Personalization, Algorithmic Dependence, and Learning

Omid Rafieian* Cornell University

Si Zuo* Cornell University

Abstract

Personalized recommendation systems are now an integral part of the digital ecosystem. However, users' increased dependence on these personalized algorithms has heightened concerns among consumer protection advocates and regulators. In this work, we bring an information-theoretic perspective to this problem and examine the underpinnings of algorithmic dependence and its downstream implications for users' preference learning and independent decision-making ability, an important construct given the growing fear of adversarial AI. We develop a utility framework where users consume experience goods and sequentially update their preference weights for product attributes. We theoretically establish regret bounds for different types of users based on their dependence on the personalized algorithm. Our theoretical results demonstrate the rationality of algorithmic dependence as the gain from following the personalized algorithm grows linearly with time periods. We then develop an empirical framework to obtain model-free measures of regret for different user types. We find that personalized algorithms generate significant welfare gains, but these gains come at the expense of users' preference learning and independent decision-making. Finally, we demonstrate that simple policy interventions can help balance the trade-off between welfare and learning, offering insights for both platforms and users.

Keywords: personalized algorithms, recommendation system, preference learning, consumer protection, linear bandits, reinforcement learning

*All authors have contributed equally. We thank Andrew Ching, John Hauser, Sylvia Hristakeva, Yufeng Huang, Gui Liberali, Song Lin, John Lynch, Ted O'Donoghue, Marcel Preuss, Jiwoong Shin, Migual Villas-Boas, Jesse Yao, Hema Yoganarasimhan, and Dennis Zhang for detailed comments that have improved the paper. We also thank the participants of the 2023 Marketing Science Conference in Miami, 2024 Theory and Practice Conference in Austin, 2024 Marketing Science Conference in Sydney, and 2025 Bass FORMS Conference. All errors are our own. Please address all correspondence to: or83@cornell.edu and sz549@cornell.edu.

1 Introduction

Personalized recommendation systems are now an integral part of the digital ecosystem. Digital platforms use massive amounts of consumer-level data to deliver personalized recommendations. One of the canonical examples of recommendation systems is the Netflix movie recommendation algorithm, which reportedly saves the company over one billion dollars annually by reducing the churn rate [Gomez-Uribe and Hunt, 2015]. Other examples include Facebook and Twitter's news feed personalization, Amazon's product recommendation, and YouTube's video recommendation algorithm.

In today's digital age, the online marketplace is saturated with many options, presenting users with the challenge of sifting through too many options to find what they truly want. Personalized recommendation systems have emerged as a solution to this problem by reducing users' search costs and simplifying decision-making. These systems are designed to effectively narrow down options in real-time and guide users towards products or services that best align with their preferences and needs. By doing so, a personalized recommendation system ensures that users can select a fitting item without the need to explore the vast digital landscape exhaustively.

However, as the adoption of and reliance on personalized recommendation systems grow, there are increasing concerns regarding users' algorithmic dependence [Buçinca et al., 2021]. Prior research has shown the pitfalls of algorithmic dependence by documenting the users' tendency even to take clearly incorrect recommendations [Spatharioti et al., 2023], risks to user well-being [Banker and Khetani, 2019], and inefficiencies in users' decision-making [McLaughlin and Spiess, 2022]. Regarding the mechanism behind algorithmic dependence, prior work in economics and marketing has suggested the presence of time-inconsistent preferences [Allcott et al., 2022] and search costs [Ursu, 2018]. However, little is known about the algorithm's information advantage and how it feeds algorithmic dependence.

In this work, we bring an information-theoretic perspective to this problem and examine the underpinnings of algorithmic dependence and its downstream implications for users' preference learning. In particular, we study contexts where users consume content sequentially and may have uncertainty about their preferences given the vast space of product features, which they resolve through experience. Personalized recommendations influence the process through which users learn their preferences by affecting the decisions they make. For example, a news reader interested in social justice topics may rarely explore other content if the algorithm correctly identifies her taste and only exposes her to this type of content. Thus, dependence on algorithmic recommendations can have important implications for user learning.

Understanding algorithmic dependence and its downstream effects on users' learning is crucial from a consumer protection perspective, as users who heavily rely on algorithms without fully learning their preferences are more vulnerable to digital manipulation by adversarial AI, a topic of growing concern among policy-makers [Kusnezov et al., 2023]. In particular, when users lack a clear understanding of their own preferences, they make poorer decisions in the absence of recommendation systems, creating a feedback loop in which they become increasingly dependent on these systems while learning less about their own preferences. In this paper, we study the interplay between personalization, algorithmic dependence, and preference learning and aim to answer the following questions:

1. How does the algorithm's information advantage translate into better recommendations? Are there information-theoretic guarantees on the regret performance of personalized recommendation systems?

2. How does the dependence on a personalized algorithm affect user learning? How can we quantify the impact?

3. What consumer protection policies could generate good recommendations while helping users learn their preferences?

To answer these questions, we face several challenges. First, we need a theoretical framework that allows us to formally characterize the personalized algorithm's information advantages and disadvantages relative to individual users. In particular, we need our framework to capture learning by the algorithm and users separately from any given prior. Second, we need to theoretically compare the outcomes for users who follow personalized algorithms with those who do not. However, since algorithms and users operate under different conditions, the theoretical guarantees, such as regret bounds, often include parameters that make them inherently incomparable. Hence, we need to obtain comparable regret bounds that allow us to assess the rationality of algorithmic dependence. Third, we need an empirical framework to evaluate the performance of different algorithms without necessarily deploying those policies online. In particular, we want the evaluation process to be model-free so we do not need to rely on model-based outcome estimates.

To address our first set of challenges, we build a general linear utility framework, where users have some preference parameters over the space of product features, and products have search and experience features, consistent with Nelson [1974]. Given the high dimensionality of the feature space and the continuous introduction of new features and styles in many applied settings (e.g., movies, food), we allow users to exhibit any level of uncertainty about

their own preference parameters and quantify this uncertainty using the well-known Shannon entropy measure [Shannon, 1948].[1] To characterize users' learning, we turn to the literature on Bayesian learning and model users who update the posterior distribution of their preference parameters [Ching et al., 2013]. For users' decision-making, we use the Thompson Sampling approach as it incorporates users' Bayesian learning and is behaviorally plausible [Schulz et al., 2019, Mauersberger, 2022, Ding et al., 2022]. Under Thompson Sampling algorithms, users sample from the posterior distribution of preferences to choose an item and update the posterior distribution according to their experience. We then characterize the personalized recommendation system and its decision-making process as a low-rank model that mimics the reality of personalized algorithms used by platforms and parsimoniously accounts for the platform's information advantage over a single user who only has access to search features.

For the second challenge, we turn to the literature on bandits [Lattimore and Szepesvári, 2020]. In particular, given the central role of information advantage in our study, we focus on the bandits literature that offers information-theoretic regret bounds for the performance of different adaptive learning algorithms [Russo and Van Roy, 2016]. To isolate the impact of recommendation systems on outcomes, we focus on two types of users: (1) self-exploring users who make decisions on their own as though there is no recommendation system, and (2) recommendation-system-dependent (RS-dependent henceforth) users who follow the recommendations provided by the recommendation system. We find that the regret for self-exploring users has a lower bound that scales linearly in time periods because self-exploring users can, at best, identify the first-best product based on the search features. Conversely, the RS-dependent user has an information-theoretic upper bound for regret that grows sub-linearly in time periods. Together, our theoretical findings suggest that the welfare gain from following the RS grows linearly in time periods, highlighting a mechanism for rational dependence solely through the algorithm's information advantage.

Third, to empirically examine the impact of personalized algorithms on both welfare and learning outcomes, we use MovieLens data, which is the main public data set used as a benchmark for personalized recommendation systems. To facilitate a model-free regret evaluation, we focus on a small sample of users who have provided many ratings so we can use their observed outcomes instead of estimating them from models. We then take a hold-out subset of size 20 from the movies they have rated and exclude it from the training part. This creates a subset of products for each user in the test set, which allows us to evaluate how

[1] It is worth emphasizing that our information-theoretic framework encompasses the classical economics assumption that users have full knowledge of their preferences as a special case, which makes our approach more general and agnostic regarding the degree of uncertainty in users' preferences.

different each model's prescription is from the observed first-best in the set. Lastly, we embed a state-of-the-art matrix factorization algorithm to simulate a personalized algorithm that captures the reality of algorithms used by the platforms [Cortes, 2018].

We apply our empirical framework to the MovieLens data and examine regret and learning outcomes. We first focus on the algorithm's information advantage and welfare gains from following the algorithm. We find a significantly better regret performance by the personalized algorithm. Notably, even compared to the self-exploring user who knows her own preferences, the personalized algorithm has a persistent advantage. Further, our results show that the algorithm quickly surpasses the performance of the self-exploring user with known preferences, emphasizing its efficiency due to low-rank learning. This is an important empirical finding, as we do not impose any specific rank constraint that forces the problem to be low-rank. Together, the algorithm's better long-run performance and faster learning rationalize algorithmic dependence.

Next, we focus on the implications of algorithmic dependence for learning. We first show that the absolute amount of preference learning is higher for self-exploring users than for RS-dependent users when they start from an identical prior. Motivated by the trade-off between welfare and learning, we develop a new regret measure defined as *counterfactual regret*, which helps quantify the potential welfare consequences of insufficient learning. The counterfactual regret measures the expected regret incurred by a user in each period if they make decisions independently without the help of the personalized algorithm. As such, a user with insufficient learning would make worse decisions on their own. The key advantage of this measure is that it has the same unit as our main regret measure, which offers greater interpretability and facilitates joint optimization of welfare and learning outcomes. We compute counterfactual regret for RS-dependent users and show that while these users enjoy a low regret due to following personalized recommendations, they have a higher counterfactual regret than self-exploring users because they become worse independent decision-makers in the absence of personalized algorithms. Therefore, our examination of learning outcomes suggests that although personalized algorithms help users make better decisions that increase their welfare, these algorithms can act as a barrier to user learning since these algorithms can limit the organic exploration process users engage in.

An immediate question that arises, given the trade-off between welfare and learning, is whether there are policies that can balance the trade-off. We consider a simple class of policies whereby the recommendation system is probabilistically unavailable for a proportion of time periods. Notably, we find that there are policies in this class of policies that push

the Pareto frontier of welfare and learning, find the right balance in the trade-off between these two outcomes, and achieve good outcomes in terms of both regret and counterfactual regret. Specifically, we find that with a small amount of randomization in the availability of the algorithm, users will learn almost the same amount as self-exploring users while only sacrificing a small amount of welfare. Our findings offer important implications for platforms and users who want to self-regulate.

In summary, our paper makes several contributions to the literature. Substantively, we present a comprehensive study of the information advantage of personalized algorithms—how this advantage enhances consumer welfare while creating a cycle of algorithmic dependence that hinders consumers' preference learning and independent decision-making. Through a series of theoretical and empirical analyses, we show that the information advantage alone can rationalize algorithmic dependence, even in the absence of search costs or time-inconsistent preferences, which are often used to justify such dependence. While concerns related to privacy, fairness, and polarization have been extensively studied in the literature on personalized recommendation systems, the topic of algorithmic dependence and its downstream effects on users' learning has received less attention. Our work extends this policy debate by uncovering the underpinnings of algorithmic dependence and its negative consequences for users' independent decision-making, offering insights that are increasingly relevant given growing concerns about adversarial AI. Methodologically, we introduce a framework that allows us to establish theoretical regret bounds for users based on their dependence on personalized algorithms. A key innovation of our approach is its ability to capture the information advantage of personalized algorithms through a low-rank assumption and access to experience features unavailable to users. Additionally, we develop a counterfactual regret measure that serves as a valuable benchmark for evaluating the effects of adversarial AI. Finally, from a policy standpoint, we identify exploration-based policies that are both simple to implement and effective in achieving desirable outcomes in terms of consumer welfare and learning.

2 Related Literature

First, our paper relates to the literature on personalization. Prior methodological work in this domain has offered a variety of methods to generate personalized policies, such as low-rank matrix factorization models for collaborative filtering and a host of causal machine learning methods [Linden et al., 2003, Mazumder et al., 2010, Athey and Imbens, 2019, Koren et al., 2021, Rafieian and Yoganarasimhan, 2023]. Related applied work in this domain has focused on different aspects of personalization, such as developing personalized algorithms tailored

to specific problems [Hauser et al., 2009, Urban et al., 2014, Liberali and Ferecatu, 2022, Rafieian, 2023, Rafieian et al., 2023, Lu et al., 2025], the interplay between personalization and consumer protection policies [Goldfarb and Tucker, 2011, Johnson et al., 2020, Rafieian and Yoganarasimhan, 2021, Johnson et al., 2023, Bondi et al., 2023, Despotakis and Yu, 2023, Ning et al., 2025], and the tension between content homogenization vs. content diversity as a result of personalization [Fleder and Hosanagar, 2009, Nguyen et al., 2014, Song et al., 2019, Holtz et al., 2020, Aridor et al., 2020, Anwar et al., 2024]. Our work extends this stream of work by bringing an information-theoretic view and focusing on the foundations of algorithmic dependence and its negative impact on users' learning. In particular, we combine the insights from the literature on matrix factorization with bandits to establish theoretical regret bounds for users as a function of their dependence.

