers and Waldfogel, 2026). Experimental evidence comparing personalized recommendations to non-personalized benchmarks similarly finds that personalization can increase engagement while dispersing aggregate consumption across products (Holtz et al., 2020). On the other hand, these technologies reduce search costs and thus concentrate consumption among superstars (Rosen, 1981)

[1]The microfoundation for this assumption comes from the literature in recommender systems documenting a *popularity bias* – an inherent tendency of deployed algorithms to favor popular titles because they learn from interaction data, which popular titles generate far more of (Celma and Cano, 2008; Abdollahpouri et al., 2017; Klimashevskaia et al., 2024).

and polarize the product distribution at the expense of middle-tail titles (Bar-Isaac et al., 2012). Empirically, Salganik et al. (2006) shows that popularity information can increase superstar concentration, while Fleder and Hosanagar (2009) and Calvano et al. (2025) show that popularity bias in classic RecSys algorithms can concentrate aggregate consumption on popular products.

These distributional effects are not fixed properties of recommendation systems. Tucker and Zhang (2011) show that popularity information reinforces superstar concentration under homogeneous preferences but can benefit niche products when preferences are heterogeneous. Our contribution is to identify a second, dynamic dimension – algorithm maturity and the volume of title-level interaction data – that changes which titles benefit as recommendation quality improves.[2] In our empirical context, Zielnicki et al. (2026) show using a structural model that previous RecSys improvements lift engagement by better identifying users with high incremental consumption from recommendations, and that this value is largest for middle-tail titles. We provide experimental evidence that continued improvements to the RecSys further shift consumption toward precisely this part of the distribution: away from superstars and toward middle-tail titles, while lifting engagement overall.

These findings have important implications for online platforms. By identifying which titles become more targetable as recommendation quality improves, our results speak to the frontier of products that benefit the most from algorithmic recommendation. In particular, continued improvements in recommendation technology mainly amplify products whose audiences can be reached through better targeting, but whose demand is difficult to realize under less mature systems or platforms with limited scale. This interpretation echoes supply-side evidence that digitization and streaming have increasingly supported middle-tail products: Benner and Waldfogel (2023) document the sustainable production of medium-budget "middle-tail" films, while Lüke (2025) finds that middle-tail books are overrepresented in online bookstores relative to brick-and-mortar retailers.

2 Conceptual Model

To tie together our empirical findings on the effects of improving recommendation technology, we provide a stylized conceptual model.

2.1 Setup

We consider a representative user i whose true utility from consuming title j is given by:

$$U_{ij} = \mu_j + \theta_{ij},$$

where $\mu_j \in \mathbb{R}$ is baseline popularity and θ_{ij} is drawn i.i.d. across (i, j) with bounded, strictly positive density on $\mathbb{R}$ and finite mean. Thus, each title's utility is a combination of a vertical component (μ_j) and a horizontal, idiosyncratic component (θ_{ij}), so a randomly selected user would be more likely to prefer a more popular than a less popular title, but for some users a less popular title can provide higher utility.

We consider a stylized setup where there are three "representative" titles – long-tail (L), middle-tail (M),

[2] Tucker (2023) highlights that algorithmic systems systematically underserve those for whom interaction data are sparse – in our context, this manifests as a built-in disadvantage for less popular titles, consistent with broader evidence of diminishing returns to data (Bajari et al., 2019; Peukert et al., 2024). Our results show that this disadvantage is not permanent but diminishes as the algorithm matures.

and superstar (S) – and the user consumes at most one title.[3] We assume that $\mu_L < \mu_M < \mu_S$ and that user i has a title j that provides the highest utility – we denote this by $j_i^* = \arg\max_j U_{ij}$. Under this formulation, $p_j := \Pr(j_i^* = j) \in (0,1)$ denotes the probability that title j is user i's best title. Thus, the goal of the RecSys is to identify this title.

User Choice. The representative user receives a singleton recommendation $\mathcal{R}_i \subset \{L, M, S\}$ and chooses from the three titles and an outside option. The user's decision utility V_{ij} is influenced by the recommendation via a "utility bonus" $b > 0$ that makes them more likely to consume the recommended title. The bonus captures, in reduced form, the role of recommendations in reducing informational frictions (Aridor et al., 2023; Zielnicki et al., 2026). The user's decision utility for title j is therefore given by:

$$V_{ij} \coloneqq U_{ij} + b \cdot \mathbb{1}\{j \in \mathcal{R}_i\},$$

where $b > 0$. The user chooses the title with the highest decision utility or the outside option (normalized to $V_{i0} = 0$), and we write the choice as $c_i = \text{argmax}_{j \in \{0,S,L,M\}} V_{ij}$.[4]

Platform Recommendation Technology. The platform's recommendation technology is indexed by a single parameter δ and we make the following assumptions about it.

Assumption 1 (Recommendation Technology)**.** *The RecSys attempts to identify the user's best title with a classifier whose accuracy is summarized by the index δ. We assume:*

- ***Popularity default.*** *Conditional on $j_i^* = j$ and the user's full utility, the recommender identifies j with probability $q_j(\delta)$ and otherwise outputs S (i.e., for user $i, \Pr(\mathcal{R}_i = \{j\} \mid U_i, j_i^\star = j) = q_j(\delta)$).*
- ***Monotonicity.*** *The functions q_L, q_M are differentiable and weakly increasing in δ.*
- ***Middle-tail dominance.*** *Over the empirically relevant range of δ, the marginal precision gain is largest for the middle-tail: $p_M\, q'_M(\delta) > p_L\, q'_L(\delta) \geq 0$.*

The primary justification for this assumption comes from the *popularity bias* widely documented in the recommendation systems literature (Celma and Cano, 2008; Klimashevskaia et al., 2024), by which deployed algorithms over-recommend popular titles because less popular titles generate too little data to extract a reliable signal. Appendix C motivates the assumption with a Bayesian recommender that generates this bias endogenously and shows the popularity bias and, over the relevant range of δ, that the weighted identification gain $p_j q'_j(\delta)$ is largest for the middle-tail: the superstar's $q'_S(\delta)$ is small, since it is already well targeted, and while the long-tail's $q'_L(\delta)$ may be sizeable, too few users rank it first for the identification gain to matter.

2.2 Empirical Predictions

Our experimental intervention involves an increase in the effectiveness of recommendation technology (δ) and we characterize directional predictions from this change, holding all other parameters fixed. Each prediction relies on Assumption 1 with proofs deferred to Appendix B.[5]

[3] While the treatment can, in principle, shift the popularity distribution, we validate in Appendix Figure A.1 that the treatment in our empirical context is incremental enough that it does not dramatically change the title classification.

[4] We assume that, as our experiment is conducted over 60 days, the experimental intervention only changes *which* titles enter into $\mathcal{R}_i$ and not how users internalize the signal provided by the recommendations. Thus, the value of b remains constant throughout the experiment.

[5] The empirical measures and predictions are those that are derived from our theory – they are not the core product metrics that Netflix optimizes for.

The first natural consequence of the improvement in recommendation technology is that it shifts recommendation share away from the popularity default and to the tail of the distribution. We define the recommendation probability $f_j(\delta) \coloneqq \Pr(j \in \mathcal{R}_i(\delta))$ and characterize how it changes:

Prediction 1 (Hump-shaped recommendations)**.** *An increase in δ results in a hump-shaped redistribution of recommendation share with the largest gains for the middle-tail (M): $f'_M(\delta) > f'_L(\delta) > f'_S(\delta)$.*

Intuitively, Prediction 1 follows since some users prefer M to S, but the platform observes too noisy of a signal for θ_{iM} in order to recommend M over S. Increasing δ provides additional precision for the platform's estimated score for title M, leading to an increased fraction of users who prefer M to receive it in the recommendations instead of S. A similar pattern holds for users who prefer L, but the magnitude is smaller since the popularity gap between L and S is bigger and the signal is still noisy even after the technology improvement.

If recommendations change (Prediction 1) and recommendations matter for choice ($b > 0$), then naturally consumption should shift to the newly recommended titles. We denote $\pi_j(\delta) \coloneqq \Pr(c_i = j)$ as the consumption probability for title j and characterize how it changes across the three titles:

Prediction 2 (Play Share Rotation)**.** *An increase in δ shifts consumption from the superstars to the middle-tail and long-tail: $\pi'_M(\delta) > 0$ and $\pi'_L(\delta) \geq 0$, while $\pi'_S(\delta) < 0$.*

Predictions 1 and 2 establish that recommendations and consumption shift toward the middle-tail, but redistribution alone does not reveal whether users are better served: an algorithm could diversify recommendations across the middle-tail without those being titles users want to consume (Chen et al., 2024). The key test is whether the redistribution raises overall *engagement* – the likelihood that a user does not take the outside option – since users consume the newly recommended titles only if they prefer them to the outside option:

Prediction 3 (Engagement Levels)**.** *An increase in δ increases total engagement: $\Pr(c_i \neq 0)$ is strictly increasing in δ.*

The main force behind this prediction is that some users who preferred M or L over S were previously recommended S due to the popularity bias, which led them to consume nothing; as the RecSys shifts to their preferred title, they consume it and the outside-option share falls. Because this newly realized consumption comes from the recommended title, the share of consumption originating from recommendations should rise as well:

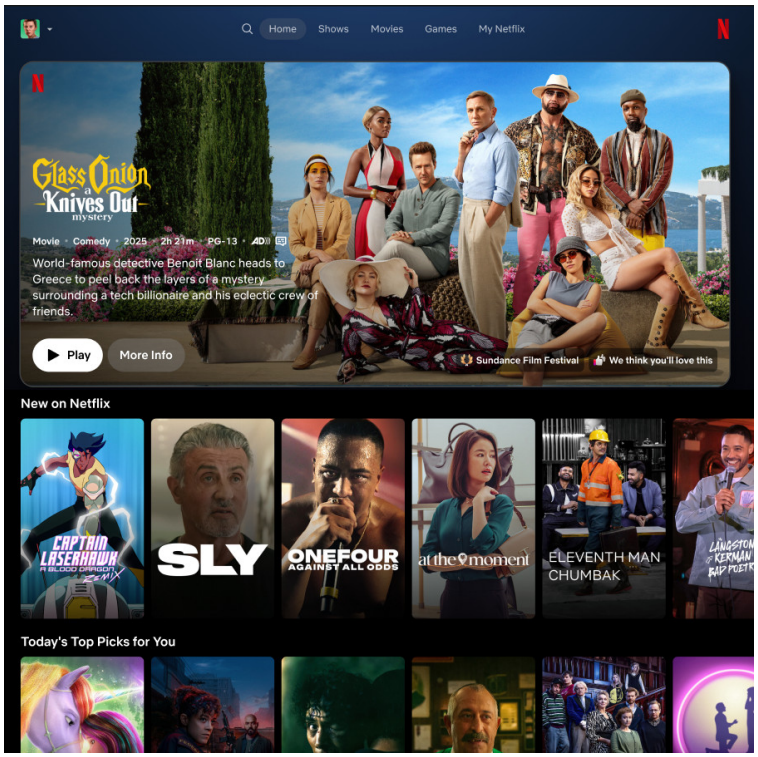
Prediction 4 (Share of Plays from Recommendation)**.** *An increase in δ results in a larger share of consumption originating from recommendations: $\Pr(c_i \in \mathcal{R}_i \mid c_i \neq 0)$ is strictly increasing in δ.*

Finally, it is important to understand whether, conditional on consumption, the improved recommendations lead users to better-matched titles. This is theoretically ambiguous in our model as there are two counteracting forces: *inframarginal switching* may increase average match quality if the improved recommendations shift a user's consumption from S to their preferred title, but *extensive-margin entry* may decrease average match quality by inducing users who would otherwise consume the outside option to consume marginal titles. This leads to our final prediction (which is formally stated in Appendix B), which ultimately will be an empirical question to understand which force wins out:

Prediction 5 (Match Quality)**.** *An increase in δ results in weakly better average match quality if and only if the match-quality gains from inframarginal switching weakly exceed the composition effect from extensive-margin entry.*

3 Data and Field Experiment

Figure 1: Netflix Homepage



Data and Empirical Setting. Our empirical context is the video-streaming platform Netflix. Users pay a subscription fee to access Netflix and can then costlessly play any title on the platform. We therefore operationalize consumption in our empirical context as playing a title, and define engagement as the overall volume of consumption. The set of titles consists of television shows (which have multiple episodes and seasons) and movies. The platform's homepage – which is personalized to each user – is shown in Figure 1. Users can directly play titles from the recommendations (defined as titles on the homepage), continue watching titles they had started before, or manually search for titles using the search bar. Our analysis uses platform data on plays and recommendations for each user in the experiment.

Field Experiment. To evaluate the empirical implications of our model, we analyze an experiment on Netflix on 8,559,252 subscribers over a 60-day period from February to April 2025. This experiment was designed as a "cumulative holdback" test to measure the collective impact of multiple innovations to personalization algorithms during this period. The control arm of the experiment corresponds to the state of the recommendation system prior to the start of the test, whereas the treatment arm corresponds to the "production" state of Netflix that is experienced by the majority of Netflix members and continuously updated during the test as new innovations (e.g., algorithmic updates) are released. These releases are held back from the control group, which continues to experience the pre-experiment version of the recommendation system. Such holdback tests are run periodically on a small fraction of the Netflix member base in order to measure long-term progress on personalization (Kohavi et al., 2020).

The algorithmic innovations deployed during the experiment encompass twelve major changes spanning roughly four categories: dedicated rankers for specific homepage rows, feature engineering, revised model

architectures such as modeling certain outcomes jointly rather than independently, and different model weights on prediction targets. Most were deployed near the start of the intervention, with a smaller number released closer to the end. These innovations are purely algorithmic and do not change the user interface, thus map to increases in δ in the language of our model.

This experiment represents an ideal testing ground for our model for several reasons. First, the treatment includes all innovations made to personalization algorithms during the experiment timeframe. As such, it is broadly representative of ongoing efforts to improve personalized recommendations at Netflix. Second, because each subrelease is individually A/B tested and chosen based on evidence of a better member experience, we can be highly confident that the treatment represents a significant improvement to personalized recommendations. Lastly, the relatively long duration of the experiment helps us confirm that the following impacts are stable and persistent.

4 Distributional Shifts in Recommendations and Plays

In this section we characterize the *distributional* shifts as a result of the technology increase by first exploring its impact on the distribution of recommendations (Section 4.1) and then its subsequent impact on the distribution of played titles (Section 4.2).

A key challenge is measuring distributional impacts without letting our intervention contaminate the play distribution itself. Unfortunately, as the environment is dynamic with many titles entering and exiting the platform, pre-experimental shares are not representative of shares during the experiment. As such, we partition titles into the same buckets as in the conceptual model and assign them based on their percentile on the aggregate play distribution across the treatment and control groups: "long-tail" (L), "middle-tail" (M), or "superstar" (S) buckets – where "superstar" is defined as the top 5 percentile, the "long-tail" as the bottom 50 percentile, and the "middle-tail" as the set of titles in between.[6]

Our primary specification measures how treatment reallocates a user's recommendations and plays across title buckets. Let c^o_{ik} be the total number of instances for outcome o for a given user i and titles in bucket k. The outcome of interest is the within-user share $s^o_{ik} = c^o_{ik} / \sum_{k'} c^o_{ik'}$ for users with at least one outcome. We estimate, separately for each outcome o and each bucket k:

$$s^o_{ik} = \alpha^o_k + \beta^o_k T_i + \varepsilon^o_{ik}, \tag{1}$$

where T_i is an indicator for user i being assigned to the treatment arm, α^o_k is the mean share in the control and β^o_k is the average treatment effect on that share. We compute robust standard errors.

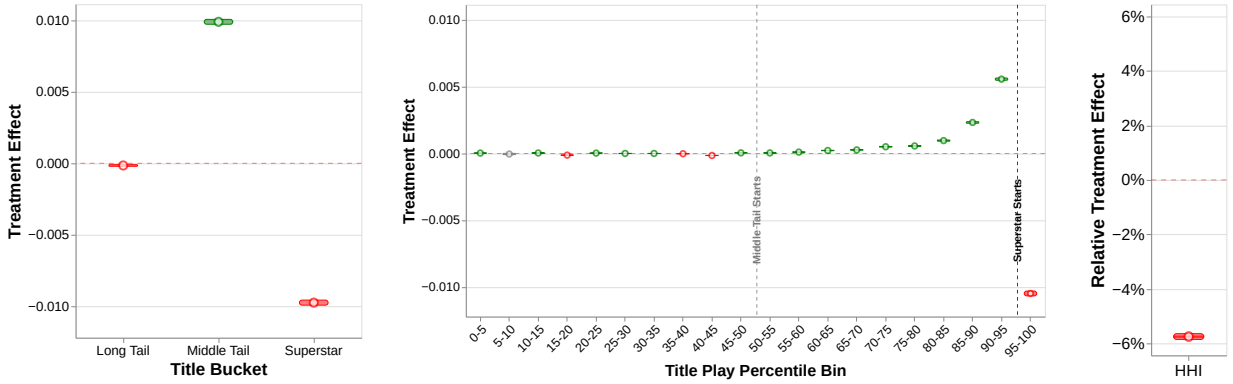
4.1 Recommendations Shift to the Middle-Tail

The first implication of an increase in the efficacy of recommendation technology, as stated in Prediction 1, is that recommendations redistribute away from superstars toward the middle-tail, with minimal effects on the long-tail. We test this by estimating specification (1) on the recommendations users observe during the experiment for each of the title buckets $\{L, M, S\}$.

Figure 2a confirms the hump shape predicted by Prediction 1: recommendation share shifts away from superstar titles and primarily gets redirected to middle-tail titles, with quantitatively small changes to the

[6] While the treatment can influence the empirical popularity distribution, we empirically confirm in Figure A.1 that these buckets remain similar between our treatment and control groups.

Figure 2: Treatment Effects on Recommendation Shares



(a) Shares by Bucket (b) Shares by Ventile Bins (c) HHI

NOTES: Estimated Average Treatment Effects (ATEs) on the distribution of recommendations. Panel (a) presents coefficient estimates from (1) with the dependent variable of the share of an individual's recommendations accounted for by each bucket. Panel (b) presents treatment effects on recommendation share within each pooled ventile, where the ventiles are defined using the pooled play distribution across the treatment and control groups. Panel (c) presents coefficient estimates for the changes in the recommendations HHI, where we compute standard errors using a delete-one jackknife. Bars denote 95% confidence intervals.

long-tail. Figure 2b shows that this same pattern persists using more granular ventile buckets and Appendix Figure A.3 shows that these patterns replicate when we take the most conservative approach of defining the percentiles purely in terms of the play percentile in the control group.[7]

An important consequence of this redistribution is that it shifts recommendation share from a small number of superstar titles and disperses it to a large number of middle-tail titles that are better fits for users. As such, the shift to the middle-tail should decrease the amount of concentration in recommendations across users. We estimate the effects of the technology improvement on the Herfindahl–Hirschman Index ($\text{HHI} = \sum_{j=1}^{N} s_j^2$) – a standard measure for the degree of concentration that we compute at the title level – to test this. Figure 2c confirms that HHI has a statistically and economically significant decrease of 5.7%.

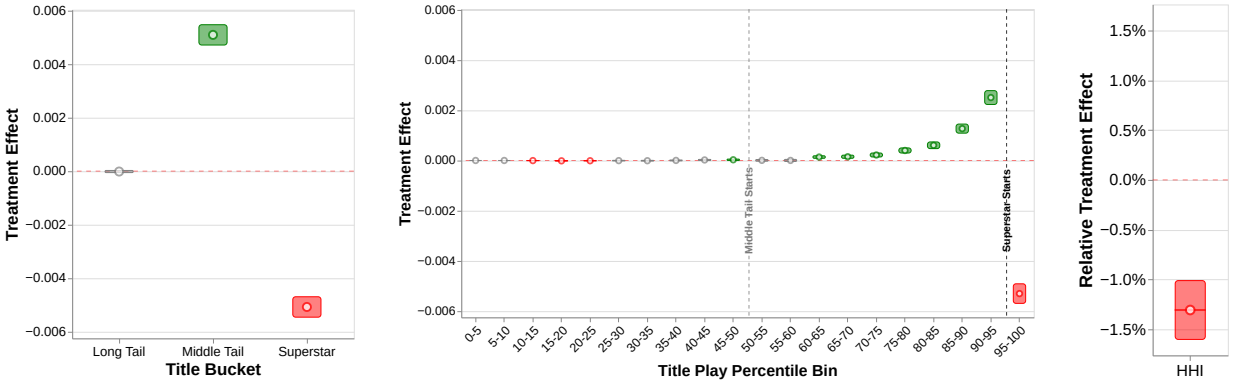
4.2 Plays Shift to the Middle-Tail

A natural implication of the shift in the distribution of recommendations is that it may also shift the distribution of plays. As shown in our conceptual model, plays depend on the title's recommendation probability *and* the probability that the user follows the recommendation. If the RecSys diversified the recommendations by replacing superstar titles with middle-tail titles that were not appealing, then users could still seek out the superstars via search or decide not to play anything on the platform at all, leading to minimal differences in the play share. Prediction 2 indicates that play share shifts if the recommendations are sufficiently consequential in user choices (i.e., $b > 0$).

We test Prediction 2 by estimating specification (1) for each title bucket, reporting the estimates in Figure 3a. The results confirm the prediction: superstar play share falls with most of the redistribution going to the middle-tail, with the long-tail quantitatively unchanged. Figure 3b shows that this same pattern persists using more granular ventile buckets and Appendix Figure A.5 shows that these patterns replicate when we take the most conservative approach of defining the percentiles purely in terms of the play

[7] Appendix Figure A.2 additionally shows that within the superstars the effect is most stark for the highest percentile.

Figure 3: Treatment Effects on Play Shares



(a) Shares by Bucket (b) Shares by Ventile Bins (c) HHI

NOTES: Estimated ATEs on the distribution of title plays during the experiment. Panel (a) presents coefficient estimates from specification (1) with the dependent variable of the share of an individual's plays accounted for by each bucket. Panel (b) presents treatment effects on play share for each pooled ventile, where ventiles are defined using the pooled play distribution across the treatment and control groups. Panel (c) presents coefficient estimates for the changes in the Herfindahl-Hirschman Index of plays, where we compute standard errors using a delete-one jackknife. Bars denote 95% confidence intervals.

percentile in the control group.[8]

Finally, we measure whether this corresponds to a reduction in play concentration: Figure 3c shows that HHI falls by a statistically and economically significant 1.2%, consistent with viewership dispersing from superstar titles to more personalized middle-tail titles.

5 Engagement Levels and Match Quality

While the results in Section 4 document that both recommendations and plays shift toward the middle-tail and subsequently become less concentrated, the key test of whether this shift is a "technological improvement" is if these distributional shifts translate into a corresponding improvement in the consumer experience that is driven by the recommendations. We assess this in this section by first characterizing whether and why the improvement leads to an increase in the *overall level* of engagement (Section 5.1) and then assessing whether the played titles had similar match quality (Section 5.2).