Second, our paper relates to the literature on consumer search and personalized rankings [Jeziorski and Segal, 2015, Ursu, 2018, Dzyabura and Hauser, 2019, Yoganarasimhan, 2020, Korganbekova and Zuber, 2023, Donnelly et al., 2024]. Most papers in this literature build a sequential search model akin to Weitzman [1979] to model consumers' search behavior and estimate structural parameters such as search costs. Under this modeling framework, consumers do not learn their preferences through experience but realize the match value of each item upon a costly search. The exception is Dzyabura and Hauser [2019], which allows for learning, but that paper does not allow for experiential learning. A common theme in the stream of work on consumer search is that personalized algorithms create value by reducing consumers' search costs. In that sense, search cost is the main driver behind algorithmic dependence. Our work differs from this stream of literature as we focus on a channel separate from search cost that drives algorithmic dependence: the algorithm's information advantage. In particular, we characterize the personalized algorithm's information advantage in the context of experience goods, which makes it a better predictor of whether the user likes the product or not than the users themselves.

Third, our paper relates to the literature on consumer learning. Understanding consumer learning dynamics has been of great interest to researchers in marketing [Roberts and Urban, 1988]. Ever since the seminal paper by Erdem and Keane [1996] who modeled forward-looking consumers who make decisions under uncertainty and engage in an exploration-exploitation trade-off, numerous studies have focused on choice contexts where dynamic learning plays an important role [Ackerberg, 2003, Crawford and Shum, 2005, Hitsch, 2006, Ching et al., 2013]. An important issue in this stream of work is computational complexity, which has motivated researchers to adopt heuristic-based strategies, which yield performance similar

to the dynamic programming strategy, while having the advantage of computational and cognitive simplicity [Lin et al., 2015, Tehrani and Ching, 2023]. We extend this stream of literature by offering a Thompson Sampling approach for characterizing consumer choice and learning process, which is another cognitively simple alternative to the typical dynamic programming solution to the exploration-exploitation trade-off and has been shown to be a behavioral plausible framework to model consumer behavior [Schulz et al., 2019, Ding et al., 2022]. Our work differs from this stream of literature as we consider a setting where users learn their preferences, rather than the product attributes. We further demonstrate how the increased flexibility offered by the Thompson Sampling approach can help researchers study settings with high-dimensional learning and establish information-theoretic regret bounds.

Fourth, our work relates to the vast literature on adaptive learning and multi-armed bandits. Prior research in this domain has offered a variety of algorithms to use [Lattimore and Szepesvári, 2020]. Although Thompson sampling has been around since the work by Thompson [1933], it has only recently gained traction after providing a remarkable empirical performance better than state-of-the-art benchmarks [Chapelle and Li, 2011]. Since then, many researchers have attempted to provide theoretical guarantees on Thompson sampling for a variety of adaptive learning problems [Agrawal and Goyal, 2012, 2013, Russo and Van Roy, 2014, 2016]. For a comprehensive review of Thompson sampling, please see Russo et al. [2018]. In marketing, a growing body of research explores the applications of Thompson Sampling across various domains [Schwartz et al., 2017, Jain et al., 2024, Ye et al., 2024, Waisman et al., 2025]. Most of the literature in this domain focuses on a single learner that optimizes the action and updates parameters upon experience. Our work extends this single-agent framework to a setting with both a learning recommendation system and an agent, offering new insights for modeling general principal-agent problems in contexts with decision-making under uncertainty. We offer theoretical regret bounds for different users based on their level of dependence on the algorithm.

3 Modeling Framework

We consider a general principal-agent model, where the principal is a platform that designs a Recommendation System (RS) that offers personalized product recommendations and the agent is a user of the platform who wants to consume products on the platform. The products available are experience products, meaning that only a subset of their features are available prior to consumption, while other features are only realized after consumption [Nelson, 1970, Villas-Boas, 2006]. There are numerous examples of such contexts, including movie recommendations, news personalization, and content recommendation on social media

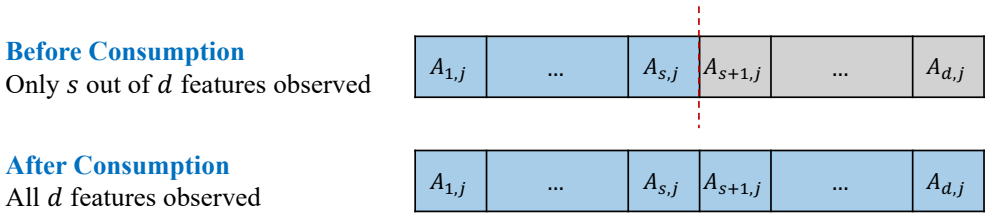


Figure 1: Illustration of search and experience features

apps. In this section, we first characterize the user's utility model in §3.1 and then describe users' preference learning in §3.2. In §3.3, we discuss the user's choice in two different regimes with and without the personalized RS.

3.1 Users' Utility Model

We propose a general utility framework in which user i derives utility from selecting product (action) A_j from the product (action) set $\mathcal{A}$. In this context, each action corresponds to consuming a distinct experience product, represented by a d-dimensional set of attributes, i.e., $\mathcal{A} \subset \mathbb{R}^d$. Following the literature on experience goods [Nelson, 1970], we categorize the product features into two types: (1) *search features*, which are attributes known before consumption (e.g., a movie's runtime, genre), and (2) *experience features*, which are realized only after consumption (e.g., the presence of a surprise ending). For convenience, we denote the first s features as search features and the remaining $d - s$ features as experience features. Figure 1 illustrates the distinction between these two feature types. We denote user i's utility from consuming product A_j by $u_i(A_j)$ and characterize it by a user-specific vector of preferences $\theta_i \in \mathbb{R}^d$ in a linear specification as follows:

$$u_i(A_j) = \theta_i^T A_j + \epsilon_{i,j}, \tag{1}$$

where $\epsilon_{i,j}$ is the error term, drawn from a normal distribution with mean zero and known variance σ_ϵ^2. Although we assume linearity for theoretical simplicity, this assumption is not overly restrictive, as a rich set of features could be used to approximate user utility in a linear manner. To account for the distinction between search and experience features, we decompose utility as follows:

$$u_i(A_j) = \underbrace{\sum_{k=1}^{s} \theta_{i,k} A_{k,j}}_{\text{utility from search features}} + \underbrace{\sum_{l=s+1}^{d} \theta_{i,l} A_{l,j}}_{\text{utility from experience features}} + \epsilon_{i,j}, \tag{2}$$

where the utility from search features is the part available to the user prior to consumption, but the utility from the experience features is only realized after consumption.

3.2 Users' Preference Learning

In high-dimensional settings with numerous product features, users are likely to experience some degree of uncertainty about their preference parameters [Branco et al., 2016, Yao et al., 2022]. This uncertainty becomes particularly pronounced when new product features are continually introduced, as users may be unfamiliar with these features and, therefore, uncertain about their importance [Lee et al., 2024]. In such cases, users gradually resolve their uncertainty by consuming products and learning their preferences over time. For instance, a movie viewer may be unsure about their preference for a specific sub-genre or a distinct visual style (e.g., the aesthetics of a Wes Anderson film) and develop a clearer understanding by watching multiple films within that category. Similarly, in the context of food, a user who has never tried a particular ingredient may initially be uncertain about their preference for it but gradually learn it through experience.

In this section, we extend our framework to sequential settings where users learn their preference parameters through sequential consumption of products. We adopt an information-theoretic approach that accommodates any level of uncertainty in users' preferences. A key advantage of this flexible characterization is that it encompasses the classical economic assumption of fully known preferences as a special case while also capturing more general scenarios where users face uncertainty about their preferences.

Let t denote each time period and $A_{i,t}$ the product chosen by user i in period t. For notational brevity, we define $U_{i,t} = u_i(A_{i,t})$ and let $\mathcal{H}_{i,t}$ denote the history (sequence) of products (actions) consumed and utility outcomes up until period t, that is, $\mathcal{H}_{i,t} = (A_{i,1}, U_{i,1}, A_{i,2}, U_{i,2}, \ldots, A_{i,t}, U_{i,t})$. We assume θ_i is drawn from a Normal distribution $N(\mu_{i,0}, \Sigma_{i,0})$. The user starts with a prior $\tilde{\theta}_{i,0} \sim N(\mu_{i,0}, \Sigma_{i,0})$ and updates the preference parameters at the end of each time period t given the prior sequence $\mathcal{H}_{i,t}$ according to the following rule:

$$\mu_{i,t} = \mathbb{E}[\theta_i \mid \mathcal{H}_{i,t}] \tag{3}$$

$$\Sigma_{i,t} = \mathbb{E}\left[(\theta_i - \mu_{i,t})(\theta_i - \mu_{i,t})^T \mid \mathcal{H}_{i,t}\right] \tag{4}$$

The sequential nature of learning indicates that users update their parameters in every time period in a Bayesian fashion. Given the Normal prior, we can apply the Bayes rule for linear Gaussian systems and present the user's parameter updating from $\mu_{i,t}$ and $\Sigma_{i,t}$ to $\mu_{i,t+1}$ and $\Sigma_{i,t+1}$ in Algorithm 1 (please see the analytical derivations of posterior mean and variance in

Web Appendix A). Algorithm 1 determines user learning given any consumption sequence $\mathcal{H}_{i,t}$. As such, for any given prior, we can examine how different consumption sequences can result in different levels of learning.

Algorithm 1 Bayesian Updating

Input: $\mu_{i,t}, \Sigma_{i,t}, A_{i,t}, U_{i,t}$

Output: $\mu_{i,t+1}, \Sigma_{i,t+1}$

1: $\Sigma_{i,t+1} \leftarrow \left(\Sigma_{i,t}^{-1} + \frac{1}{\sigma_\epsilon^2} A_{i,t} A_{i,t}^T\right)^{-1}$

2: $\mu_{i,t+1} \leftarrow \Sigma_{i,t+1}\left(\Sigma_{i,t}^{-1}\mu_{i,t} + \frac{1}{\sigma_\epsilon^2} A_{i,t} U_{i,t}\right)$

In summary, it is important to emphasize that the presence of uncertainty does not imply that users have entirely uninformative priors. For instance, a user may already have a reasonably well-calibrated belief about their enjoyment of a new visual style in movies or a certain ingredient in food. Further, it is worth noting that our learning procedure is somewhat different from the typical approach in the Bayesian learning literature in marketing and economics where consumers learn about product attributes [Erdem and Keane, 1996]. However, although preference learning may appear different from attribute learning, "the two are equivalent as long as one assumes that the utility function is linear in attributes and preference weights" (see footnote 35 in Ching et al. [2013]). Therefore, one could find an equivalence result between our analysis and one with attribute learning.

3.3 User Choice

We now discuss the user's decision-making process that determines the consumption sequence $\mathcal{H}_{i,t}$. To do so, we need to characterize the choice architecture in each period. For any product set $\mathcal{A} = \{A^{(1)}, A^{(2)}, \cdots, A^{(K)}\}$, we consider the following choice architecture when the recommendation system is available:

$$\underbrace{A^{(p)}}_{\text{recommended}}, \underbrace{A^{(1)}, A^{(2)}, \cdots, A^{(K)}}_{\text{not recommended}},$$

where one product from the set is recommended and the rest of the products are not recommended.[2] We now characterize the user's decision-making process in the definition below:

[2]Having only one recommended action is only for simplicity and one could easily extend the framework to cases with multiple recommended actions.

Definition 1. *Let $\mathcal{I}_{i,t}$ denote all the information available to user i at time t. The user's decision-making process is characterized by the policy $\pi(\cdot \mid \mathcal{I}_{i,t})$, which is a probability distribution over products conditional on the information and products available.*

To isolate the impact of personalized algorithms on users, we consider two types of users: (1) *self-exploring user*, who makes decisions on their own without the personalized RS, and (2) *RS-dependent user* who follows the personalized recommendation in every time period.[3] In what follows, we first characterize the self-exploring user's choice in §3.3.1, and then present how the RS provides personalized recommendations and characterize the consumption sequence for the RS-dependent users in §3.3.2.

3.3.1 Self-Exploring User

In the absence of the personalized recommendation, users make decisions on their own. Given the utility framework in Equation (1), a forward-looking utility-maximizing user wants to optimize the overall utility over T periods. This naturally motivates users to learn their preference parameters through experience and balance good decision-making with proper exploration of their own preference parameters. We can define the objective function for a forward-looking user as maximizing the discounted expected utility stream as follows:

$$\underset{\pi}{\operatorname{argmax}}\, \mathbb{E}\left[\sum_{t=0}^{T} \delta^t U_{i,t} \mid \mu_{i,0}, \Sigma_{i,0}, \pi\right], \tag{5}$$

where δ is the discount factor and the expectation is taken over the randomness in products $A_{i,t}$ and utilities $U_{i,t}$. Typical approaches to find the optimal sequence of choices by users involve solving a dynamic programming problem, which is known to be an NP-hard problem. The lack of cognitive simplicity of dynamic programming solutions has motivated researchers to study the simpler heuristic-based strategies as the underlying learning process [Lin et al., 2015, Tehrani and Ching, 2023].