For most of this analysis, we compare *user-level* outcomes across the experimental period and estimate average treatment effects using the following regression:

$$Y_i = \alpha + \beta T_i + \varepsilon_i, \tag{2}$$

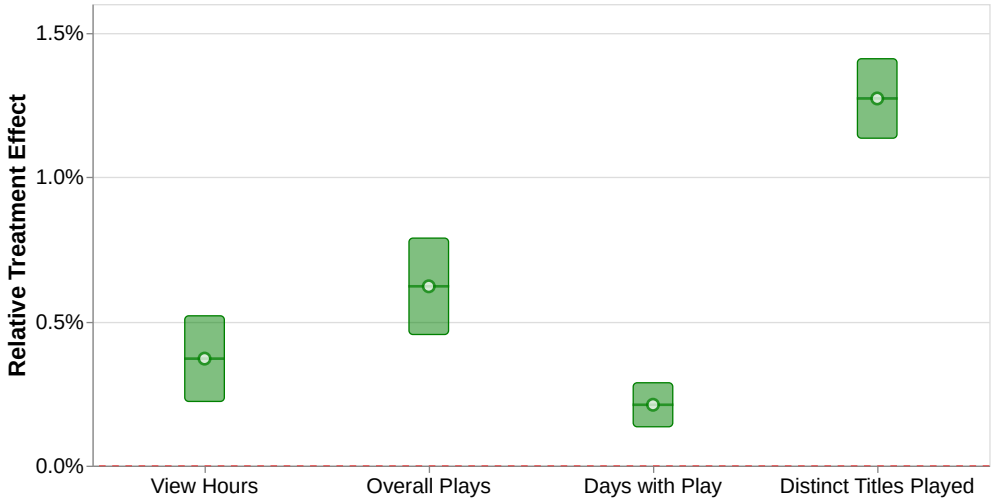
with similar notation as before and compute robust standard errors.

5.1 Overall Engagement and Recommendation Dependence

The simple conceptual model from Section 2 predicts that the recommendations should be more relevant to consumers, inducing increased engagement (Prediction 3) that is driven by increased play share coming from the recommendations (Prediction 4). In this section, we test these predictions.

[8] Appendix Figure A.4 additionally shows that within the superstars the effect is most stark for the highest percentile.

Figure 4: Treatment Effects on Overall Engagement



NOTES: Estimated ATEs from (2) on account-level engagement outcomes: view hours, overall plays, days with at least one play, and distinct titles played. All effects are reported as relative percent change against the control mean. Bars denote 95% confidence intervals computed via robust standard errors.

Figure 4 presents the estimates of β from (2) across four different measures of engagement: total view hours, number of plays, the number of days with at least one play, and the number of distinct titles played. The results confirm Prediction 3: view hours increased by 0.37%, the total number of titles played by 0.62%, the distinct number of titles played by 1.2%, and the number of days with at least one play by 0.21%.[9,10] As context for the magnitudes, Zielnicki et al. (2026) estimate via a structural model that reverting to the predominant RecSys algorithm from 10 years ago (matrix factorization) would reduce days with a play by 4%, with economic effects comparable to those of reverting from matrix factorization to popularity-based ranking (then valued at more than $1 billion by Gomez-Uribe and Hunt (2015)). Our estimate of the effects of roughly two months of algorithmic innovation – a 0.21% increase in days with a play – represents about 5% of this magnitude. Crucially, it implies that the shift away from superstars documented in Section 4 is not diversification for diversification's sake, but actually coincides with improved engagement.

The intuition from our conceptual model for *why* the improvement leads to an increase in overall engagement is simple: the RecSys replaces less well-matched superstars with better matched middle-tail titles which results in users being more likely to find content to play from the recommender system. In our empirical context, we can validate this by measuring whether the reliance on the recommendations for generating plays increases – which is the crux of Prediction 4.

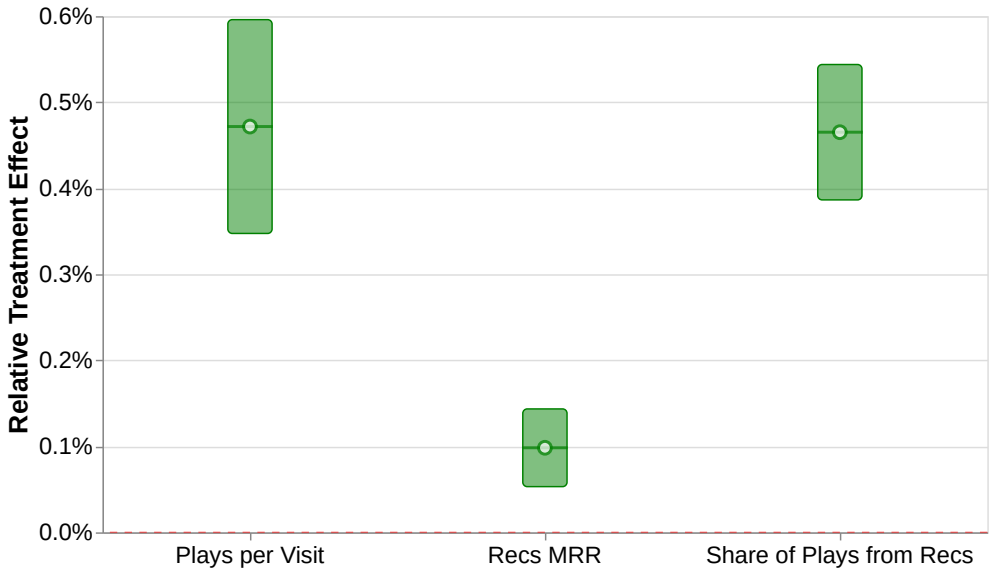
To test Prediction 4, we restrict focus to the first play per session in order to rule out effects coming from continuation play and classify each play as either coming from recommendations, manual search, play continuation, or previously saved (e.g., my list). We then estimate (2) for the ratio of plays to visits, the share of plays that originate from recommendations, and the mean reciprocal rank (MRR) of plays.[11]

[9]For television shows, total and distinct title counts can diverge, since each episode is counted separately toward the total but does not count as a distinct title. Appendix Figure A.7 shows that the increase in engagement is for both TV shows and movies, but leads to larger increases for TV shows.

[10]Appendix Figure A.6 shows that the treatment effects are stable or increasing, indicating sustained improvements in engagement throughout the intervention period.

[11]MRR measures the reciprocal of the row number on the homepage from which the play was derived and is a common

Figure 5: Treatment Effects on Reliance on Recommendations



NOTES: Estimated ATEs from (2) for measures of reliance on recommendations: plays per visit, mean reciprocal rank (MRR), and the share of first-in-session plays sourced from recommendations rather than other navigation. MRR is computed using the reciprocal of the row at which the played title appears on the homepage. All effects are reported as relative percent change against the control mean. Bars denote 95% confidence intervals computed using robust standard errors.

Figure 5 shows that the ratio of plays to visits increases by 0.47%, the share of plays originating from the recommendations increases by 0.5%, and MRR increases by 0.1% – all consistent with Prediction 4.

5.2 Match Quality

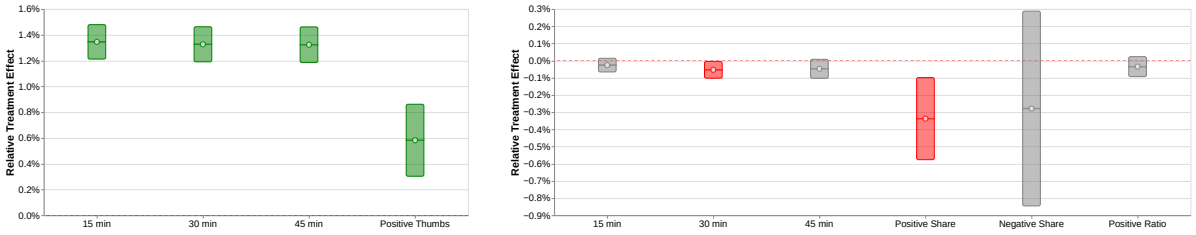
We now evaluate Prediction 5 – how match quality changes due to the treatment. We rely on two proxies for match quality. The first is whether users continue watching the titles they start. As titles have different runtimes, we denote a title as *completed* if the user watched it for at least $\min\{X, \text{runtime}\}$ minutes for varying values of X on the first day of viewing the title. The second proxy is whether a user provided a positive "thumbs" rating to a title. We restrict ourselves to titles for which the user was shown the thumbs prompt, which asks the user for a negative thumb (thumbs down) or a positive thumb (single or double thumbs up).

First, we measure whether users find more titles that they complete or thumb positively. We estimate (2) on the number of titles satisfying each of these measures and report the results in Figure 6a: users complete 1.3% more titles at each of the $X = \{15, 30, 45\}$-minute thresholds and provide 0.57% more positive thumbs.

A natural question – and the crux of Prediction 5 – is whether the average title that a user plays is of comparable match quality. For each user, we compute the share of these outcomes out of their plays and estimate (2) for each. Figure 6b shows statistically significant and negative effects on the probability of completing at least 30 minutes and on the share of positive thumbs, precise nulls on the probabilities of the 15- and 45-minute completion shares and the positive thumbs ratio, and insignificant effects on the share of

measure used to assess recommendation quality (Gunawardana and Shani, 2009) with a larger value indicating that the played title was ranked higher in the recommendations.

Figure 6: Treatment Effects on Match Quality



(a) Total Number of Played Titles

(b) Fraction of Played Titles

NOTES: Estimated ATEs from (2) on match-quality measures: shares of titles thumbed positively/negatively, and shares watched at least 15, 30, or 45 minutes (or full runtime, if shorter) on the first viewing day. Watch-duration shares use all plays; thumbs shares are restricted to plays long enough to trigger the thumbs prompt. Panel (a) shows estimates for the total titles per user reaching the watch-time threshold or receiving a positive thumbs; Panel (b) shows the user-level fraction of titles meeting each duration, receiving a positive/negative thumbs, or the positive-to-negative thumbs ratio. Bars are 95% CIs from robust standard errors.

negative thumbs. From our conceptual model, these results are a combination of the counteracting forces of extensive-margin entry and inframarginal switching.[12] While we parse these two effects in the Appendix, overall, our results indicate that users found more "high quality" titles to play and that the average title is of comparable quality, especially once we account for extensive-margin selection into additional plays due to the treatment.

6 Conclusion

This paper studied how advances in recommendation technology reallocate the concentration of consumption. Using a unique experiment on Netflix's recommender system encompassing 8.5 million users, we show that such improvements redistribute plays from a small number of superstar titles to a larger number of middle-tail titles while also increasing overall engagement and user reliance on recommendations. These results stand in contrast to previous work that hypothesized that increased personalization would lead to an increasing consumption share of long-tail and superstar titles at the expense of the middle-tail. Our model provides a rationalization for this: in markets with horizontal preference heterogeneity, improving the RecSys allows it to extract more precise signals out of limited data, which attenuates the popularity bias inherent in existing algorithms.

These results have important managerial implications, as they suggest that the returns of middle-tail products to platforms depend on algorithmic maturity and the ability of the RecSys to exploit limited data. As recommendation technology continues to improve, it becomes more feasible for platforms to target an increasingly broad set of middle-tail titles that earlier generations of recommendation technologies were less able to personalize. This frontier remains bounded, however: long-tail titles generate very sparse interaction data, making targeting difficult for even mature technologies. Furthermore, it may be difficult to predict *ex ante* which titles will fall into this margin. Even so, our findings reframe the long-running debate over

[12] To isolate the latter, we temporally order each user's played titles and plot the match quality measures for the k-th played title in Appendix Figure A.8. Users play titles of lower match quality as they play more titles, suggesting the negative point estimates could be due to playing marginal titles that users would not have played without the improved recommendations. To conservatively account for extensive-margin entry and isolate the effects of inframarginal switching, we compute Lee bounds (Lee, 2009) on match quality for the k-th played title in Figure A.9 which suggest that we cannot consistently sign whether the newly played titles are higher or lower quality. The assumption of monotonicity required for the validity of Lee bounds – treatment inducing additional consumption compared to the control – is reasonable given the results in Section 5.1.

whether recommendation systems concentrate or disperse consumption: the answer is neither fixed nor universal but evolves with algorithmic maturity. At the current frontier, continued improvement lifts overall engagement and reallocates it toward the middle-tail.

References

Abdollahpouri, H., R. Burke, and B. Mobasher (2017). Controlling popularity bias in learning-to-rank recommendation. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*, pp. 42–46.

Aguiar, L. and J. Waldfogel (2018). Quality predictability and the welfare benefits from new products: Evidence from the digitization of recorded music. *Journal of Political Economy 126*(2), 492–524.

Aridor, G., D. Goncalves, D. Kluver, R. Kong, and J. Konstan (2023). The economics of recommender systems: Evidence from a field experiment on movielens. In *Proceedings of the 24th ACM Conference on Economics and Computation*, pp. 117–117.

Bajari, P., V. Chernozhukov, A. Hortaçsu, and J. Suzuki (2019). The impact of big data on firm performance: An empirical investigation. *AEA Papers and Proceedings 109*, 33–37.

Bar-Isaac, H., G. Caruana, and V. Cuñat (2012). Search, design, and market structure. *American Economic Review 102*(2), 1140–1160.

Benner, M. J. and J. Waldfogel (2023). Changing the channel: Digitization and the rise of "middle tail" strategies. *Strategic Management Journal 44*(1), 264–287.

Bennett, J. and S. Lanning (2007). The netflix prize.

Brynjolfsson, E., Y. Hu, and M. D. Smith (2003). Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management Science 49*(11), 1580–1596.

Calvano, E., G. Calzolari, V. Denicolo, and S. Pastorello (2025). Artificial intelligence, algorithmic recommendations and competition.

Celma, Ò. and P. Cano (2008). From hits to niches? or how popular artists can bias music recommendation and discovery. In *Proceedings of the 2nd KDD Workshop on Large-Scale Recommender Systems and the Netflix Prize Competition*, pp. 1–8.

Chen, G., T. Chan, D. Zhang, S. Liu, and Y. Wu (2024). The impact of a more diversified recommender system on digital platforms: Evidence from a large-scale field experiment.

Fleder, D. and K. Hosanagar (2009). Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management Science 55*(5), 697–712.

Goldfarb, A. and C. Tucker (2019). Digital economics. *Journal of Economic Literature 57*(1), 3–43.

Gomez-Uribe, C. A. and N. Hunt (2015). The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS) 6*(4), 1–19.

Gunawardana, A. and G. Shani (2009). A survey of accuracy evaluation metrics of recommendation tasks. *Journal of Machine Learning Research 10*(12).

He, X., L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua (2017). Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web*, pp. 173–182.

Holtz, D., B. Carterette, P. Chandar, Z. Nazari, H. Cramer, and S. Aral (2020). The engagement-diversity connection: Evidence from a field experiment on spotify. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pp. 75–76.

Hsiao, K.-J., Y. Feng, and S. Lamkhede (2025). Foundation model for personalized recommendation. *Netflix Tech Blog*.

Klimashevskaia, A., D. Jannach, M. Elahi, and C. Trattner (2024). A survey on popularity bias in recommender systems. *User Modeling and User-Adapted Interaction 34*(5), 1777–1834.

Kohavi, R., D. Tang, and Y. Xu (2020). *Trustworthy online controlled experiments: A practical guide to a/b testing*. Cambridge University Press.

Koren, Y., R. Bell, and C. Volinsky (2009). Matrix factorization techniques for recommender systems. *Computer 42*(8), 30–37.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies 76*(3), 1071–1102.

Lüke, D. (2025). Tales of tails: sales distribution and the role of retail channels in the german book market. *Journal of Cultural Economics*, 1–23.

Peukert, C., A. Sen, and J. Claussen (2024). The editor and the algorithm: Recommendation technology in online news. *Management Science 70*(9), 5816–5831.

Quadrana, M., P. Cremonesi, and D. Jannach (2018). Sequence-aware recommender systems. *ACM Computing Surveys 51*(4), 1–36.

Reimers, I. and J. Waldfogel (2026). Ai and the quantity and quality of creative products: Have llms boosted creation of valuable books?

Resnick, P. and H. R. Varian (1997). Recommender systems. *Communications of the ACM 40*(3), 56–58.

Rosen, S. (1981). The economics of superstars. *American Economic Review 71*(5), 845–858.

Salganik, M. J., P. S. Dodds, and D. J. Watts (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science 311*(5762), 854–856.

Sarwar, B., G. Karypis, J. Konstan, and J. Riedl (2001). Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International Conference on World Wide Web*, pp. 285–295.

Tucker, C. (2023). Algorithmic exclusion: The fragility of algorithms to sparse and missing data. *Report, Brookings Center on Regulation and Markets, Washington, DC*, 1–26.

Tucker, C. and J. Zhang (2011). How does popularity information affect choices? a field experiment. *Management Science 57*(5), 828–842.

Waldfogel, J. (2017). How digitization has created a golden age of music, movies, books, and television. *Journal of Economic Perspectives 31*(3), 195–214.

Zentner, A., M. Smith, and C. Kaya (2013). How video rental patterns change as consumers move online. *Management Science 59*(11), 2622–2634.

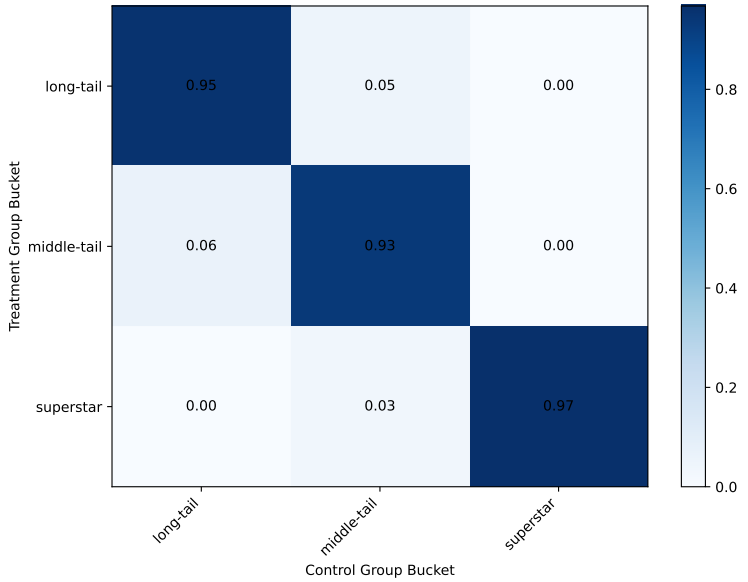
Zhao, Z., W. Fan, J. Li, Y. Liu, X. Mei, Y. Wang, Z. Wen, F. Wang, X. Zhao, J. Tang, et al. (2024). Recommender systems in the era of large language models (llms). *IEEE Transactions on Knowledge and Data Engineering 36*(11), 6889–6907.

Zielnicki, K., G. Aridor, A. Bibaut, A. Tran, W. Chou, and N. Kallus (2026). The value of personalized recommendations: Evidence from netflix. *International Journal of Industrial Organization*, 103304.

Online Appendix

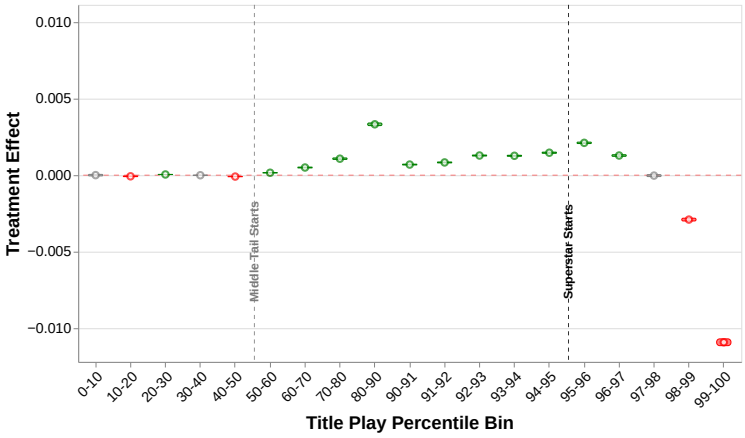
A Additional Results, Figures and Tables

Figure A.1: Confusion Matrix of Title Bucket Assignment across Treatments



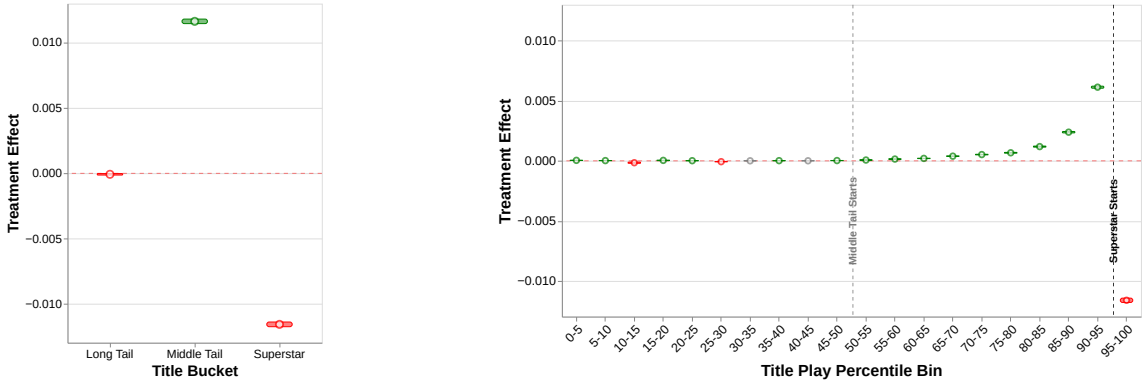
NOTES: This figure shows the confusion matrix for title categorization into long-tail, middle-tail, and superstars. For each treatment group, we compute each title's percentile rank in the empirical play distribution within that group. We then classify titles as long-tail if they fall in the bottom 50% of the distribution, middle-tail if they fall in the next 45%, and superstars if they fall in the top 5%. The confusion matrix shows the agreement in categorization across groups.

Figure A.2: Recommendation Percentile-Bin Treatment Effects



NOTES: Estimated ATEs on recommendation share within percentile buckets defined using the pooled play distribution across the treatment and control groups.

Figure A.3: Treatment Effects on Recommendation Shares (Control Group Ordering)

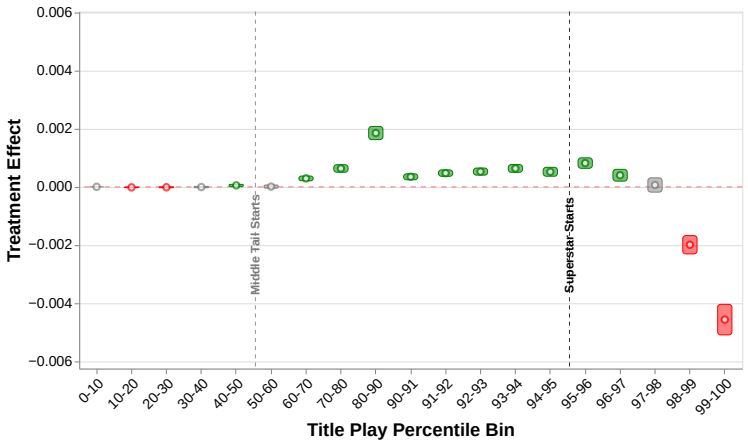


(a) Shares by Bucket

(b) Shares by Ventile Bins

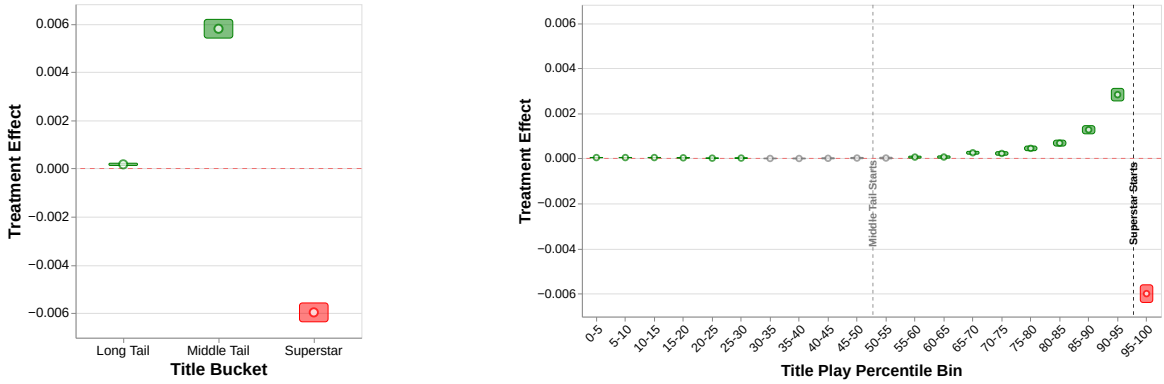
NOTES: Estimated ATEs on the distribution of recommendations. Panel (a) presents coefficient estimates from (1) with the dependent variable of the share of an individual's recommendations accounted for by each bucket. Panel (b) presents treatment effects on recommendation share within each ventile. In both panels the buckets and ventiles are defined using the control group play distribution. Bars denote 95% confidence intervals.