We draw inspiration from this stream of literature and assume that users employ a Thompson Sampling approach that is a simple and intuitive heuristic-based strategy consistent with Bayesian learning [Thompson, 1933]. In addition, the prior literature has documented Thompson Sampling algorithm's behavioral plausibility as a framework to model consumer choice and learning [Schulz et al., 2019, Mauersberger, 2022, Ding et al., 2022] and its excellent

[3]We choose this stylized approach and compare these two levels of dependence and independence to offer critical and sharp insights into this problem in a tractable manner. However, as we later demonstrate in this paper, our framework can also be used to examine outcomes under more hybrid user types.

empirical performance in terms of welfare [Chapelle and Li, 2011].[4]

Thompson Sampling aims to find the right balance between exploration and exploitation in the decision-making process. The algorithm starts by initializing the user's prior belief distribution about the preference weights $N(\mu_{i,0}, \Sigma_{i,0})$. It then draws $\tilde{\theta}_{i,0}$ from this distribution and computes the utility from search features for all possible products by plugging in $\tilde{\theta}_{i,0}$ for $\theta_{i,0}$ in Equation (2). It is important to note that the user cannot use experience features because those features are not available prior to consumption. In the next step, the algorithm chooses the product that maximizes the estimated utility and observes the utility $U_{i,0}$ for that instance. Finally, the algorithm applies Bayesian updating procedure in Algorithm 1 using the new instance and updates the posterior distribution of preference weights. The Thompson Sampling algorithm continues this process for $\mathcal{T}$ periods.

Algorithm 2 Choice and Learning for the Self-Exploring User

Input: $\mu_{i,0}, \Sigma_{i,0}, \mathcal{A}, \mathcal{T}$
Output: $\mathcal{H}_{i,\mathcal{T}}, \{\mu_{i,t}, \Sigma_{i,t}\}_{t=1}^{\mathcal{T}}$

for $t = 0 \rightarrow \mathcal{T}$ **do**
  $\tilde{\theta}_{i,t} \sim N(\mu_{i,t}, \Sigma_{i,t})$ ▷ Sampling preference weights
  $A_{i,t} \leftarrow \text{argmax}_{A_j \in \mathcal{A}} \sum_{k=1}^{s} \tilde{\theta}_{i,k,t} A_{k,j}$ ▷ Selecting action based on search features
  $\Sigma_{i,t+1} \leftarrow \left(\Sigma_{i,t}^{-1} + \frac{1}{\sigma_\epsilon^2} A_{i,t} A_{i,t}^T\right)^{-1}$ ▷ Updating posterior variance
  $\mu_{i,t+1} \leftarrow \Sigma_{i,t+1}\left(\Sigma_{i,t}^{-1}\mu_{i,t} + \frac{1}{\sigma_\epsilon^2} A_{i,t} U_{i,t}\right)$ ▷ Updating posterior mean
end for

Assuming that the self-exploring user uses Thompson Sampling has several key advantages. First, it is a commonly used heuristic strategy for this dynamic problem, and early literature has shown it to be nearly optimal [Chapelle and Li, 2011]. Second, it is computationally light, making it advantageous in our later empirical analysis using the MovieLens data set. Third, due to its simplicity, it is easy to incorporate it in cases where there is a recommendation system present in the problem. We discuss this issue in the following section.

3.3.2 RS-Dependent User

We now focus on the user's choice in the presence of the recommendation system. To do so, we first introduce a personalized RS that aims to simplify the user's decision-making problem. Since we want to quantify the impact of the personalized RS on user-specific outcomes, we

[4]The prior literature has documented lower regret for Thompson Sampling compared to the alternatives across several empirical domains.

assume that the recommendation system's objective is the same as that of the user.[5] A natural difference between the personalized RS and a single user is the fact that the system has access to the data of all other users. To understand how the recommendation system's data advantage manifests itself in better decision-making capabilities, we first introduce some notations. Let $\Theta_{[d\times N]}$ denote the matrix of preference weights for a group of N users, i.e., $\Theta_{[d\times N]} = [\theta_1 \mid \theta_2 \mid \ldots \mid \theta_N]$. In all major platforms, N is a very large number. Similarly, let $A_{[d\times J]}$ denote the matrix of attributes for all J products ($J = |\mathcal{A}|$) where each column represents the vector of attributes for a product. We can define the matrix analog of the user's utility in Equation 1 as follows:

$$U = \Theta^T A + E, \tag{6}$$

where $U_{N\times J}$ represent the utility for each pair of user and product and $E_{N\times J}$ is a matrix of i.i.d error term drawn from a mean-zero Normal distribution with variance σ_ϵ^2. The recommendation system has access to $U^{\text{obs}}_{N\times J}$, which is an incomplete realization of matrix U as each user reveals utility for a subset of items. The main question is how the data from other users $U^{\text{obs}}_{N\times J}$ help the personalized algorithm learn about the new user i with vector of preferences θ_i.

In principle, if the prior data from users do not inform us about the new user, the recommendation system's strategy will be the same as the self-exploring user, presumably with a less informed prior because users know more about their own preferences than the RS. However, the prior empirical work on personalized recommendation systems suggests otherwise as it documents extensive similarities in user preferences [Koren et al., 2021]. The most common approach to characterize the similarities in user preferences is to use a factor model, which suggests that the matrix $\Theta^T A$ can be factorized into two low-rank matrices. In particular, we make the following low-rank assumption:

Assumption 1. *The expected utility matrix $\Theta^T A$ can be decomposed as follows:*

$$\Theta^T_{[d\times N]} A_{[d\times J]} = \Gamma^T_{[r\times N]} F_{[r\times J]}, \tag{7}$$

where $F_{[r\times J]} = [F_1 \mid F_2 \mid \ldots \mid F_J]$ is the matrix of product-specific factors for all J products,

[5]It is worth emphasizing that the only reason we make this assumption is to ensure that the impact of the recommendation system is not driven by the misalignment in objectives. It is generally easy to show that in cases where the objectives are misaligned, the extent of harm by the recommendation system will be larger [Kleinberg et al., 2022]. In that sense, our results will provide a lower bound for the negative impact caused by the RS.

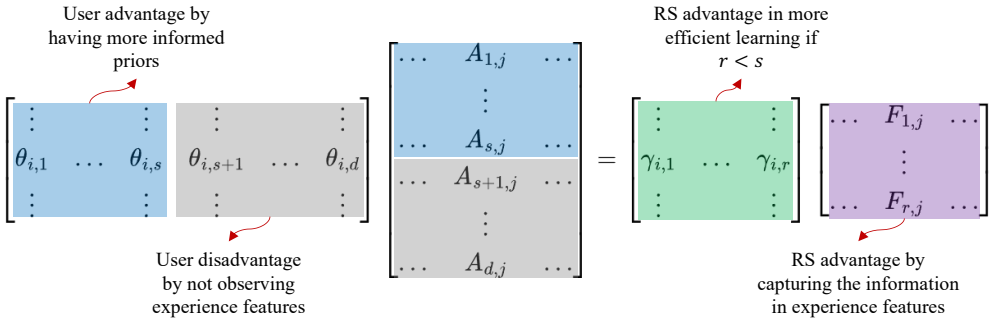


Figure 2: Comparison of information advantage by the RS and users

and $\Gamma_{[r\times N]} = [\gamma_1 \mid \gamma_2 \mid \ldots \mid \gamma_N]$ *presents the matrix of user-specific factor weights for all* N *users.*

We can now view the recommendation system's data advantage in light of Assumption 1. Because the recommendation system has other users' prior data, it has access to an accurate estimate of the product-specific matrix F, which captures not only search features, but also experience features. As such, the recommendation system's task of learning user i's preference parameters will turn into the task of learning user i's weights for r-dimensional products because we have: $\theta_i^T A_j = \gamma_i^T F_j$. Hence, the recommendation system's data advantage manifests in two ways: (1) learning r-dimensional weights as opposed to the self-exploring user's learning of d parameters, and (2) access to a low-rank embedding of product space that contains both search and experience features. Figure 2 compares the information advantage of the RS and users, emphasizing the potential for users to have more informed priors. These insights are essential for establishing theoretical guarantees for different algorithms.

Before presenting the procedure that determines choice and learning for the RS-dependent user, we make the following assumption about the user's learning:

Assumption 2. *The only channel through which users learn their preferences is consumption.*

This assumption is inspired by the fact that users do not observe experience features without consuming the product. A direct implication of this assumption is that the RS-dependent user cannot learn from the recommendations beyond their own experience. In our context, this assumption is reasonable as recommendation systems are often very complex and it is not realistic to assume that users can learn further by observing that a product is recommended.[6]

[6]Prior work has proposed clever approaches to allow for consumer inference from the recommendation

We now present the choice and learning processes for the RS-dependent user in Algorithm 3. In this setting, not only is there a user who learns her preference parameters through experience, there is also a recommendation system that learns user preferences and offers recommendations. Both players learn user i's preference parameters, but they operate in different spaces: user i learns her own parameters θ_i in the d-dimensional space, whereas the recommendation system learns user i's preference weights for factors in the r-dimensional space. To distinguish between these two learning processes, we use superscripts (θ) and (γ) to refer to the parameters of both players' prior distributions: $\mu_{i,0}^{(\theta)}$, $\Sigma_{i,0}^{(\theta)}$, $\mu_{i,0}^{(\gamma)}$, and $\Sigma_{i,0}^{(\gamma)}$.

The RS moves first in each time period. Since the recommendation system has access to F, it only wants to learn γ_i. As such, it engages in a Linear-Gaussian Thompson Sampling procedure, where it first draws $\tilde{\gamma}_{i,t}$ from the posterior distribution and then recommends the product with the highest expected utility (lines 2 and 3). The RS-dependent user always follows the recommended action (line 4). Once the utility is realized, both the user and recommendation system update parameters of their posterior distribution $\mu_{i,t+1}^{(\theta)}$, $\Sigma_{i,t+1}^{(\theta)}$, $\mu_{i,t+1}^{(\gamma)}$, and $\Sigma_{i,t+1}^{(\gamma)}$. The algorithm repeats this process for $\mathcal{T}$ periods.

Algorithm 3 Choice and Learning for the RS-Dependent User

Input: $\mu_{i,0}^{(\theta)}, \Sigma_{i,0}^{(\theta)}, \mu_{i,0}^{(\gamma)}, \Sigma_{i,0}^{(\gamma)}, F, \mathcal{A}, \mathcal{T}$

Output: $\mathcal{H}_{i,\mathcal{T}}, \{\mu_{i,t}^{(\theta)}, \Sigma_{i,t}^{(\theta)}\}_{t=1}^{\mathcal{T}}, \{\mu_{i,t}^{(\gamma)}, \Sigma_{i,t}^{(\gamma)}\}_{t=1}^{\mathcal{T}}$

1: **for** $t = 0 \to \mathcal{T}$ **do**

2: $\tilde{\gamma}_{i,t} \sim N(\mu_{i,t}^{(\gamma)}, \Sigma_{i,t}^{(\gamma)})$ ▷ RS: Sampling low-dimensional weights

3: $j^* \leftarrow \text{argmax}_j \sum_{k=1}^{r} \tilde{\gamma}_{i,k,t} F_{k,j}$ ▷ RS: Selecting recommendation

4: $A_{i,t} \leftarrow A_{j^*}$ ▷ User: Following recommendation

5: $\Sigma_{i,t+1}^{(\theta)} \leftarrow \left(\left(\Sigma_{i,t}^{(\theta)}\right)^{-1} + \frac{1}{\sigma_\epsilon^2} A_{i,t} A_{i,t}^T \right)^{-1}$ ▷ User: Updating posterior variance

6: $\mu_{i,t+1}^{(\theta)} \leftarrow \Sigma_{i,t+1}^{(\theta)} \left(\left(\Sigma_{i,t}^{(\theta)}\right)^{-1} \mu_{i,t}^{(\theta)} + \frac{1}{\sigma_\epsilon^2} A_{i,t} U_{i,t} \right)$ ▷ User: Updating posterior mean

7: $\Sigma_{i,t+1}^{(\gamma)} \leftarrow \left(\left(\Sigma_{i,t}^{(\gamma)}\right)^{-1} + \frac{1}{\sigma_\epsilon^2} F_{j^*} F_{j^*}^T \right)^{-1}$ ▷ RS: Updating posterior variance

8: $\mu_{i,t+1}^{(\gamma)} \leftarrow \Sigma_{i,t+1}^{(\gamma)} \left(\left(\Sigma_{i,t}^{(\gamma)}\right)^{-1} \mu_{i,t}^{(\gamma)} + \frac{1}{\sigma_\epsilon^2} F_{j^*} U_{i,t} \right)$ ▷ RS: Updating posterior mean

9: **end for**

Lastly, an important point about the RS-dependent user is that the preference learning

information [Shin and Yu, 2021]. As far as user's choice is concerned, the RS-dependent user in our context infers the superiority of RS by definition and always follows that. As for preference learning, one could relax Assumption 2 by allowing the user to solve a system of inequalities that indicate the recommended product is better than other products and learn about only search features of each product. However, this approach would be computationally intractable and cognitively complex for users.

process is the same for both user types. As such, any difference in the final learning outcomes comes from the consumption sequence. We utilize this fact when we examine the learning implications of RS-dependence in our empirical framework in §5.3.