Figure A.4: Play Percentile-Bin Treatment Effects



NOTES: Estimated ATEs on play share within the percentile buckets defined using the pooled play distribution across the treatment and control groups.

Figure A.5: Treatment Effects on Play Shares (Control Group Ordering)

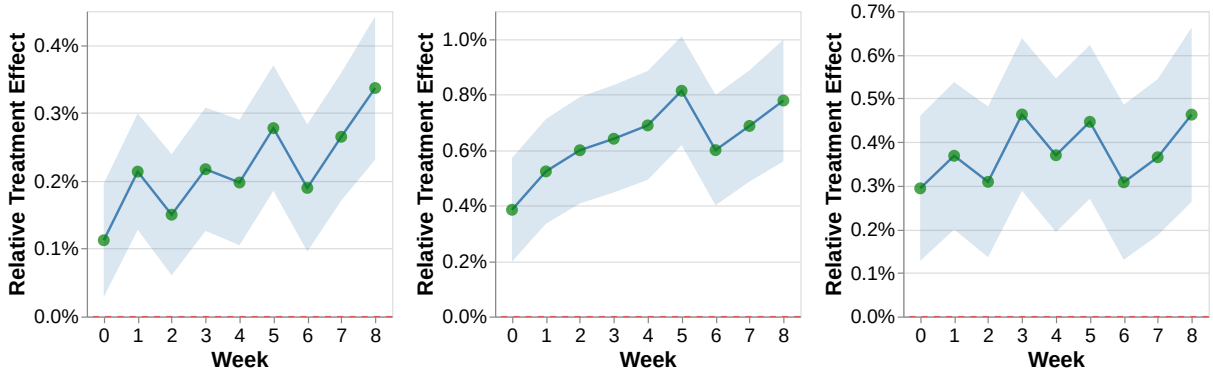


(a) Shares by Bucket

(b) Shares by Ventile Bins

NOTES: Estimated ATEs on the distribution of title plays during the experiment. Panel (a) presents coefficient estimates from specification (1) with the dependent variable of the share of an individual's plays accounted for by each bucket. Panel (b) presents treatment effects on play share for each pooled ventile. In both panels the buckets and ventiles are defined using the control group play distribution. Bars denote 95% confidence intervals.

Figure A.6: Overall Engagement Treatment Effects over Time



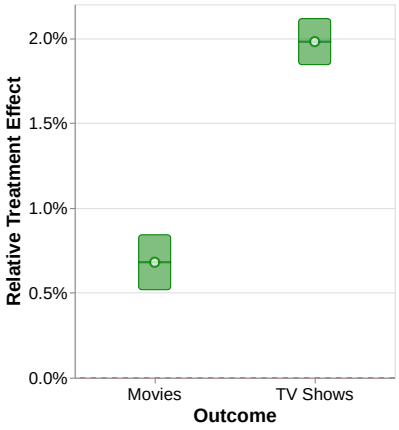
(a) Days with Play

(b) Overall Plays

(c) View Hours

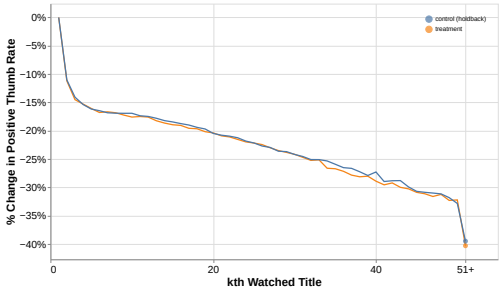
NOTES: Estimated ATEs from specification (2) for the three primary engagement outcomes – days with play, overall plays, and view hours – estimated separately each week of the experimental intervention. Shaded bands denote 95% confidence intervals computed via robust standard errors.

Figure A.7: Engagement by Title Type

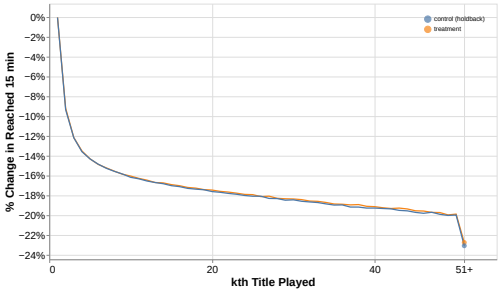


NOTES: The figure decomposes the distinct-titles-played in Figure 4 by content type, reporting relative treatment effects on the number of distinct movies and the number of distinct TV shows watched per account. Bars denote 95% confidence intervals computed via robust standard errors.

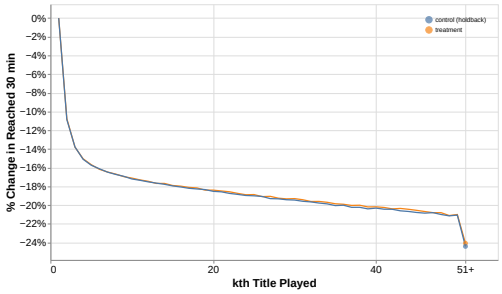
Figure A.8: Match Quality for First-k Plays



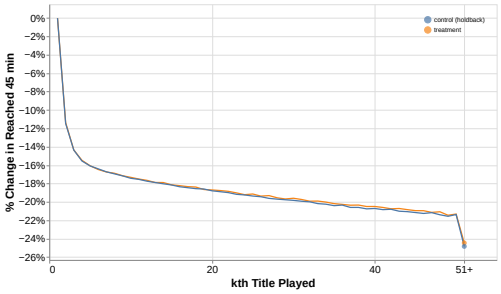
(a) Positive Thumb Rate



(b) Share Watched At Least 15 Minutes



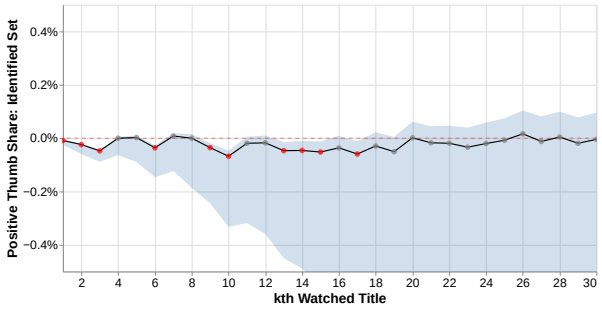
(c) Share Watched At Least 30 Minutes



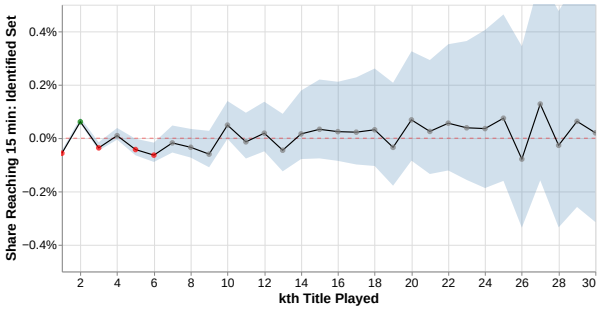
(d) Share Watched At Least 45 Minutes

NOTES: The figure plots the underlying level of positive thumb rate (Panel a) and the share of plays watched at least 15, 30, and 45 minutes during the first day of viewing (Panels b–d) against the exact kth title played, one line per treatment arm, shown up to $k = 50$ and then a single pooled 51+ point. The positive thumb rate is only computed for titles that were watched sufficiently long to have the thumbs prompt appear. TV shows are considered one title, regardless of how many episodes were watched. For each measure, we report percent deviations from the values for $k = 1$.

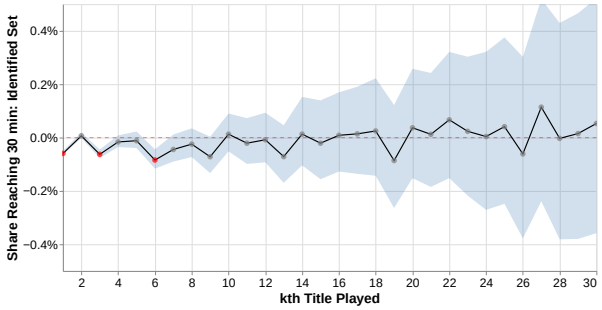
Figure A.9: Lee Bounds on Match Quality for First-k Plays



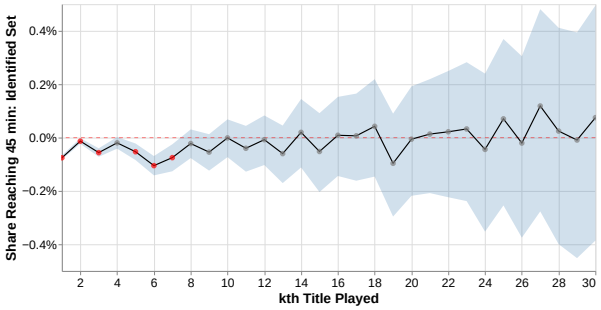
(a) Positive Thumb Share



(b) Share Watched At Least 15 Minutes



(c) Share Watched At Least 30 Minutes



(d) Share Watched At Least 45 Minutes

NOTES: The figure presents the estimated Lee bounds (Lee, 2009) for each k-th play on positive thumb share (Panel a) and the share of plays watched at least 15, 30, and 45 minutes during the first day of viewing (Panels b–d). For panel a, k indexes the account's k-th distinct title across all content that was watched sufficiently long to trigger the prompt for querying whether a user dislikes the title (negative thumb) or likes the title (single or double positive thumbs up). For panels b–d, k indexes any title that was started. TV shows are considered one title, regardless of how many episodes were watched. The shaded area represents the Lee bounds for each k.

B Omitted Proofs

Note that for all the proofs, because the utility shocks have a continuous density, ties occur with probability zero.

Proof of Prediction 1. For any parameter δ, and any title $k \in \{L, M, S\}$, the law of total probability applied to the partition $\{j_i^\star = j\}_{j\in\{L,M,S\}}$ gives

$$f_k(\delta) = \Pr(k \in \mathcal{R}_i(\delta)) = \sum_{j\in\{L,M,S\}} \Pr(k \in \mathcal{R}_i(\delta) \mid j_i^\star = j)\, p_j. \tag{3}$$

By Assumption 1 (popularity default), conditional on $j_i^\star = j$ the recommender outputs either j (with probability $q_j(\delta)$) or the superstar S (with probability $1 - q_j(\delta)$).

Writing (3) for $k \in \{L, M, S\}$, we obtain

$$f_L(\delta) = p_L \, q_L(\delta), \tag{4}$$
$$f_M(\delta) = p_M \, q_M(\delta), \tag{5}$$
$$f_S(\delta) = p_S + p_M \, (1 - q_M(\delta)) + p_L \, (1 - q_L(\delta)). \tag{6}$$

(4) holds since L is recommended only when $j_i^\star = L$, and similarly for (5). (6) reflects the fact that S is recommended both when $j_i^\star = S$ (in which case $\Pr(S \in \mathcal{R}_i(\delta) \mid j_i^\star = S) = 1$) and when $j_i^\star \in \{L, M\}$ in the case where the recommender defaults to S (which happens respectively with probability $(1 - q_L(\delta))$ and $(1 - q_M(\delta))$).

By Assumption 1 (monotonicity), we can differentiate f_L, f_M, and f_S with respect to δ, which gives that

$$f'_M(\delta) - f'_L(\delta) = p_M \, q'_M(\delta) - p_L \, q'_L(\delta) > 0,$$
$$f'_S(\delta) - f'_L(\delta) = -p_M \, q'_M(\delta) - p_L \, q'_L(\delta) - p_L q'_L(\delta) < 0,$$

where both inequalities follow from Assumption 1, hence the result. □

Proof of Prediction 2. We defined $\pi_k(\delta) = \Pr(c_i = k)$ as the probability that user i consumes title k. Applying the law of total probability to the partition $\{j_i^\star = j\}_{j \in \{L,M,S\}}$, we can write

$$\pi_k(\delta) = \sum_{j \in \{L,M,S\}} p_j \, \Pr(c_i = k \mid j_i^\star = j).$$

For $j \in \{L, M\}$, conditional on $j_i^\star = j$ the recommender outputs j with probability $q_j(\delta)$ and S with probability $1 - q_j(\delta)$; for $j = S$, the recommender always outputs S regardless of δ. Therefore, we have that for $k \in \{L, M, S\}$

$$\Pr(c_i = k \mid j_i^\star = j) = q_j(\delta) \, \Pr(c_i = k \mid j_i^\star = j, \mathcal{R}_i = \{j\}) + (1 - q_j(\delta)) \Pr(c_i = k \mid j_i^\star = j, \mathcal{R}_i = \{S\}).$$

The conditional consumption probabilities $\Pr(c_i = k \mid j_i^\star = j, \mathcal{R}_i = \{r\})$ depend only on $(U_{iL}, U_{iM}, U_{iS}, b)$, not on δ. Differentiating, we obtain

$$\pi'_k(\delta) = \sum_{j \in \{L,M\}} p_j \, q'_j(\delta) \big[\Pr(c_i = k \mid j_i^\star = j, \mathcal{R}_i = \{j\}) - \Pr(c_i = k \mid j_i^\star = j, \mathcal{R}_i = \{S\}) \big]. \tag{7}$$

We use a key observation to simplify (7). Conditional on $j_i^\star = j$, we have $U_{ij} \geq U_{ik}$ for all k. Under $\mathcal{R}_i = \{j\}$, the bonus on j reinforces this ranking, so the user consumes either j or the outside option. Under $\mathcal{R}_i = \{S\}$, the bonus on S can lift S above j but cannot lift any third title above j, so the user consumes either j, S, or the outside option. In particular, for $k \notin \{j, S\}$,

$$\Pr(c_i = k \mid j_i^\star = j, \mathcal{R}_i = \{j\}) = \Pr(c_i = k \mid j_i^\star = j, \mathcal{R}_i = \{S\}) = 0.$$

The bracketed term in (7) is therefore nonzero only when $k \in \{j, S\}$.

This observation partitions the analysis into two structurally distinct cases. When k is a non-default title ($k \in \{L, M\}$), only the $j = k$ term in (7) survives: the only way a precision gain changes the consumption of k is by correctly identifying a k-type user and redirecting them from the popularity default S to their best

title k. When k is the popularity default ($k = S$), both $j = L$ and $j = M$ terms contribute: every precision gain redirects mass away from S as some L- and M-type users are correctly identified and shifted away. We handle these two cases in turn.

Case $k \in \{L, M\}$. Only the $j = k$ term contributes:

$$\pi'_k(\delta) = p_k\, q'_k(\delta)\big[\Pr(c_i = k \mid j_i^\star = k, \mathcal{R}_i = \{k\}) - \Pr(c_i = k \mid j_i^\star = k, \mathcal{R}_i = \{S\})\big].$$

Conditional on $\{j_i^\star = k\}$, the bonus on k under $\mathcal{R}_i = \{k\}$ makes $c_i = k$ weakly more likely than under $\mathcal{R}_i = \{S\}$, where the bonus instead pushes consumption toward S. To establish strict inequality, consider the event $\{j_i^\star = k,\, -b < U_{ik} < 0,\, U_{iS} + b < 0\}$, which has positive measure by the bounded, strictly positive density of θ. On this event:

- Under $\mathcal{R}_i = \{S\}$, $U_{iS} + b < 0$ and $U_{ik} < 0$, so the user takes the outside option and $c_i = 0$.
- Under $\mathcal{R}_i = \{k\}$, $V_{ik} = U_{ik} + b > 0$, so the user consumes k.

Hence the bracketed term is strictly positive. By Assumption 1, $p_M q'_M(\delta) > 0$ and $p_L q'_L(\delta) \geq 0$, hence $\pi'_M(\delta) > 0$ and $\pi'_L(\delta) \geq 0$.

Case $k = S$. Both $j = L$ and $j = M$ contribute:

$$\pi'_S(\delta) = \sum_{j \in \{L,M\}} p_j\, q'_j(\delta)\big[\Pr(c_i = S \mid j_i^\star = j, \mathcal{R}_i = \{j\}) - \Pr(c_i = S \mid j_i^\star = j, \mathcal{R}_i = \{S\})\big].$$

Conditional on $j_i^\star = j$ for $j \in \{L, M\}$, the bonus under $\mathcal{R}_i = \{j\}$ is on j rather than S, so S is weakly less likely to be consumed than under $\mathcal{R}_i = \{S\}$. Each bracketed term is nonpositive.

The $j = M$ term is strictly negative. Consider the event $\{j_i^\star = M,\, 0 < U_{iM} - U_{iS} < b,\, U_{iS} + b > 0\}$, which has positive measure by the bounded, strictly positive density of θ. On this event:

- Under $\mathcal{R}_i = \{S\}$, the bonus lifts $U_{iS} + b$ above U_{iM} (since $U_{iM} - U_{iS} < b$) and above U_{iL} (since $U_{iL} \leq U_{iM}$) and above 0 (since $U_{iS} + b > 0$), so the user consumes S.
- Under $\mathcal{R}_i = \{M\}$, the bonus is on M instead, so $V_{iM} = U_{iM} + b > V_{iS} = U_{iS}$, and the user does not consume S.

Hence $\Pr(c_i = S \mid j_i^\star = M, \mathcal{R}_i = \{S\}) > \Pr(c_i = S \mid j_i^\star = M, \mathcal{R}_i = \{M\})$, and the $j = M$ bracketed term in $\pi'_S(\delta)$ is strictly negative. Since $p_M q'_M(\delta) > 0$ by Assumption 1, we obtain $\pi'_S(\delta) < 0$. □

Proof of Prediction 3. The overall strategy is to decompose engagement by the user's best title $j_i^\star$ and show that improvements in δ raise the engagement of L- and M-type users by redirecting them from the popularity default S to their true best title, where the recommendation bonus $b > 0$ makes consumption more likely.

Let $E(\delta) \coloneqq \Pr(c_i \neq 0)$ denote the engagement probability. Applying the law of total probability to the partition $\{j_i^\star = j\}_{j \in \{L,M,S\}}$, we can write

$$E(\delta) = \Pr(c_i \neq 0) = \sum_{j \in \{L,M,S\}} p_j\, \Pr(c_i \neq 0 \mid j_i^\star = j).$$

The S-type term contributes nothing to $E'(\delta)$: when $j_i^\star = S$, the popularity default coincides with the user's best title, hence $\mathcal{R}_i = \{S\}$ with probability one regardless of δ. Therefore $\Pr(c_i \neq 0 \mid j_i^\star = S) = \Pr(U_{iS} + b > 0 \mid j_i^\star = S)$ is independent of δ.

The main consequence of the increase in δ comes from the L- and M-type terms, where δ shifts the recommendation away from the default S toward the user's actual best title.

For $j \in \{L, M\}$, conditional on $j_i^\star = j$ the recommender outputs j with probability $q_j(\delta)$ and S with probability $1 - q_j(\delta)$:

$$\Pr(c_i \neq 0 \mid j_i^\star = j) = q_j(\delta)\,\Pr(c_i \neq 0 \mid j_i^\star = j, \mathcal{R}_i = \{j\}) + (1 - q_j(\delta))\,\Pr(c_i \neq 0 \mid j_i^\star = j, \mathcal{R}_i = \{S\}).$$

The conditional consumption probabilities given $\mathcal{R}_i$ are functions of $(U_{iL}, U_{iM}, U_{iS}, b)$ alone and do not depend on δ. Differentiating with respect to δ gives that

$$\frac{d}{d\delta}\Pr(c_i \neq 0 \mid j_i^\star = j) = q_j'(\delta)\big[\Pr(c_i \neq 0 \mid j_i^\star = j, \mathcal{R}_i = \{j\}) - \Pr(c_i \neq 0 \mid j_i^\star = j, \mathcal{R}_i = \{S\})\big].$$

Since the S-type term in $E(\delta)$ is constant in δ, multiplying by p_j and summing over $j \in \{L, M\}$ gives that

$$E'(\delta) = \sum_{j \in \{L,M\}} p_j\, q_j'(\delta)\big[\Pr(c_i \neq 0 \mid j_i^\star = j, \mathcal{R}_i = \{j\}) - \Pr(c_i \neq 0 \mid j_i^\star = j, \mathcal{R}_i = \{S\})\big]. \tag{8}$$

Each term measures the engagement gain from redirecting a j-type user's recommendation away from S toward j, weighted by the marginal precision gain $p_j q_j'(\delta)$.

We now show that each redirection gain is strictly positive. For any $j \in \{L, M\}$, conditional on $\{j_i^\star = j\}$, we have $U_{ij} \geq U_{iS}$, thus recommending j (which gets the bonus) gives a higher decision utility V_{ij} than recommending S. The recommendation bonus thus makes consumption weakly more likely under $\mathcal{R}_i = \{j\}$. To get strict inequality, we identify a set of users for whom the recommendation directly flips the consumption decision: those whose true best title j has utility just below the outside option, while S remains below even with the bonus. Since the θ_{ij} have bounded, strictly positive density on $\mathbb{R}$,

$$\Pr\big(-b < U_{ij} \leq 0,\ U_{iS} + b \leq 0 \,\big|\, j_i^\star = j\big) > 0.$$

On this event, $U_{ik} \leq U_{ij} \leq 0$ for all k and $U_{iS} + b \leq 0$, hence under $\mathcal{R}_i = \{S\}$ the user takes the outside option ($c_i = 0$), while under $\mathcal{R}_i = \{j\}$ the bonus pushes $V_{ij} = U_{ij} + b > 0$ and the user consumes j ($c_i \neq 0$). Therefore

$$\Pr(c_i \neq 0 \mid j_i^\star = j, \mathcal{R}_i = \{j\}) > \Pr(c_i \neq 0 \mid j_i^\star = j, \mathcal{R}_i = \{S\}).$$

By Assumption 1, $p_L q_L'(\delta) \geq 0$ and $p_M q_M'(\delta) > 0$. Combining it with the bracketed term in (8) being strictly positive, the L-type term is nonnegative and the M-type term is strictly positive. Therefore $E'(\delta) > 0$. □

Proof of Prediction 4. The strategy is to show that improvements in δ raise the probability of recommendation-driven consumption faster than they raise overall engagement, so the conditional share of consumption coming from recommendations rises.