4 Theoretical Analysis

In this section, we theoretically examine the algorithms described earlier. We begin by defining our primary outcomes of interest in §4.1. We then present the regret bounds for self-exploring users who follow Algorithm 2 in §4.2. Next, in §4.4, we present the regret bounds for RS-dependent users who follow Algorithm 3. Finally, in §4.4, we theoretically compare the two regret bounds, establish bounds for welfare gains from RS dependence and examine the possibility of rational dependence.

4.1 Main Outcomes

As discussed earlier, we are interested in two key user-level outcomes: welfare and learning. In this section, we define these two key outcomes and formally present the measures we use for them. We can define user welfare for user i for T periods as the total sum of utility from actions chosen under policy π as follows:

$$\texttt{Welfare}_i(T;\pi) = \mathbb{E}\left[\sum_{t=0}^{T} u_i(A_{i,t})\right], \tag{8}$$

where $\texttt{Welfare}_i$ is a user-specific function that depends on user i's preference parameters θ_i.[7] Another closely tied measure that is often studied in the literature on sequential decision-making and linear bandits is *expected regret*, which takes the difference between the expected utility from some notion of first-best in each period and user welfare. We formally define *expected regret* as follows:

Definition 2. *Suppose that $A_i^* \in \text{argmax}_{a\in\mathcal{A}}\,\mathbb{E}[u_i(a) \mid \theta_i]$ is the optimal action (product) given θ_i. For the sequence of actions $\{A_{i,t}\}_{t=0}^{T}$ chosen according to policy π, the **expected regret** is given as follows:*

$$\texttt{Regret}_i(T;\pi) = \mathbb{E}\left[\sum_{t=0}^{T} \left(u_i(A_i^*) - u_i(A_{i,t})\right)\right], \tag{9}$$

where the expectation is taken over the randomness in the actions, utilities, and the prior distribution over θ_i.

[7] It is worth noting that we set the discount factor δ to one as we consider a finite horizon problem.

This notion of expected regret is often referred to as the Bayes regret or Bayes risk. Consistent with the prior bandits literature, one advantage of using regret instead of welfare is the possibility of obtaining statistical bounds. We later use these bounds to conduct a theoretical analysis of our problem.

The second outcome we are interested in is user learning. Intuitively, the greater uncertainty the user has over her own preference parameters, the lower the degree of learning by the user. As such, we turn to the well-established concept of Shannon entropy that measures the amount of uncertainty or surprise in random variables [Shannon, 1948].

Definition 3. *Let $A^*_{i,t}$ denote the random variable corresponding to the optimal action (product) given the prior sequence $\mathcal{H}_{i,t-1}$. We measure a user's preference learning based on the **Shannon entropy** of $A^*_{i,t}$, which is defined as follows:*

$$H(A^*_{i,t}) = -\sum_{k=1}^{|\mathcal{A}|} P(A^*_i = A_k \mid \mathcal{H}_{i,t-1}) \log\left(P(A^*_i = A_k \mid \mathcal{H}_{i,t-1})\right). \tag{10}$$

According to this definition, a higher entropy means that the user is more uncertain as to what the optimal action is. For example, if the user deterministically chooses one action (maximum certainty and learning), the Shannon entropy of the optimal action will be equal to zero, i.e., $H(A_{i,t}) = 0$. On the other hand, when the user is maximally uncertain between actions, each action has an equal probability, and the Shannon entropy of optimal action will take its maximum value $H(A_{i,t}) = \log(|\mathcal{A}|)$.

A key challenge in using Shannon entropy as the primary measure of learning is that it is not directly comparable to regret, which serves as the main measure of welfare. To address this, we introduce the concept of *counterfactual regret*—the regret a user would incur at any given time period if they were to make decisions independently, without the influence of the personalized recommendation system. This measure allows us to capture the trade-off between welfare and learning: if lower regret actions come at the expense of user learning, we expect counterfactual regret to be high for that user. We formally define the *expected counterfactual regret* as follows:

Definition 4. *Let $\tilde{A}_{i,t}$ denote the counterfactual action (product), which is the random variable corresponding to the optimal action (product) given the prior sequence $\mathcal{H}_{i,t-1}$. This is the optimal action the user would choose based on her past learning through experience if the personalized algorithm was not available at period t only. For the sequence of counterfactual actions $\{\tilde{A}_{i,t}\}_{t=0}^{T}$ that would have been chosen in each period under π in the absence of RS,*

the expected ***counterfactual regret*** *is given as follows:*

$$\texttt{CounterfactualRegret} = \mathbb{E}\left[\sum_{t=0}^{T}\Big(u_i(A_i^*) - u_i(\tilde{A}_{i,t})\Big)\right], \tag{11}$$

where the expectation is taken over the randomness in the actions, utilities, and the prior distribution over θ_i*.*

The main benefit of using the notion of *counterfactual regret* is its direct comparability with the actual *regret*. In cases where users follow the personalized algorithms to choose the product, we expect the counterfactual product they would choose without the personalized algorithm to be different from the one offered by the algorithm. As such, there will be a discrepancy between the *counterfactual regret* and the actual *regret*. The gap between the two highlights the loss in independent decision-making ability due to RS-dependence.

4.2 Regret Bounds for Self-Exploring Users

We start by deriving the regret bounds for the self-exploring user. As shown in Algorithm 2, the user starts with a prior distribution $N(\mu_{i,0}, \Sigma_{i,0})$ and samples from the prior to select the product and then updates the posterior distribution based on the realized utility. The user repeats the procedure until convergence to the first-best action (product), which is the best product identifiable by the user. We denote the first-best for the self-exploring user by $A_i^{*,s}$ and define it as the product with the highest utility from *search* features if the preference parameters are known, that is, $A_i^{*,s} \in \text{argmax}_{A_j \in \mathcal{A}} \sum_{k=1}^{s} \theta_{i,k} A_{k,j}$. As such, the first-best identifiable by the self-exploring user is different from the first-best A_i^* in the regret equation (Definition 2), which allows us to write the following decomposition:

$$\begin{aligned}\texttt{Regret}_i(T;\pi) &= \mathbb{E}\left[\sum_{t=0}^{T}\left(u_i(A_i^*) - u_i(A_{i,t})\right)\right] \\ &= \mathbb{E}\left[\sum_{t=0}^{T}\left(u_i(A_i^*) - u_i(A_i^{*,s})\right)\right] + \mathbb{E}\left[\sum_{t=0}^{T}\left(u_i(A_i^{*,s}) - u_i(A_{i,t})\right)\right]\end{aligned} \tag{12}$$

where the first element in the equation above is a summation over a constant gap between the expected utility from the two first-bests, and the second term is another notion of regret for the self-exploring user. Specifically, we can show that the second term is the same as the linear bandit regret with an s-dimensional vector of preference weights, which allows us to use information-theoretic regret bounds for linear banidts established in the prior literature

[Russo and Van Roy, 2016].[8] We use this insight to arrive at the following proposition:

Proposition 1. *Let π_{SE} denote the policy for the self-exploring user who follows the Thompson Sampling algorithm for choice and learning as in Algorithm 2. Further, let g denote the gap in expected utility between the first-best action and the first-best action based on only search features, that is, $g = \mathbb{E}\left[u_i(A_i^*) - u_i(A_i^{*,s})\right]$. The regret bound for this user is as follows:*

$$gT \leq \texttt{Regret}_i(T; \pi_{SE}) \leq gT + \sqrt{2(\sigma_\epsilon^2 + \sigma_x^2) s H(A_{i,0}^*) T}, \tag{13}$$

where $H(A_{i,0}^*)$ is the Shannon entropy of the prior distribution of optimal action for self-exploring user, defined at period 0, σ_x^2 is the variance of the utility that comes from experience features, and s is the dimensionality of the search features.

Proof. Please see Web Appendix C.3.1. □

As shown in this proposition, both lower and upper bounds contain a term linear in T. The upper bound further includes an information-theoretic bound, which is closely related to the one established in Russo and Van Roy [2016]. The lower bound happens in the event where the user has no uncertainty about the preferences (i.e., $H(A_{i,0}^*) = 0$).

Next, we examine the gap g, which is defined as the difference between the first-best and the first-best based on only search features. Since this gap appears in the regret lower bound, it is crucial to understand how large of a magnitude it has and the theoretical conditions under which this term converges to zero. To do so, we need to introduce new notations: let $U_{i,j}^s$ and $U_{i,j}^x$ denote the utility the user derives from the search features and experience features, respectively. We can write the following proposition to characterize the lower bound for the gap g:

Proposition 2. *Suppose that the search utility and experience utility across products are Normally distributed, such that $U_{i,j}^s \sim N(\mu_{i,s}, \sigma_{i,s}^2)$ and $U_{i,j}^x \sim N(\mu_{i,x}, \sigma_{i,x}^2)$. If search utility and experience utility are independent across products, the expected gap between the first-best and first-best based on the search features has the following lower bound:*

$$\mathbb{E}\left[u_i(A_i^*) - u_i(A_i^{*,s})\right] \geq \sqrt{2\log(|\mathcal{A}|)} \left(\left(\sqrt{\sigma_{i,s}^2 + \sigma_{i,x}^2} - \sigma_{i,s} \right) - O\left(\frac{\sqrt{\sigma_{i,s}^2 + \sigma_{i,x}^2}}{\log(|\mathcal{A}|)} \right) \right) \tag{14}$$

[8] We present the preliminaries from information theory and important lemmas from this paper in Web Appendix B.

Proof. Please see Web Appendix C.3.2. □

As shown in Proposition 2, the gap depends on the variance of the utility from experience features and the number of products available. In particular, the gap grows as the experience features play a larger role in determining a user's utility. Further, as the action space grows, the term $O\left(\sqrt{\sigma_{i,s}^2 + \sigma_{i,x}^2}/\log(|\mathcal{A}|)\right)$ converges to zero, which suggests a substantially higher gap when the action space is larger.

4.3 Regret Bounds for RS-Dependent Users

We now turn to establishing regret bounds for the RS-dependent user. The regret analysis for the RS-dependent users is a bit more subtle. In this case, the user follows the recommendation from the personalized RS. As such, we need to find the regret bound for the RS. One could view the RS-dependent algorithm as a Thompson Sampling algorithm that operates in an r-dimensional environment, as it has access to factor information. Under Assumption 1, the first-best action that the RS could obtain is the same as the first-best action in Definition 2. Therefore, the proposition below characterizes the following regret bound for the RS-dependent user:

Proposition 3. *Let π_{RS} denote the policy for the RS-dependent user who consistently follows the personalized recommendations. Further, let $H(A_{i,0}^{RS})$ denote the Shannon entropy of the prior distribution of optimal action for the personalized RS. The regret bound for the RS-dependent user is as follows:*

$$\texttt{Regret}(T; \pi_{RS}) \leq \sqrt{2\sigma_\epsilon^2 r H(A_{i,0}^{RS}) T}, \tag{15}$$

where r is the number of factors. The expected regret for the RS-dependent user is equal to the expected regret for the personalized RS.

Proof. Please see Web Appendix C.3.3. □

As shown in Proposition 3, the regret bound for the RS-dependent user does not depend on the dimensionality of the feature space but on the rank r. In many practical settings, r is much smaller than the dimensionality of the feature space (d) or search features (s) [Udell and Townsend, 2019]. On the other hand, the main challenge for the RS is the initial lack of information about user preferences. In the extreme case, one could assume that the algorithm has an uninformative prior such that the prior distribution of optimal action gives all actions the same probability, which makes the Shannon entropy of the prior distribution of optimal

action equal to $\log(|\mathcal{A}|)$. However, modern RS algorithms try to overcome the cold-start problem by better exploring the side information about the user [Farias and Li, 2019]. Our information-theoretic bound accommodates these realistic scenarios where RS algorithms have a prior better than a random guess.

4.4 Welfare Gains from RS Dependence

As highlighted earlier, both the user and the RS have some information advantages. The regret bounds in Propositions 1 and 3 illustrate their dependence on factors associated with each type of information advantage and disadvantage, as depicted in Figure 2. On the one hand, the platform's algorithm benefits from access to data from other users, enabling it to efficiently identify the actual first-best in contrast to the self-exploring user, who can only determine the first-best based on search features. On the other hand, in certain scenarios, users may have greater knowledge of their own parameters than the algorithm, resulting in a very small Shannon entropy of the prior distribution of the optimal action ($H(A^*_{i,0}) \approx 0$), which is substantially lower than that of the RS-dependent user ($H(A^*_{i,0}) \ll H(A^{RS}_{i,0})$).

Our goal in this section is to compare the regret bounds for self-exploring and RS-dependent algorithms established earlier. Combining the lower bound for the self-exploring user's regret with the upper bound for the RS-dependent user's regret, we arrive at the following corollary:

Corollary 1. *The difference in regret between self-exploring users and RS-dependent users has a lower bound given by:*

$$\texttt{Regret}(T;\pi_{SE}) - \texttt{Regret}(T;\pi_{RS}) \geq gT - \sqrt{2\sigma_\epsilon^2 r H(A^{RS}_{i,0})T} \tag{16}$$

A few key insights emerge from Corollary 1. First, the difference in Equation (16) directly corresponds to the welfare gain from following the RS, as the first-best terms in both regret values cancel out. Second, we observe that the positive term grows linearly in T, while the negative term grows sub-linearly. This implies that with a nonzero gap g, the welfare gain from following the RS increases linearly over time. Figure 3 highlights the region where the lower bound for welfare gain in Equation (16) is positive for different numbers of time periods, gap values and a set of initial parameter, in an instance where the error variance and rank parameters are calibrated based on the prior work on movie recommendations ($\sigma_\epsilon = 0.5$ and $r = 40$), and the Shannon entropy of the prior is fully uninformative, that is, $H(A^{\text{RS}}_{i,0}) = \log(|\mathcal{A}|)$. As shown in this figure, despite the algorithm's uninformative prior and the user's full certainty about their preferences, the region where the lower bound for the

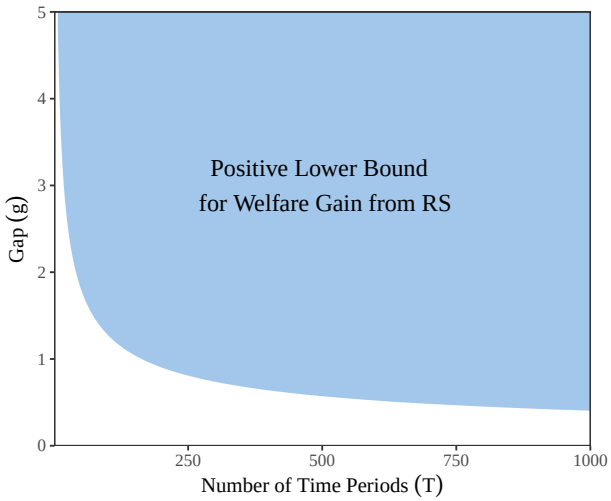


Figure 3: Region with a guaranteed welfare gain from RS dependence across model parameters

welfare gain is positive expands rapidly as the number of time periods grows, making even a small gap sufficient to make RS dependence a more utility-maximizing choice than the self-exploring alternative.

Together, our results highlight the welfare gains from the personalized algorithm created solely through its information advantage. We specifically abstract away from users' preference uncertainty and search costs to isolate the welfare gain offered by the RS purely due to its information advantage through access to other users' data. If we account for users' preference uncertainty, the region with a positive lower bound for welfare gain from the personalized RS expands, as the user must incur some exploration cost. Likewise, if the user incurs a search cost while exploring products, the illustrated region in Figure 3 will shift downward on the y-axis. Isolating the information advantage channel is important, as it suggests that even in the absence of search costs or preference uncertainty, users will rationally develop a dependence on the recommendation system (RS).

5 Empirical Framework

In this section, we first provide a broader discussion of our theoretical analysis to identify key empirical questions that warrant further investigation. As established in Corollary 1, the welfare benefits of RS tend to outweigh its downsides as the number of periods T increases. However, in real-world settings, users do not engage with the system for an infinite number of periods. As a result, self-exploring users may achieve better welfare outcomes within finite periods when (1) the gap g is small, and/or (2) the RS initially provides poor recommendations and requires time to stabilize. The latter scenario is more likely if the

effective rank r is high or if the RS starts with an uninformative prior. Therefore, assessing the net welfare impact of RS dependence in finite samples remains an empirical question, contingent on the actual values of g and r.

Second, another key outcome of interest in our study is learning. In particular, we want to know how following RS influences users' preference learning and independent decision-making ability, measured by *counterfactual regret*. In principle, if the personalized recommendations create enough variation in products to learn preference parameters, we do not expect it to affect user learning. We know that the algorithm needs to explore in the beginning, which naturally creates some variation in product features consumed. However, comparing the upper bounds in Propositions 1 and 3, we know that the RS scales the regret in the order of $\sqrt{r/s}$-fraction of the self-exploring user if they start from the same prior, that is, $H(A^*_{i,0}) = H(A^{\text{RS}}_{i,0})$. As such, the RS algorithm is likely to reach the exploitation stage quickly, potentially limiting the extent of exploration necessary for effective user learning. Specifically, if the exploitation phase primarily recommends a narrow set of products, users may experience insufficient variation to fully learn their preferences. Importantly, whether or not the exploitation stage of the algorithm generates enough variation for user learning largely depends on the distribution of user preferences and, therefore, is an empirical question.

In our empirical framework, we aim to examine the finite-sample properties of self-exploring and RS-dependent algorithms in terms of both welfare and learning measures. To accomplish these objectives, we require an empirical setting that meets the following criteria: (1) availability of preference data on user-product pairs, (2) access to a rich set of search features for products that allow us to estimate user preferences and simulate different learning patterns under various algorithms, and (3) the ability to integrate a state-of-the-art personalized recommendation system capable of learning complex user preferences and providing relevant recommendations.

To meet these three criteria, we utilize the MovieLens 1M dataset.[9] The MovieLens dataset includes over one million user ratings from 6,040 distinct users on 3,706 unique movies, which serve as our measure of user preferences. Beyond ratings, the dataset provides detailed information about the movies, including genres, themes, and an extensive collection of tags associated with each movie, which we leverage as the set of search features. Lastly, the MovieLens dataset has been widely used as a benchmark for research on personalized recommendation systems, ensuring that we can integrate state-of-the-art recommendation

[9]This dataset is publicly available and can be accessed at `https://grouplens.org/datasets/movielens/1m/`.

algorithms into our empirical framework.

In this section, we begin by outlining our empirical strategy in §5.1. We then analyze the welfare gains derived from the personalized RS due to its information advantage in §5.2. Next, in §5.3, we investigate the learning outcomes for self-exploring and RS-dependent users. Finally, in §5.4, we explore the policy implications and identify potential strategies that balance the trade-off between welfare and learning objectives.

5.1 Empirical Strategy

In order to satisfactorily answer the empirical questions, we face several challenges. Before discussing these challenges in great detail, we need to state a few assumptions upfront. The first assumption we make is about the relationship between ratings and user utility. While our theoretical framework uses the concept of user utility, we only observe user ratings. Let $Y_{[N\times J]}$ denote the rating matrix for N users and J movies. We present our assumption that links ratings to utility as follows:

Assumption 3. *User i's utility from watching movie j is well-approximated by user ratings, i.e., $U_{[N\times J]} \approx Y_{[N\times J]}$.*

Our next assumption pertains to the set of available search features. Since our data contain detailed movie tags, we use these tags as the primary search features for each movie. These tags include attributes such as originality, great ending, and good soundtrack–features that are visible to users before consumption, as they appear in the movie profile on the website (for a complete list of movie tags, see Web Appendix D.1). Naturally, one could use all tags to characterize the dimensionality of user preferences. However, it is reasonable to assume that only a subset of these tags are important for user utility. Specifically, we assume that the top k most frequently used tags capture the search features. In our analysis, we set $k = 50$ and formalize this assumption as follows:

Assumption 4. *The matrix of a movie's search features is defined as $A_{[51\times J]} = [A_1 \mid A_2 \mid \cdots \mid A_J]$, where A_j represents features of movie j in terms of the aggregated rating and top 50 tags selected from the data.*

We emphasize that the choice of k is flexible. While we use $k = 50$ for our main analysis for modeling reasons, all qualitative insights hold for other characterizations. The final assumption we make is the following about the stability of true user preferences:

Assumption 5. *For any user i, the true user vector of preferences is represented as θ_i for all time periods, independent of the movies watched.*

It is important to note that, in our setting, the user can still learn their preferences through experience. While the actual preferences remain stable, the user's belief about these preferences can evolve over time. With all the assumptions clearly defined, we discuss these challenges along with their corresponding empirical strategy in the following sections.

5.1.1 Data Sparsity

For each user in our data, we only observe ratings for a subset of products, which makes it challenging to estimate user-level parameters θ_i. To overcome this challenge, we select a test sample of 400 users who have rated over 420 movies. Having a rich set of ratings for a user allows us to estimate parameters at the individual-level, by simply regressing the outcomes on movie features. As stated in Assumption 4, we use 51 features in our analysis that contain search features that users could see on the TMDB website, including the aggregated rating for each movie as well as 50 other tag information.[10]

5.1.2 Model-free Regret Measurement

Empirically measuring regret for algorithms different from the one used in the data is fundamentally challenging. One approach is to impute ratings for unrated movies, but this requires highly accurate model-based estimates and risks information leakage between the imputation model and the algorithms being evaluated. Alternatively, restricting analysis to the set of movies each user has already rated avoids this issue but limits the ability of algorithms to identify strong recommendations outside the observed choices. Moreover, the movie recommendation setting differs slightly from the theoretical framework, where the action set $\mathcal{A}$ is fixed. In reality, once content is consumed, users may not derive the same utility from repeated consumption. This makes the set of available products for each policy different, making an apples-to-apples comparison between algorithms impossible.

To address these challenges, we randomly select a hold-out set of 20 movies for each user in our test set, ensuring that these movies have been rated by the user so that we have access to the actual outcomes and the first-best option among them. We then exclude these 20 movies from the user's training set, preventing the algorithm from accessing any information about the user's ratings for these movies. Notably, while the hold-out set remains fixed for each user, it varies across users in our test data. This setup allows us to train any algorithm without the 20 hold-out movies and evaluate its regret and counterfactual regret in a model-free manner by directly using the user's observed ratings for these hold-out movies.

[10]We only use 50 features to facilitate the OLS estimation of parameters at the user level. As a robustness check, we extend this to a richer set of features and more advanced learning algorithms. All qualitative insights remain unchanged.

5.1.3 Integrating the Personalized RS

To evaluate outcomes for RS-dependent users, we require a personalized RS that continuously updates its recommendations as it acquires more information about each user. As such, this has to be a personalized RS that can handle the cold-start problem. For each user i in the test set, we use the existing data of 5040 other users in the training set and combine it with the information about user i available at the beginning of each time period t. Following Cortes [2018], we use a matrix factorization algorithm for our personalized recommendation system, which is widely used in the industry. We then train this recommendation system and predict ratings for a hold-out set of 20 movies. Based on these predictions, the RS recommends the movie with the highest predicted rating, which the RS-dependent user subsequently consumes.

5.2 Algorithm's Information Advantage and Regret Performance

We now turn to examining the algorithm's information advantage and the welfare gains from following the RS. This empirical analysis complements the theoretical results from §4 by using real data to assess the finite-sample performance of different algorithms. Specifically, we aim to address two key questions: (1) how quickly the RS learns to make high-quality recommendations on the hold-out set of 20 movies and (2) whether these recommendations outperform those that users could identify based on observable search features. The first question pertains to the efficiency of low-dimensional learning through the factor model, while the second captures the inherent gap g—the difference between the first-best option identified by the RS and the best choice the user could make based on their search features.

For each user in our test set, we use the actual sequence of ratings from the data for the training set of movies, excluding the hold-out set. At each time period, our personalized RS updates its parameters and generates new predictions for the hold-out movies. These predictions allow us to evaluate the RS algorithm's performance in terms of regret and other relevant metrics. To contextualize the RS algorithm's effectiveness, we compare its performance against the following benchmarks:

- **Aggregate Rating:** This algorithm predicts ratings based on aggregated ratings (average rating) for each user without leveraging specific movie features. This serves as a benchmark as the personalized RS also starts from aggregated ratings and updates as more information arrives. Given the absence of a learning mechanism in this benchmark, its performance measures remain constant over time.
- **User with Known Preferences:** This algorithm predicts ratings using the user's known

Figure 4: The regret performance of different algorithms on the hold-out set

51-dimensional preference vector, which consists of a linear preference over the top 50 most frequently mentioned tags and the aggregated rating. This preference vector is estimated by regressing the user's actual ratings for all watched movies on these 51 features. The user's preference is assumed to be fully known from the outset and remains constant over time. This serves as the best performance achievable by a self-exploring user.[11] Thus, any gap between this algorithm and the personalized RS reflects the gap factor g characterized in our theoretical analysis.

- **User with Learning:** This algorithm is similar to the previous one, with the key difference being that the user learns according to Algorithm 1. Like the RS algorithm, we use the actual sequence of ratings in the data and update the preference parameters for search features over time. It starts from the prior, assigning weight one to the average rating and zero weights to the remaining features, and updates as new ratings are observed. The prior variance is set to be one for all preference parameters. At any time period, the algorithm predicts ratings for all movies in the hold-out set based on their features, allowing us to evaluate its performance.

Figure 4 displays the performance of all four algorithms over 400 periods, aggregated across 100 simulations. Several key insights emerge from this figure. First, we observe that the RS outperforms all three benchmarks in the hold-out test set, demonstrating its ability to identify and leverage key patterns in user behavior. Second, there is a consistent gap between the regret under *RS* and the *User with Known Preferences*, which highlights the intrinsic

[11] If the model specification is linear, this is the best-in-class model under Empirical Risk Minimization principle. However, we relax this assumption and consider PCA-based features as a robustness check in Web Appendix D.2.1 and arrive at qualitatively similar results.

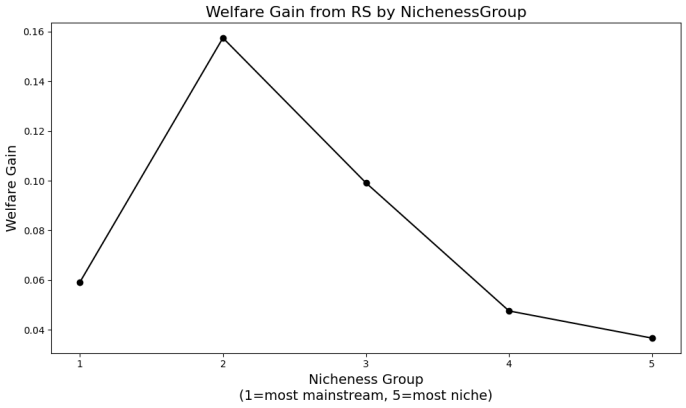


Figure 5: Welfare gains from RS across nicheness of user preferences

information advantage of the RS and the gap between the first-best and the first-best based on search features. Third, we see that the RS quickly outperforms the user with known preferences, emphasizing its efficiency due to low-rank learning. This finding is particularly important as we do not impose any specific rank assumption that forces r to be lower than 51; rather, the personalized RS algorithm identifies the rank in a data-driven manner. In Web Appendix D.2.2, we use other performance metrics to demonstrate the performance of RS compared to benchmarks.