Let $\rho(\delta) \coloneqq \Pr(c_i \in \mathcal{R}_i \mid c_i \neq 0)$ denote the recommendation share. Since $\{c_i \in \mathcal{R}_i\} \subset \{c_i \neq 0\}$,

$$\rho(\delta) = \frac{\Pr(c_i \in \mathcal{R}_i, c_i \neq 0)}{\Pr(c_i \neq 0)} = \frac{\Pr(c_i \in \mathcal{R}_i)}{E(\delta)}.$$

We first show that $\frac{d}{d\delta}\Pr(c_i \in \mathcal{R}_i) \geq E'(\delta)$. From (8) in the proof of Prediction 3, we can write

$$E'(\delta) = \sum_{j \in \{L,M\}} p_j \, q_j'(\delta) \big[\Pr(c_i \neq 0 \mid j_i^\star = j, \mathcal{R}_i = \{j\}) - \Pr(c_i \neq 0 \mid j_i^\star = j, \mathcal{R}_i = \{S\}) \big]. \tag{9}$$

An analogous decomposition of $\Pr(c_i \in \mathcal{R}_i)$ gives

$$\frac{d}{d\delta}\Pr(c_i \in \mathcal{R}_i) = \sum_{j \in \{L,M\}} p_j \, q_j'(\delta) \big[\Pr(c_i \in \mathcal{R}_i \mid j_i^\star = j, \mathcal{R}_i = \{j\}) - \Pr(c_i \in \mathcal{R}_i \mid j_i^\star = j, \mathcal{R}_i = \{S\}) \big]. \tag{10}$$

We compare the two derivatives term by term. Conditional on $\{j_i^\star = j\}$ and $\mathcal{R}_i = \{j\}$, the user either consumes the recommended title j or takes the outside option, thus the events $\{c_i \in \mathcal{R}_i\}$ and $\{c_i \neq 0\}$ coincide:

$$\Pr(c_i \in \mathcal{R}_i \mid j_i^\star = j, \mathcal{R}_i = \{j\}) = \Pr(c_i \neq 0 \mid j_i^\star = j, \mathcal{R}_i = \{j\}).$$

Conditional on $j_i^\star = j$ and $\mathcal{R}_i = \{S\}$, the user may consume S (the recommended title) or some other title outside the recommendation, thus consuming the recommendation is weakly less likely than consuming anything:

$$\Pr(c_i \in \mathcal{R}_i \mid j_i^\star = j, \mathcal{R}_i = \{S\}) \leq \Pr(c_i \neq 0 \mid j_i^\star = j, \mathcal{R}_i = \{S\}).$$

Subtracting the second from the first, the bracketed term in (10) weakly exceeds the bracketed term in (9). Since the weights $p_j q_j'(\delta)$ are nonnegative with $p_M q_M'(\delta) > 0$,

$$\frac{d}{d\delta}\Pr(c_i \in \mathcal{R}_i) \geq E'(\delta) > 0.$$

Applying the quotient rule, we can write

$$\rho'(\delta) = \frac{\frac{d}{d\delta}\Pr(c_i \in \mathcal{R}_i) \cdot E(\delta) - \Pr(c_i \in \mathcal{R}_i) \cdot E'(\delta)}{E(\delta)^2}.$$

Substituting $\frac{d}{d\delta}\Pr(c_i \in \mathcal{R}_i) \geq E'(\delta)$ and using $E'(\delta) > 0$ from Prediction 3, we can write

$$\rho'(\delta) \geq \frac{E'(\delta) \cdot E(\delta) - \Pr(c_i \in \mathcal{R}_i) \cdot E'(\delta)}{E(\delta)^2} = \frac{E'(\delta)\big[E(\delta) - \Pr(c_i \in \mathcal{R}_i)\big]}{E(\delta)^2}.$$

We finally show $E(\delta) > \Pr(c_i \in \mathcal{R}_i)$, which gives the strict inequality. Consider users with $j_i^\star = M$ and $\mathcal{R}_i = \{S\}$. By the bounded, strictly positive density of θ, there is a positive-measure set of such users with $U_{iM} > \max\{U_{iS} + b, U_{iL}, 0\}$. On this event, the user consumes M — outside the singleton recommendation $\{S\}$ — hence $c_i \neq 0$ while $c_i \notin \mathcal{R}_i$. This event occurs with positive probability whenever $q_M(\delta) < 1$, thus

$$E(\delta) > \Pr(c_i \in \mathcal{R}_i). \tag{11}$$

Combined with $E'(\delta) > 0$, (11) yields $\rho'(\delta) > 0$, and we can conclude that $\Pr(c_i \in \mathcal{R}_i \mid c_i \neq 0)$ is strictly increasing in δ. □

Formal condition for Prediction 5. For any $j \in \{L, M\}$, define the event

$$\mathcal{S}_j \coloneqq \{U_{iS} > -b,\ 0 < U_{ij} - U_{iS} < b\}.$$

Conditional on $\{j_i^\star = j\}$, the event $\mathcal{S}_j$ identifies users who undergo an inframarginal consumption switch when the recommendation is corrected from S to j. The condition $U_{ij} - U_{iS} < b$ means that the recommendation bonus on S overturns the user's true-utility ranking when S is recommended. The condition $U_{iS} > -b$ ensures that the user consumes rather than taking the outside option.

For any $j \in \{L, M\}$, define the event

$$\mathcal{E}_j \coloneqq \{U_{iS} < -b < U_{ij} < 0\}.$$

On this event, $U_{ij} < 0$ and $U_{iS} + b < 0$, hence the user takes the outside option when S is recommended. However, when j is recommended, $U_{ij} + b > 0$, hence the user consumes j.

We now introduce

$$\begin{aligned} r_{\text{inf}}(\delta) &\coloneqq \sum_{j \in \{L,M\}} p_j q_j'(\delta) \Pr(\mathcal{S}_j \mid j_i^\star = j), \\ g_{\text{inf}}(\delta) &\coloneqq \frac{\sum_{j \in \{L,M\}} p_j q_j'(\delta) \mathbb{E}\left[(U_{ij} - U_{iS}) \mathbb{1}\{\mathcal{S}_j\} \mid j_i^\star = j\right]}{r_{\text{inf}}(\delta)}, \\ u_{\text{ext}}(\delta) &\coloneqq \frac{\sum_{j \in \{L,M\}} p_j q_j'(\delta) \mathbb{E}\left[U_{ij} \mathbb{1}\{\mathcal{E}_j\} \mid j_i^\star = j\right]}{E'(\delta)}. \end{aligned} \tag{12}$$

Prediction 3 establishes $E'(\delta) > 0$, and we show in the proof that $r_{\text{inf}}(\delta) > 0$, hence the ratios are well-defined. We refer to $r_{\text{inf}}(\delta)$ as the rate of inframarginal switching (from S to the user's best title), to $g_{\text{inf}}(\delta)$ as the expected true-utility gain conditional on an inframarginal switch, and to $u_{\text{ext}}(\delta)$ as the expected true utility conditional on extensive-margin entry. These expectations aggregate across best-title realizations using the marginal precision gains $p_j q_j'(\delta)$. These definitions weight best-title realizations by the same marginal precision gains $p_j q_j'(\delta)$ that determine which recommendations are marginally corrected.

Let $Q(\delta) \coloneqq \mathbb{E}\left[\sum_{k \in \{L,M,S\}} U_{ik} \mathbb{1}\{c_i = k\}\right]$ denote expected realized true utility, assigning zero utility when the user takes the outside option. Average match quality can then be written as

$$\overline{U}(\delta) = \frac{\mathbb{E}\left[\sum_{k \in \{L,M,S\}} U_{ik} \mathbb{1}\{c_i = k\}\right]}{E(\delta)}.$$

Prediction 5 (Formal). $\overline{U}(\delta)$ is weakly increasing in δ if and only if

$$\underbrace{\frac{r_{\text{inf}}(\delta)}{E'(\delta)}}_{\substack{\text{inframarginal switches} \\ \text{per extensive-margin entry}}} \qquad \underbrace{g_{\text{inf}}(\delta)}_{\substack{\text{true-utility gain per} \\ \text{inframarginal switch}}} \quad \geq \quad \underbrace{\overline{U}(\delta) - u_{\text{ext}}(\delta)}_{\substack{\text{match-quality gap per} \\ \text{extensive-margin entry}}}. \tag{13}$$

Average match quality is strictly increasing when the inequality is strict, locally unchanged when it holds with equality, and decreasing when it is reversed.

Proof of Prediction 5. The overall strategy is to decompose the effect of a marginal increase in δ into an inframarginal change in which title the user consumes and an extensive-margin change in whether the user

consumes. We then apply the quotient rule to the average match quality. Let $c_i(r)$ denote the user's choice when the recommendation is $\mathcal{R}_i = \{r\}$.

Fix $j \in \{L, M\}$ and condition on $j_i^\star = j$. Conditional on $\{j_i^\star = j\}$, we have that $\{U_{ij} > U_{ik}$ for any $k \neq j\}$ almost surely.

When $\mathcal{R}_i = \{j\}$, the recommendation bonus reinforces the user's true-utility ranking, hence the user chooses either j or the outside option:

$$c_i(j) = \begin{cases} j, & \text{if } U_{ij} + b > 0, \\ 0, & \text{if } U_{ij} + b < 0. \end{cases}$$

When $\mathcal{R}_i = \{S\}$, the recommendation bonus can lift S above j, but no third title can be chosen, as shown previously. Therefore

$$c_i(S) = \begin{cases} S, \text{ if } & U_{iS} + b > \max\{U_{ij}, 0\}, \\ j, & \text{if } U_{ij} > \max\{U_{iS} + b, 0\}, \\ 0, & \text{if } 0 > \max\{U_{ij}, U_{iS} + b\}. \end{cases}$$

Conditional on $\mathcal{S}_j$ and $j_i^\star = j$, correcting the recommendation changes which title the user consumes but not whether the user consumes, hence

$$\mathbb{1}\{c_i(j) \neq 0\} - \mathbb{1}\{c_i(S) \neq 0\} = 0. \tag{14}$$

The resulting increase in realized true utility is $U_{ij} - U_{iS} \in (0, b)$.

Conditional on $\mathcal{E}_j$ and $j_i^\star = j$, correcting the recommendation raises engagement by one, following

$$\mathbb{1}\{c_i(j) \neq 0\} - \mathbb{1}\{c_i(S) \neq 0\} = 1, \tag{15}$$

and changes the realized true utility by $U_{ij} \in (-b, 0)$ (decrease).