Next, we examine the heterogeneity in welfare gains from the recommendation system (RS) relative to the *User with Learning*, based on the nicheness of user preferences. We quantify nicheness using the correlation between a user's actual movie ratings and the aggregated ratings in the training sample. Users are then categorized into five nicheness groups, with the most niche group comprising the bottom 20% in terms of correlation with the aggregate rating (details provided in Web Appendix D.2.3). Figure 5 visualizes the welfare gains from RS across these groups. Interestingly, we observe an inverted U-shaped pattern: users with moderately niche preferences experience the highest welfare gains from personalized RS. For highly mainstream users, the marginal benefit of personalization is limited since aggregate ratings already serve as strong predictors. On the other hand, for highly niche users, the RS struggles to leverage data from other users effectively, as their preferences deviate significantly from the majority. This pattern aligns with the fundamental mechanics of personalized algorithms, which rely on identifying similarities in the joint space of users and products.

In summary, our findings highlight the information advantage of the personalized RS algorithm, offering a rational explanation for why users become dependent on algorithms. This is significant because prior literature has often cited factors such as time-inconsistent

preferences and search costs to justify algorithmic dependence [Allcott et al., 2022, Ursu, 2018]. However, as our analysis suggests, there is a clear limitation for users in the context of experience goods, which rationally forces them to rely on personalized RS, even absent search costs or time-inconsistent preferences.

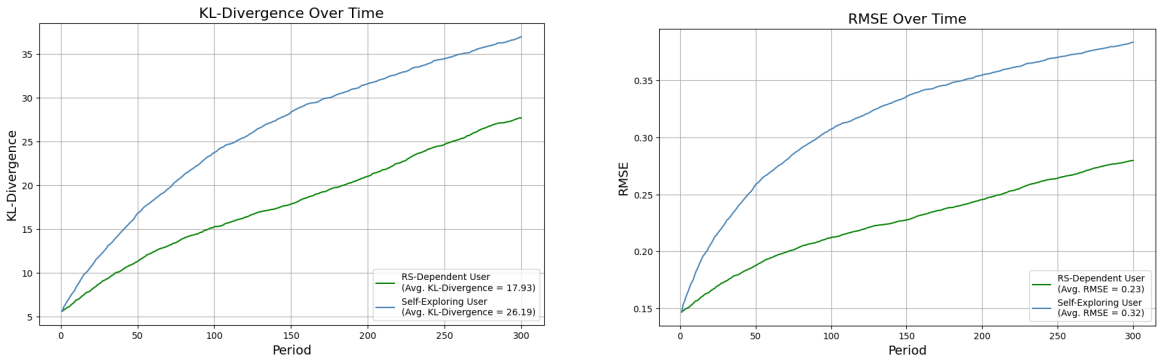
5.3 Implications for Users' Preference Learning

As shown in our theoretical and empirical analysis so far, RS-dependence brings substantial welfare gains for users, making it a rational strategy. However, this dependence also has implications for users' preference learning, as it alters the prior experience from which they resolve uncertainty about their preferences. In this section, we explore the implications of RS-dependence for users' learning and their ability to make independent decisions without the assistance of the RS.

We focus on the two user types studied in this paper: self-exploring users and RS-dependent users. Unlike in §5.2, where we maintain the user's consumption sequence, in this analysis, we modify the sequence according to Algorithms 2 and 3 to examine how each algorithm influences users' preference learning. However, we still remain within the set of movies that the user has actually rated in the data to avoid model-based evaluation. Specifically, for each user, we only allow consumption of the movies that they have already watched, excluding the hold-out set of 20 movies.

We start with a prior distribution of preferences that reflects the average taste, meaning the user's initial weight for the aggregated rating feature is set to one, while the weights for all other features are set to zero. The prior variance for all preference parameters is initialized at one. In each round, both self-exploring and RS-dependent users choose a movie from the available set, consume it, realize the utility, and update their preference parameters accordingly. We then measure the extent of learning by comparing the updated preferences with the prior preferences. If the system is learning effectively, we expect to see a larger difference between the updated preferences and the prior. To quantify this difference, we use two key measures: (1) Kullback-Leibler (KL) divergence, which is commonly used to assess the difference between two probability distributions and measures the divergence between the posterior distribution at each point from the prior distribution, and (2) Root Mean Square Error (RMSE), which calculates the square root of the average squared differences between the mean preference parameters. These measures help evaluate how much the user has learned over time.

Figure 6 illustrates the learning outcomes for both self-exploring and RS-dependent users, measured using KL divergence and RMSE. As both figures demonstrate, the difference from



(a) KL Divergence (b) RMSE

Figure 6: User learning under different algorithms

the prior increases at a faster rate for the self-exploring user compared to the RS-dependent user. This suggests that the self-exploring user learns more efficiently over time. It is important to note that in this analysis, both user types begin with the same prior preferences, so the observed differences reflect the relative learning rates of the two approaches. We further validate the robustness of these findings in Web Appendix D.2.4, where we consider scenarios in which users begin with more informed priors. Across all cases, the self-exploring user consistently exhibits greater learning. This suggests that actively exploring a diverse range of movies accelerates preference learning, whereas the RS-dependent user, despite benefiting from personalized recommendations, experiences a slower learning trajectory.

The finding in Figure 6 indeed highlights that following personalized recommendations comes at the expense of the user's ability to learn. To quantify the impact of this trade-off by a metric comparable to welfare, we turn to the measure of *counterfactual regret*, as outlined in Equation (11) and Definition 4. This measure allows us to assess how the RS algorithm influences the user's independent decision-making ability by comparing their performance with different strategies. We consider three separate user types for this analysis:

- **Self-Exploring User:** This user chooses movies on their own and updates their preferences based on their realized utility.
- **RS-Dependent User:** This user relies on the personalized RS for their movie choices and updates their preferences upon consuming the algorithm's recommendations.
- **RS-Dependent User Counterfactual:** This user follows the personalized RS recommendations in all prior periods, but in each period of the counterfactual evaluation, they make their own choices (without the algorithm's help).

If RS-dependence truly acts as a barrier to preference learning, we should observe that the

RS-Dependent User Counterfactual experiences higher regret compared to the Self-Exploring User and the RS-Dependent User. This would suggest that although RS-dependence provides welfare gains by offering better recommendations, it limits the user's ability to independently make decisions, thereby hindering their learning.

The results presented in Figure 7 underscore the trade-off between the benefits of following personalized recommendations and the cost in terms of independent decision-making ability. As shown in the figure, the two lines for the RS-Dependent User and Self-Exploring User resemble the results in Figure 4, but with the key difference that the sequence of consumed movies has changed due to the different algorithms being applied. The red line, representing counterfactual regret, illustrates the amount of regret an RS-Dependent User incurs when making decisions independently (i.e., without relying on the personalized recommendation system). This line shows a persistent gap between the expected regret (from following the RS) and the counterfactual regret (from independent decision-making).

Figure 7: Comparison of regret and counterfactual regret for different algorithms

Importantly, the expected counterfactual regret for the RS-Dependent User is consistently higher than the expected regret for the Self-Exploring User.[12] This finding suggests that although following personalized recommendations reduces expected regret, it comes at the expense of users' independent decision-making ability. This finding is important for several reasons. First, it highlights the trade-off between welfare gains and learning ability. Second, it raises concerns about dependency on personalized RS systems, particularly in contexts where independent decision-making is critical (e.g., to avoid adversarial AI attacks or to

[12]It is worth clarifying that the counterfactual regret for the self-exploring user is the same as the expected regret.

maintain a high level of user autonomy). Thus, our analysis of learning outcomes emphasizes the need to carefully balance the benefits of personalized algorithms with the potential risks of over-dependence.

5.4 Policy Implications

Our findings from §5.2 and §5.3 reveal a trade-off between welfare and learning. This raises the question: is it possible to design a policy that effectively balances these two objectives?

We explore a specific class of policies that introduce randomness into the availability of the recommendation system. With this random availability, RS-dependent users are required to make their own decisions during certain periods. This is expected to boost users' preference learning while still maintaining some of the benefits provided by the personalized RS. Specifically, we define these policies using a single parameter p, which controls the probability that the recommendation system will be unavailable. When the personalized recommendation system is available, the RS-dependent user follows the algorithm; when it is unavailable, the user makes a decision independently based on her updated preferences. Notably, when the personalized recommendation system is always available (i.e., $p = 0$), the outcomes are consistent with those for the RS-dependent user. Conversely, when the recommendation system is unavailable (i.e., $p = 1$), the outcomes align with those of the self-exploring user. From a theoretical perspective, random availability policies can be interpreted as a mixed strategy in which the user probabilistically alternates between following the RS recommendations and making independent choices. Please see Web Appendix E for the detailed description of the random availability algorithm.

As in §5.2 and §5.3, we measure both regret and counterfactual regret using the hold-out set of 20 movies for each user. For each random availability policy, we adjust the order of consumption according to the policy and assess its performance in terms of regret and counterfactual regret on the hold-out set. Figure 8 presents the results of this analysis. The figure clearly demonstrates the trade-off between regret and counterfactual regret, which is closely related to the classic exploration-exploitation dilemma. We show the Pareto Frontier for different random availability policies, where certain policies Pareto dominate the self-exploring strategy, achieving both lower expected and counterfactual regret. This finding suggests that random availability policies can provide superior alternatives to stringent data protection and privacy policies that entirely prohibit personalized recommendation systems.

Importantly, our results show that the Pareto Frontier approaches the optimal levels of expected regret achieved by the RS-dependent user. For instance, the policy with $p = 0.3$ achieves what can be considered the best of both worlds, by achieving counterfactual regret

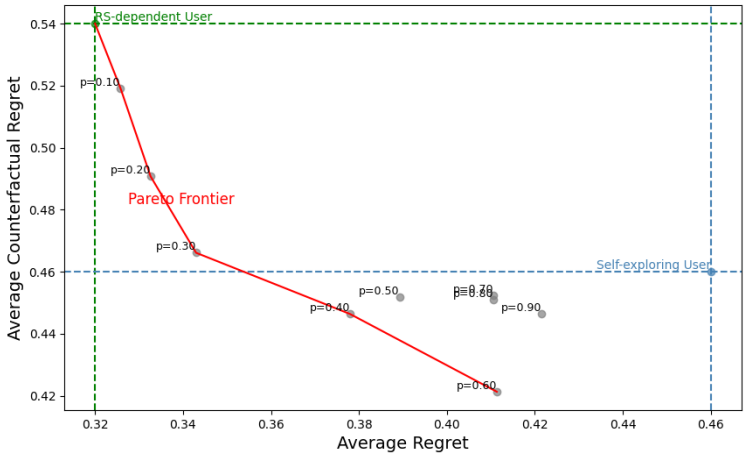


Figure 8: The regret and counterfactual regret performance of random availability policies

comparable to that of the self-exploring user while maintaining expected regret similar to that of the RS-dependent user. These findings indicate that even relatively simple policies, like random availability, can effectively balance the trade-off between welfare and learning. This offers valuable implications for platforms and users aiming to self-regulate their use of recommendation systems.

6 Discussion and Conclusion

Personalized recommendation systems have become a cornerstone of the digital ecosystem. However, growing user dependence on these algorithms has raised concerns among consumer protection advocates and regulators. In this work, we take an information-theoretic approach to examine the foundations of algorithmic dependence and its implications for users' preference learning and independent decision-making—an increasingly critical issue given rising concerns about adversarial AI. We develop a utility framework in which users consume experience goods and sequentially learn their preferences, allowing us to analyze how personalized algorithms influence the learning process. Our theoretical results establish regret bounds for different user types based on their level of dependence on personalized recommendations. We show that algorithmic dependence is rational, as the welfare gains from following the personalized algorithm relative to self-exploration grow linearly over time. To complement our theoretical findings, we introduce an empirical framework that provides model-free measures of regret across different user types. We find that personalized algorithms generate significant welfare gains, but these gains come at the cost of users' preference learning and independent

decision-making. Finally, we demonstrate that simple policy interventions can help balance the trade-off between welfare and learning, offering insights for both platforms and users.

In summary, our study makes several important contributions to the literature. From a substantive standpoint, we provide a comprehensive analysis of how personalized recommendation systems influence both welfare and learning outcomes, combining theoretical and empirical models. While these algorithms improve welfare by offering better product recommendations, they can also hinder users' learning and independent decision-making ability. Therefore, our work contributes to the policy debate around personalized algorithms by uncovering the foundations of algorithmic dependence and emphasizing its negative effects on users' ability to make independent decisions, offering insights that are particularly pertinent in light of the growing concerns surrounding adversarial AI.

Methodologically, we develop a theoretical framework to study algorithmic dependence and establish theoretical regret bounds based on users' reliance on personalized algorithms. A key innovation of our approach is its ability to capture the information advantage of these algorithms through a low-rank assumption and access to experience features unavailable to users. Additionally, we introduce a counterfactual regret measure that serves as a valuable benchmark for assessing the impact of adversarial AI. From a policy perspective, we demonstrate that straightforward exploration-based strategies can effectively balance the trade-off between welfare and learning. Our findings provide valuable insights for both managers and consumers. For managers, we propose that achieving a balance between performance and user learning is feasible at a low business cost. Similarly, our results offer self-regulation insights for consumers, enabling them to manage their own dependence on recommendation systems by taking simple steps.

Nevertheless, our paper has certain limitations that open avenues for future research. First, our findings are largely based on the established theories on user learning. One could design a long-run randomized experiment and verify these findings in the field. Second, our paper focuses on experiential preference learning. Future research can extend our work to different forms of learning (e.g., learning how to perform a task) and study the impact of algorithms on those learning outcomes. Finally, our paper studies exogenous levels of dependence on personalized algorithms to quantify the downstream consequence of this dependence. Future work can endogenize this aspect and examine the mechanisms behind this algorithmic dependence.

Competing Interests Declaration

Author(s) have no competing interests to declare.

References

D. A. Ackerberg. Advertising, learning, and consumer choice in experience good markets: an empirical examination. *International Economic Review*, 44(3):1007–1040, 2003.

S. Agrawal and N. Goyal. Analysis of thompson sampling for the multi-armed bandit problem. In *Conference on learning theory*, pages 39–1. JMLR Workshop and Conference Proceedings, 2012.

S. Agrawal and N. Goyal. Thompson sampling for contextual bandits with linear payoffs. In *International conference on machine learning*, pages 127–135. PMLR, 2013.

H. Allcott, M. Gentzkow, and L. Song. Digital addiction. *American Economic Review*, 112 (7):2424–2463, 2022.

M. S. Anwar, G. Schoenebeck, and P. S. Dhillon. Filter bubble or homogenization? disentangling the long-term effects of recommendations on user consumption patterns. *arXiv preprint arXiv:2402.15013*, 2024.

G. Aridor, D. Goncalves, and S. Sikdar. Deconstructing the filter bubble: User decision-making and recommender systems. In *Proceedings of the 14th ACM conference on recommender systems*, pages 82–91, 2020.

S. Athey and G. W. Imbens. Machine learning methods that economists should know about. *Annual Review of Economics*, 11(1):685–725, 2019.

S. Banker and S. Khetani. Algorithm overdependence: How the use of algorithmic recommendation systems can increase risks to consumer well-being. *Journal of Public Policy & Marketing*, 38(4):500–515, 2019.

T. Bondi, O. Rafieian, and Y. J. Yao. Privacy and polarization: An inference-based framework. *Available at SSRN 4641822*, 2023.

F. Branco, M. Sun, and J. M. Villas-Boas. Too much information? information provision and search costs. *Marketing Science*, 35(4):605–618, 2016.

Z. Buçinca, M. B. Malaya, and K. Z. Gajos. To trust or to think: cognitive forcing functions can reduce overreliance on ai in ai-assisted decision-making. *Proceedings of the ACM on Human-computer Interaction*, 5(CSCW1):1–21, 2021.

S. Chang, F. M. Harper, and L. Terveen. Using groups of items for preference elicitation in recommender systems. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 1258–1269, 2015.

O. Chapelle and L. Li. An empirical evaluation of thompson sampling. *Advances in neural information processing systems*, 24, 2011.

A. T. Ching, T. Erdem, and M. P. Keane. Learning models: An assessment of progress,

challenges, and new developments. *Marketing Science*, 32(6):913–938, 2013.

D. Cortes. Cold-start recommendations in collective matrix factorization. *arXiv preprint arXiv:1809.00366*, 2018.

G. S. Crawford and M. Shum. Uncertainty and learning in pharmaceutical demand. *Econometrica*, 73(4):1137–1173, 2005.

S. Despotakis and J. Yu. Multidimensional targeting and consumer response. *Management Science*, 69(8):4518–4540, 2023.

J. Ding, Y. Feng, and Y. Rong. A behavioral model for exploration vs. exploitation: Theoretical framework and experimental evidence. *arXiv preprint arXiv:2207.01028*, 2022.

R. Donnelly, A. Kanodia, and I. Morozov. Welfare effects of personalized rankings. *Marketing Science*, 43(1):92–113, 2024.

D. Dzyabura and J. R. Hauser. Recommending products when consumers learn their preference weights. *Marketing Science*, 38(3):417–441, 2019.

T. Erdem and M. P. Keane. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing science*, 15(1):1–20, 1996.

V. F. Farias and A. A. Li. Learning preferences with side information. *Management Science*, 65(7):3131–3149, 2019.

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

A. Goldfarb and C. E. Tucker. Privacy Regulation and Online Advertising. *Management science*, 57(1):57–71, 2011.

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

J. R. Hauser, G. L. Urban, G. Liberali, and M. Braun. Website morphing. *Marketing Science*, 28(2):202–223, 2009.

G. J. Hitsch. An empirical model of optimal dynamic product launch and exit under demand uncertainty. *Marketing Science*, 25(1):25–50, 2006.

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

L. Jain, Z. Li, E. Loghmani, B. Mason, and H. Yoganarasimhan. Effective adaptive exploration of prices and promotions in choice-based demand models. *Marketing Science*, 43(5):1002–1030, 2024.

K. Järvelin and J. Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems (TOIS)*, 20(4):422–446, 2002.

P. Jeziorski and I. Segal. What makes them click: Empirical analysis of consumer demand for search advertising. *American Economic Journal: Microeconomics*, 7(3):24–53, 2015.

G. A. Johnson, S. K. Shriver, and S. Du. Consumer privacy choice in online advertising: Who opts out and at what cost to industry? *Marketing Science*, 39(1):33–51, 2020.

G. A. Johnson, S. K. Shriver, and S. G. Goldberg. Privacy and market concentration: intended and unintended consequences of the gdpr. *Management Science*, 69(10):5695–5721, 2023.

J. Kleinberg, S. Mullainathan, and M. Raghavan. The challenge of understanding what users want: Inconsistent preferences and engagement optimization. *arXiv preprint arXiv:2202.11776*, 2022.

Y. Koren, S. Rendle, and R. Bell. Advances in collaborative filtering. *Recommender systems handbook*, pages 91–142, 2021.

M. Korganbekova and C. Zuber. Balancing user privacy and personalization. *Work in progress*, 2023.

D. Kusnezov, Y. A. Barsoum, E. Begoli, et al. Risks and Mitigation Strategies for Adversarial Artificial Intelligence Threats: A DHS S&T Study, 2023. URL `https://www.dhs.gov/sites/default/files/2023-12/23_1222_st_risks_mitigation_strategies.pdf`.

T. Lattimore and C. Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

J.-Y. Lee, J. Shin, and J. Yu. Communicating attribute importance under competition. *KAIST College of Business Working Paper Series*, 2024.

G. Liberali and A. Ferecatu. Morphing for consumer dynamics: Bandits meet hidden markov models. *Marketing Science*, 2022.

S. Lin, J. Zhang, and J. R. Hauser. Learning from experience, simply. *Marketing Science*, 34 (1):1–19, 2015.

G. Linden, B. Smith, and J. York. Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing*, 7(1):76–80, 2003.

S. Lu, S. Yang, and Y. Yao. Within-category satiation and cross-category spillover in multiproduct advertising. *Journal of Marketing*, 89(2):119–140, 2025.

F. Mauersberger. Thompson sampling: A behavioral model of expectation formation for economics. *Available at SSRN 4128376*, 2022.

R. Mazumder, T. Hastie, and R. Tibshirani. Spectral regularization algorithms for learning large incomplete matrices. *The Journal of Machine Learning Research*, 11:2287–2322, 2010.

B. McLaughlin and J. Spiess. Algorithmic assistance with recommendation-dependent

preferences. *arXiv preprint arXiv:2208.07626*, 2022.

K. P. Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

P. Nelson. Information and consumer behavior. *Journal of political economy*, 78(2):311–329, 1970.

P. Nelson. Advertising as information. *Journal of political economy*, 82(4):729–754, 1974.

T. T. Nguyen, P.-M. Hui, F. M. Harper, L. Terveen, and J. A. Konstan. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *Proceedings of the 23rd international conference on World wide web*, pages 677–686, 2014.

Z. E. Ning, J. Shin, and J. Yu. Targeted advertising as implicit recommendation: strategic mistargeting and personal data opt-out. *Marketing Science*, 44(2):390–410, 2025.

O. Rafieian. Optimizing user engagement through adaptive ad sequencing. *Marketing Science*, 42(5):910–933, 2023.

O. Rafieian and H. Yoganarasimhan. Targeting and privacy in mobile advertising. *Marketing Science*, 2021.

O. Rafieian and H. Yoganarasimhan. AI and personalization. *Artificial Intelligence in Marketing*, pages 77–102, 2023.

O. Rafieian, A. Kapoor, and A. Sharma. Multi-objective personalization of marketing interventions. *Available at SSRN 4394969*, 2023.

J. H. Roberts and G. L. Urban. Modeling multiattribute utility, risk, and belief dynamics for new consumer durable brand choice. *Management Science*, 34(2):167–185, 1988.

D. Russo and B. Van Roy. Learning to optimize via posterior sampling. *Mathematics of Operations Research*, 39(4):1221–1243, 2014.

D. Russo and B. Van Roy. An information-theoretic analysis of thompson sampling. *The Journal of Machine Learning Research*, 17(1):2442–2471, 2016.

D. J. Russo, B. Van Roy, A. Kazerouni, I. Osband, Z. Wen, et al. A tutorial on thompson sampling. *Foundations and Trends® in Machine Learning*, 11(1):1–96, 2018.

E. Schulz, R. Bhui, B. C. Love, B. Brier, M. T. Todd, and S. J. Gershman. Structured, uncertainty-driven exploration in real-world consumer choice. *Proceedings of the National Academy of Sciences*, 116(28):13903–13908, 2019.

E. M. Schwartz, E. T. Bradlow, and P. S. Fader. Customer acquisition via display advertising using multi-armed bandit experiments. *Marketing Science*, 36(4):500–522, 2017.

C. E. Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

J. Shin and J. Yu. Targeted advertising and consumer inference. *Marketing Science*, 40(5):

900–922, 2021.

Y. Song, N. Sahoo, and E. Ofek. When and how to diversify—a multicategory utility model for personalized content recommendation. *Management Science*, 65(8):3737–3757, 2019.

S. E. Spatharioti, D. M. Rothschild, D. G. Goldstein, and J. M. Hofman. Comparing traditional and llm-based search for consumer choice: A randomized experiment. *arXiv preprint arXiv:2307.03744*, 2023.

S. S. Tehrani and A. T. Ching. A heuristic approach to explore: The value of perfect information. *Management Science*, 2023.

W. R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3-4):285–294, 1933.

M. Udell and A. Townsend. Why are big data matrices approximately low rank? *SIAM Journal on Mathematics of Data Science*, 1(1):144–160, 2019.

G. L. Urban, G. Liberali, E. MacDonald, R. Bordley, and J. R. Hauser. Morphing banner advertising. *Marketing Science*, 33(1):27–46, 2014.

R. M. Ursu. The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science*, 37(4):530–552, 2018.

J. M. Villas-Boas. Dynamic competition with experience goods. *Journal of Economics & Management Strategy*, 15(1):37–66, 2006.

C. Waisman, H. S. Nair, and C. Carrion. Online causal inference for advertising in real-time bidding auctions. *Marketing Science*, 44(1):176–195, 2025.

M. L. Weitzman. Optimal search for the best alternative. *Econometrica: Journal of the Econometric Society*, pages 641–654, 1979.

F. Yao, C. Li, D. Nekipelov, H. Wang, and H. Xu. Learning from a learning user for optimal recommendations. In *International Conference on Machine Learning*, pages 25382–25406. PMLR, 2022.

Z. Ye, H. Yoganarasimhan, and Y. Zheng. Lola: Llm-assisted online learning algorithm for content experiments. *arXiv preprint arXiv:2406.02611*, 2024.