In all other preference realizations, correcting the recommendation leaves both engagement and realized true utility unchanged. Therefore, it follows pointwise from (14) and (15) that

$$\mathbb{1}\{c_i(j) \neq 0\} - \mathbb{1}\{c_i(S) \neq 0\} = \mathbb{1}\{\mathcal{E}_j\},$$

and we refer to $\mathbb{1}\{\mathcal{E}_j\}$ as the *extensive-margin entry*. The change in utility can be written

$$\sum_{k \in \{L,M,S\}} U_{ik} \mathbb{1}\{c_i(j) = k\} - \sum_{k \in \{L,M,S\}} U_{ik} \mathbb{1}\{c_i(S) = k\} = \underbrace{(U_{ij} - U_{iS})\mathbb{1}\{\mathcal{S}_j\}}_{\substack{\text{true-utility gain from} \\ \text{an inframarginal switch}}} + \underbrace{U_{ij}\mathbb{1}\{\mathcal{E}_j\}}_{\substack{\text{true utility loss from} \\ \text{extensive-margin entry}}}. \tag{16}$$

We now aggregate across the preference realizations whose recommendations are marginally changed by an increase in δ. From (8) in the proof of Prediction 3, we can write

$$E'(\delta) = \sum_{j \in \{L,M\}} p_j \, q_j'(\delta) \big[\Pr(c_i \neq 0 \mid j_i^\star = j, \mathcal{R}_i = \{j\}) - \Pr(c_i \neq 0 \mid j_i^\star = j, \mathcal{R}_i = \{S\}) \big].$$

The difference in engagement probabilities is exactly the probability that correcting the recommendation

induces extensive-margin entry, hence

$$\begin{aligned}\Pr(c_i \neq 0 \mid j_i^\star = j, \mathcal{R}_i = \{j\}) - \Pr(c_i \neq 0 \mid j_i^\star = j, \mathcal{R}_i = \{S\}) &= \Pr(U_{ij} + b > 0, U_{ij} < 0, U_{iS} + b < 0 \mid j_i^\star = j) \\ &= \Pr(U_{ij} > -b, U_{ij} < 0, U_{iS} < -b \mid j_i^\star = j) \\ &= \Pr(U_{iS} < -b < U_{ij} < 0 \mid j_i^\star = j) \\ &= \Pr(\mathcal{E}_j \mid j_i^\star = j),\end{aligned}$$

and we obtain that

$$E'(\delta) = \underbrace{\sum_{j \in \{L,M\}} p_j q_j'(\delta) \Pr(\mathcal{E}_j \mid j_i^\star = j)}_{\text{rate of extensive-margin entry}}.$$

By the strictly positive density of the utility shocks, the event $\{j_i^\star = M, U_{iS} > -b,\ 0 < U_{iM} - U_{iS} < b\} = \{j_i^\star = M, \mathcal{S}_M\}$ has strictly positive probability. Since $p_M q_M'(\delta) > 0$ by Assumption 1 (middle-tail dominance), it follows that $r_{\inf}(\delta) \geq p_M q_M'(\delta) \Pr(\mathcal{S}_M \mid j_i^\star = M) > 0$, where $r_{\inf}(\delta)$ is defined in (12).

Conditional on $\{\mathcal{S}_j, j_i^\star = j\}, 0 < U_{ij} - U_{iS} < b$, hence

$$\begin{aligned}0 < \frac{\sum_{j \in \{L,M\}} p_j q_j'(\delta) \mathbb{E}\left[(U_{ij} - U_{iS}) \mathbb{1}\{\mathcal{S}_j\} \mid j_i^\star = j\right]}{r_{\inf}(\delta)} &< \frac{\sum_{j \in \{L,M\}} p_j q_j'(\delta) \mathbb{E}\left[b \mathbb{1}\{\mathcal{S}_j\} \mid j_i^\star = j\right]}{r_{\inf}(\delta)} \\ &= b \frac{\sum_{j \in \{L,M\}} p_j q_j'(\delta) \Pr(\mathcal{S}_j \mid j_i^\star = j)}{r_{\inf}(\delta)} \\ &= b,\end{aligned}$$

which implies that $0 < g_{\inf}(\delta) < b$. Conditional on $\{\mathcal{E}_j, j_i^\star = j\}, U_{ij} \in (-b, 0)$ and the same algebra as above gives that $-b < u_{\text{ext}}(\delta) < 0$.

Conditional on $j_i^\star = j \in \{L, M\}$, the recommendation is j with probability $q_j(\delta)$ and S with probability $1 - q_j(\delta)$. Hence,

$$\begin{aligned}Q(\delta) = p_S \mathbb{E}\left[\sum_{k \in \{L,M,S\}} U_{ik} \mathbb{1}\{c_i(S) = k\} \mid j_i^\star = S\right] + \sum_{j \in \{L,M\}} p_j q_j(\delta) \mathbb{E}\left[\sum_{k \in \{L,M,S\}} U_{ik} \mathbb{1}\{c_i(j) = k\} \mid j_i^\star = j\right] \\ + \sum_{j \in \{L,M\}} p_j \big(1 - q_j(\delta)\big) \mathbb{E}\left[\sum_{k \in \{L,M,S\}} U_{ik} \mathbb{1}\{c_i(S) = k\} \mid j_i^\star = j\right].\end{aligned}$$

The first term does not depend on δ, and in the remaining terms, only $q_j(\delta)$ depends on δ. Therefore,

$$\begin{aligned}Q'(\delta) &= \sum_{j \in \{L,M\}} p_j q_j'(\delta) \mathbb{E}\left[\sum_{k \in \{L,M,S\}} U_{ik} \mathbb{1}\{c_i(j) = k\} - \sum_{k \in \{L,M,S\}} U_{ik} \mathbb{1}\{c_i(S) = k\} \mid j_i^\star = j\right] \\ &= \sum_{j \in \{L,M\}} p_j q_j'(\delta) \mathbb{E}\left[(U_{ij} - U_{iS}) \mathbb{1}\{\mathcal{S}_j\} + U_{ij} \mathbb{1}\{\mathcal{E}_j\} \mid j_i^\star = j\right] \\ &= \underbrace{r_{\inf}(\delta) g_{\inf}(\delta)}_{\substack{\text{true-utility gains from} \\ \text{inframarginal switches}}} + \underbrace{E'(\delta) u_{\text{ext}}(\delta)}_{\substack{\text{true utility contributed by} \\ \text{extensive-margin entry}}}, \qquad (17)\end{aligned}$$

where we use (16) in the penultimate equality, and the definition of $r_{\inf}(\delta), g_{\inf}(\delta), u_{\text{ext}}(\delta)$ in the last equality.

Since we can write $\overline{U}(\delta) = Q(\delta)/E(\delta)$, applying the quotient rule gives

$$\overline{U}'(\delta) = \frac{Q'(\delta)E(\delta) - Q(\delta)E'(\delta)}{E(\delta)^2} = \frac{1}{E(\delta)}\left[Q'(\delta) - \overline{U}(\delta)E'(\delta)\right] = \frac{E'(\delta)}{E(\delta)}\left[\frac{Q'(\delta)}{E'(\delta)} - \overline{U}(\delta)\right].$$

Substituting (17), we obtain

$$\overline{U}'(\delta) = \frac{E'(\delta)}{E(\delta)}\Bigg[\underbrace{\frac{r_{\inf}(\delta)}{E'(\delta)}g_{\inf}(\delta)}_{\substack{\text{inframarginal true-utility gains} \\ \text{per extensive-margin entry}}} - \underbrace{\left(\overline{U}(\delta) - u_{\text{ext}}(\delta)\right)}_{\substack{\text{match-quality gap per} \\ \text{extensive-margin entry}}}\Bigg]. \tag{18}$$

Since $E(\delta) > 0$ and $E'(\delta) > 0$, the sign of $\overline{U}'(\delta)$ is exactly the sign of the difference between the two underbraced terms in (18). Therefore, average match quality is weakly increasing if and only if (13) holds. It is strictly increasing when the inequality is strict, locally unchanged when it holds with equality, and decreasing when it is reversed. □

C Bayesian Recommender System Microfoundation

In this section we describe a simple Bayesian RecSys that microfounds the two key components of Assumption 1: the popularity default and the larger marginal gains for middle-tail titles. Assumption 1 directly imposes these, but in this section we show that they can arise endogenously from a RecSys that forms beliefs about user preferences under noisy signals, where less popular titles are observed less precisely.

The Bayesian recommender system directly observes the vertical component of preferences (μ_j) and a noisy, unbiased signal about the horizontal component of preferences (θ_{ij}). For expositional ease we impose a Gaussian assumption on θ_{ij} such that $\theta_{ij} \sim \mathcal{N}(0, \sigma_\theta^2)$ and the platform observes a noisy signal $\eta_{ij} = \theta_{ij} + \epsilon_{ij}/\sqrt{\alpha_j}$ for $j \in \{L, M, S\}$, with $\epsilon_{ij} \sim \mathcal{N}(0,1)$ independent across (i,j) and across θ so that $\eta_{ij} \mid \theta_{ij} \sim \mathcal{N}(\theta_{ij}, 1/\alpha_j)$. The RecSys holds a correctly specified prior over the unobserved θ_{ij} – namely its true marginal $\mathcal{N}(0, \sigma_\theta^2)$.

The key assumption on the recommendation technology is that the signal precision is given by $\alpha_j = \delta\lambda_j$, where δ indexes technology quality as in our primary model and $\lambda_L < \lambda_M < \lambda_S$ captures the fact that more popular titles have more interaction data. The latter component captures the popularity bias that has been widely documented and studied in the RecSys literature (Celma and Cano, 2008; Abdollahpouri et al., 2017; Klimashevskaia et al., 2024).

Since μ_j is known and the prior on θ_{ij} is conjugate to the Gaussian signal, the posterior mean of utility $U_{ij} = \mu_j + \theta_{ij}$ is:

$$s_{ij} \;=\; \mu_j + w(\alpha_j)\,\eta_{ij}, \qquad w(\alpha) \;=\; \frac{\alpha\sigma_\theta^2}{1+\alpha\sigma_\theta^2}, \qquad j \in \{L, M, S\},$$

where the shrinkage weight $w(\alpha_j)$ is the posterior precision ratio from the Gaussian Bayesian update. The recommender scores each title by this posterior mean and recommends $\mathcal{R}_i = \{\arg\max_{j\in\{L,M,S\}} s_{ij}\}$.

This structure motivates each component of Assumption 1:

(i) **Popularity default.** In the limit case when the recommendation technology is very poor (i.e., δ close to 0), $w(\alpha_j) \to 0$ for all $j \in \{L, M, S\}$. In this case, the score assigned by the RecSys to each title

collapses to the vertical component ($s_{ij} \to \mu_j$) and consequently the recommender selects S for every user as $\mu_S > \mu_M > \mu_L$. Thus, in this limit case only the superstar title can be recommended to users.

As δ increases, the RecSys puts more weight on the horizontal component and decreases the degree of this popularity default but does not remove it entirely. The consequence is that a less popular title j is recommended over S only when its weighted signal overcomes the popularity advantage of S, namely when $w(\alpha_j)\,\eta_{ij} - w(\alpha_S)\,\eta_{iS} > \mu_S - \mu_j$. Since $w(\alpha_j)$ is increasing in δ, the RecSys correctly recommends j to a larger share of the users who prefer it as the technology improves. The popularity default nonetheless persists at every finite δ – a user is still recommended S whenever the signal for their preferred title is insufficiently favorable – but this bias weakens as the technology improves. This is the stylized "popularity default" of Assumption 1: the recommender defaults to S whenever the signal for a user's preferred title is insufficiently favorable, which weakens as δ increases but always persists unless the technology is perfect.

(ii) q_j **monotonicity in** δ**.** The weight $w(\alpha_j)$ is strictly increasing in δ for each $j \in \{L, M, S\}$, so the score puts increasing weight on the horizontal component θ_{ij} relative to the vertical component μ_j. As a result, the probability that the recommender identifies the user's best title, $q_j(\delta)$, is weakly increasing in δ for each j. For S, this monotonicity may be nearly flat as q_S is precise at realistic values for δ.

(iii) **Middle-tail dominance.** Assumption 1 requires the marginal gain in correctly matched users, $p_j q_j'(\delta)$ – the marginal precision gain $q_j'(\delta)$ weighted by the share of users p_j for whom j is the preferred title – to be larger for the middle-tail M than for the long-tail L in the empirically relevant range of δ.

This follows since the long-tail L has low baseline popularity μ_L – a large gap between it and the superstar title – and low signal precision λ_L – a direct consequence of L generating less interaction data. Together these cause L's effective precision $\alpha_L = \delta\lambda_L$ to advance slowly, so at empirically relevant δ its noisy signal remains far from the threshold to switch the recommendation from S to L, and $q_L'(\delta)$ stays modest. The middle-tail M, on the other hand, is closer to that threshold to begin with and, because $\lambda_M > \lambda_L$, crosses it faster as δ improves. Together these lead to

$$p_M q_M'(\delta) > p_L q_L'(\delta),$$

as long as L's marginal precision gain does not outweigh M's larger share of users for whom it is their preferred title. This need not hold for all δ – at very high δ, M saturates while L continues to advance – but should hold for the relevant portion of δ when the RecSys is not already matching M users to their preferred titles.