H. Yoganarasimhan. Search personalization using machine learning. *Management Science*, 66 (3):1045—-1070, 2020.

Appendices

A Analytical Derivation of Bayes Updating Rule with Gaussian Noise

Proof. First, we start from a fundamental result in Bayesian statistics:

Lemma 1. *(Bayes rule for linear Gaussian systems - Theorem 4.4.1 in Murphy [2012])*
Suppose we have two variables, X and Y. Let $X \in \mathbb{R}^{D_x}$ be a hidden variable and $Y \in \mathbb{R}^{D_y}$ be a noisy observation of X. Assume the following prior and likelihood:

$$\mathbb{P}(X) = \mathcal{N}(\mu_X, \mathbf{\Sigma}_X), \quad \mathbb{P}(Y \mid X) = \mathcal{N}(AX + B, \mathbf{\Sigma}_Y).$$

The posterior $\mathbb{P}(X \mid Y)$ is then given by:

$$\mathbb{P}(X \mid Y) = \mathcal{N}(\mu_{X|Y}, \mathbf{\Sigma}_{X|Y}),$$

where the posterior mean and covariance are:

$$\mathbf{\Sigma}_{X|Y}^{-1} = \mathbf{\Sigma}_X^{-1} + A^\top \mathbf{\Sigma}_Y^{-1} A,$$

$$\mu_{X|Y} = \mathbf{\Sigma}_{X|Y} \left(\mathbf{\Sigma}_X^{-1} \mu_X + A^\top \mathbf{\Sigma}_Y^{-1} (Y - B) \right).$$

We now apply Lemma 1 to our specific setting. At any given time t, we have the following prior belief about θ_i:

$$\theta_{i,t} \sim \mathcal{N}(\mu_{i,t}, \Sigma_{i,t}).$$

New observations at time t are given by the action $A_{i,t} \in \mathbb{R}^{d \times 1}$ and utility $U_{i,t}$. According to Equation (1),

$$u_i(A_j) = \theta_i^\top A_j + \epsilon_{i,j},$$

where $\epsilon_{i,j} \sim \mathcal{N}(0, \sigma_\epsilon^2)$. Therefore, the likelihood of observing $U_{i,t}$ given θ_i is:

$$U_{i,t} \mid A_{i,t}, \theta_i \sim \mathcal{N}(A_{i,t}^\top \theta_i, \sigma_\epsilon^2).$$

Thus, we clearly define the prior and likelihood as:

$$\mathbb{P}(\theta_i) = \mathcal{N}(\mu_{i,t}, \Sigma_{i,t}), \quad \mathbb{P}(U_{i,t} \mid \theta_i, A_{i,t}) = \mathcal{N}(A_{i,t}^\top \theta_i, \sigma_\epsilon^2).$$

Using Lemma 1, the posterior variance is updated as follows:

$$\Sigma_{i,t+1}^{-1} = \Sigma_{i,t}^{-1} + A_{i,t}^\top (\sigma_\epsilon^{-2}) A_{i,t}.$$

Since σ_ϵ^{-2} is a scalar, this simplifies to:

$$\Sigma_{i,t+1}^{-1} = \Sigma_{i,t}^{-1} + \frac{1}{\sigma_\epsilon^2} A_{i,t} A_{i,t}^\top.$$

Next, applying Lemma 1, we also update the posterior mean:

$$\mu_{i,t+1} = \Sigma_{i,t+1} \left(\Sigma_{i,t}^{-1} \mu_{i,t} + A_{i,t}^\top (\sigma_\epsilon^{-2}) U_{i,t} \right).$$

Simplifying further (again, since σ_ϵ^{-2} is scalar), we obtain:

$$\mu_{i,t+1} = \Sigma_{i,t+1} \left(\Sigma_{i,t}^{-1} \mu_{i,t} + \frac{1}{\sigma_\epsilon^2} A_{i,t} U_{i,t} \right).$$

Therefore, we've explicitly derived the posterior distribution:

$$\theta_i \mid U_{i,t}, A_{i,t} \sim \mathcal{N}(\mu_{i,t+1}, \Sigma_{i,t+1}),$$

with posterior mean and variance:

$$\Sigma_{i,t+1}^{-1} = \Sigma_{i,t}^{-1} + \frac{1}{\sigma_\epsilon^2} A_{i,t} A_{i,t}^\top, \quad \mu_{i,t+1} = \Sigma_{i,t+1} \left(\Sigma_{i,t}^{-1} \mu_{i,t} + \frac{1}{\sigma_\epsilon^2} A_{i,t} U_{i,t} \right).$$

□

B Preliminaries from Information Theory

Since we work with information-theoretic bounds, we need to provide some basic definitions for the concepts we use in the proofs. We start with Shannon entropy that appears in our regret analysis and is one of the most fundamental concept in information theory:

Definition 5. *For variable X, the Shannon entropy is defined as:*

$$H(X) = -\sum_{x \in \mathcal{X}} P(X = x) \log \left(P(X = x) \right) \tag{17}$$

Naturally, Shannon entropy quantifies the uncertainty or information content in a probability distribution. It measures how unpredictable or surprising an outcome is when drawn from a given set of probabilities. Another important definition is the Kullback-Leibler divergence, which is defined for any two probability measures as follows:

Definition 6. *The Kullback-Leibler (KL) divergence between two probability measures P and Q, where P is absolutely continuous with respect to Q, is defined as:*

$$D_{KL}(P \| Q) = \int \log \left(\frac{dP}{dQ} \right) dP. \tag{18}$$

Here, $\frac{dP}{dQ}$ is the Radon–Nikodym derivative of P with respect to Q, which represents the density of P relative to Q.

KL divergence measures how much one probability distribution differs from another. Intuitively, it quantifies the amount of extra information required to describe data sampled from one distribution (P) when using a different distribution (Q) as a reference. is not symmetric, meaning switching P and Q gives a different result. This reflects the fact that using Q to approximate P may lead to more or less information loss than vice versa. As expected, $D_{KL}(P\|Q)$ is zero if and only if $P = Q$ everywhere, meaning there is no difference between the two distributions.

Lastly, we define the concept of mutual information between two variables as follows:

Definition 7. *The mutual information between two random variables X and Y is defined as:*

$$I(X;Y) = D_{KL}(P(X,Y)\|P(X)P(Y)) \tag{19}$$

Alternatively, it can be expressed in terms of entropy:

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \tag{20}$$

Intuitively, mutual information (MI) measures the amount of information that one random variable provides about another. It quantifies how much knowing X reduces uncertainty about Y (or vice versa). Unlike KL divergence, the mutual information is symmetric. The independence properties nicely transfer to the concept of mutual properties. For example, if Z is independent of both X and Y, we have:

$$I(X;Y \mid Z) = I(X;Y) \tag{21}$$

Further, the probability chain rule appears in additive terms as follows:

$$I(X;(Y,Z)) = I(X;Y) + I(X;Z \mid Y) \tag{22}$$

The chain rule offers great convenience in settings with a large number of variables. Lastly, we present a KL divergence form of mutual information as follows:

$$I(X;Y) = \sum_{x\in\mathcal{X}} P(X=x) D_{KL}(P(Y \mid X=x)\|P(Y)) \tag{23}$$

For the proof of this equation, please refer to Russo and Van Roy [2016].

C Proofs

In this section, we present the proofs for the propositions in the main text. We start with some useful lemmas in Appendix C.1 that later allow us to prove the propositions in the main

text. We then present a proof for the general case of a linear reward function in Appendix C.2. We present the proofs for propositions in the main text in Appendix C.3.

C.1 Useful Lemmas for a Generic Linear Reward Case

We present a general case of a reward function R, which is defined as a linear function of action characteristics through parameters β as follows:

$$R(A_j) = \beta^T A_j + \nu_j, \tag{24}$$

where A_j is an action with d-dimensional features such that $A_j \in \mathcal{A} \subset \mathbb{R}^d$ and $\beta \in \mathbb{R}^d$ is the weights and ν_j is the idiosyncratic term with known variance σ_ν. One may notice the similarity between this reward characterization and our utility specification. The reason we define this more generally is to use it as a reference in the analysis of regret in both cases. As such, we refer to reward when discussing the general case, and utility in our main models that are specific to the context of experience products. We further drop index i for brevity. One could easily add individual-specific indices for reward parameters.

Let A^* denote the first-best action, that is, $A^* = \text{argmax}_{A_j \in \mathcal{A}} \mathbb{E}[R(A_j) \mid \beta]$, which allows us to define expected regret the same way as in Definition 2. Further, let A^{TS} denote the action chosen by the Thompson Sampling algorithm given the prior information set $\mathcal{H}$. A key property of the Thompson Sampling algorithm is its probability matching, which is reflected in the following equation:

$$P(A^{\text{TS}} = A_j \mid \mathcal{H}) = P(A^* = A_j \mid \mathcal{H}) \tag{25}$$

Intuitively, it means that the probability that the first-best is equal to each action is the same as the probability the algorithm chooses that action, given the information available. The feature above is why the Thompson Sampling is often referred to as the *probability matching* algorithm, and makes the application of this algorithm appealing in real settings [Chapelle and Li, 2011]. We use this feature in establishing the regret bounds as both the self-exploring and RS-dependent users in our context use different version of the Thompson Sampling algorithm.

We define an important matrix $M \in \mathbb{R}^{|\mathcal{A}| \times |\mathcal{A}|}$, where each element in this matrix is defined as follows:

$$M_{j,k} = \sqrt{P(A^* = A_j)}\sqrt{P(A^* = A_k)}\left(\mathbb{E}\left[R(A_j) \mid A^* = A_k\right] - \mathbb{E}\left[R(A_j)\right]\right), \tag{26}$$

where the two terms under the square root are the probabilities of the optimal action being

equal to actions corresponding to the row and column of the matrix, and the last term is the expected reward from action A_j given the information that action A_k is first-best minus the unconditional expected reward for action A_j. Though it may not be clear why we define such a matrix, its properties allow us to obtain bounds that later help us in establishing the information-theoretic regret bounds. Before we present a few important lemmas, we stress that the matrix M is not to be confused with our utility matrix as it is only defined for product pairs. Lastly, it is worth emphasizing that the matrix M can be defined for any prior history $\mathcal{H}$. We drop that from the notation for simplicity.

We start by a simple fact for any matrix and write the following lemma:

Lemma 2. *For any matrix $Q \in \mathbb{R}^{k\times k}$, the following property holds:*

$$tr(Q) \leq \|Q\|_F \sqrt{Rank(Q)}, \tag{27}$$

where $tr(Q)$ is the trace of the matrix ($tr(Q) = \sum_{i=1}^{k} Q_{i,i}$), and $\|Q\|_F$ is the Frobenius norm of the matrix, that is, $\|Q\|_F = \sqrt{\sum_{i=1}^{k}\sum_{j=1}^{k} Q_{i,j}^2}$.

Proof. Let $\tilde{r}$ denote the rank of matrix $Q_{[k\times k]}$ where $\tilde{r}$ singular values are denoted by $\tilde{\sigma}_1, \tilde{\sigma}_2, \cdots, \tilde{\sigma}_{\tilde{r}}$. Further, let $\|Q\|_*$ denote the nuclear norm of matrix Q such that $\|Q\|_* = \sum_{i=1}^{\tilde{r}} \tilde{\sigma}_i$. On the one hand, we can write:

$$\text{tr}(Q) = \text{tr}\left(\frac{1}{2}Q + \frac{1}{2}Q^T\right) \overset{(1)}{\leq} \|\frac{1}{2}Q + \frac{1}{2}Q^T\|_* \overset{(2)}{\leq} \frac{1}{2}\|Q\|_* + \frac{1}{2}\|Q^T\|_* \overset{(3)}{=} \|Q\|_*, \tag{28}$$

where inequality (1) is a result of Von Neumann's trace inequality, (2) applies the triangle inequality, and (3) uses the fact that $\|Q\|_* = \|Q^T\|_*$. On the other hand, we can write the following inequality for the nuclear norm:

$$\|Q\|_* = \sum_{i=1}^{\tilde{r}} \tilde{\sigma}_i \overset{(1)}{\leq} \sqrt{\tilde{r}}\sqrt{\sum_{i=1}^{\tilde{r}} \tilde{\sigma}_i^2} = \sqrt{\tilde{r}}\|Q\|_F, \tag{29}$$

where (1) is an application of Cauchy-Schwarz inequality. Combining Equations (28) and (29), we arrive at the following inequality and complete the proof:

$$\text{tr}(Q) \leq \sqrt{\tilde{r}}\|Q\|_F \tag{30}$$

□

We can apply Lemma 2 to matrix M and arrive at the following inequality:

$$\text{tr}(M) \leq \sqrt{\text{Rank}(M)}\|M\|_F \tag{31}$$

We now attemp to further simplify all three pieces in the equation above: (1) trace of matrix M, (2) rank of matrix M, and (3) Frobenius norm of matrix M. We do so in the following sections:

C.1.1 Trace of Matrix M

We start with the trace of the matrix. We can write the following lemma:

Lemma 3. *For the matrix M defined in Equation* (26)*, the following equality holds:*

$$tr(M) = \mathbb{E}\left[R\left(A^*\right)\right] - \mathbb{E}\left[R\left(A^{TS}\right)\right] \tag{32}$$

Proof. The proof of this lemma uses the probability matching feature of the Thompson Sampling algorithm that yields $P(A^{\text{TS}} = A_j) = P(A^* = A_j)$. We can write:

$$\begin{aligned}
\text{tr}(M) &= \sum_{A_j \in \mathcal{A}} P(A^* = A_j) \left(\mathbb{E}\left[R(A_j) \mid A^* = A_k\right] - \mathbb{E}\left[R(A_j)\right]\right) \\
&= \sum_{A_j \in \mathcal{A}} P(A^* = A_j) \mathbb{E}\left[R(A_j) \mid A^* = A_k\right] - \sum_{A_j \in \mathcal{A}} P(A^* = A_j) \mathbb{E}\left[R(A_j)\right] \\
&= \sum_{A_j \in \mathcal{A}} P(A^* = A_j) \mathbb{E}\left[R(A_j) \mid A^* = A_k\right] - \sum_{A_j \in \mathcal{A}} P(A^{\text{TS}} = A_j) \mathbb{E}\left[R(A_j) \mid A^{\text{TS}} = A_k\right] \\
&= \mathbb{E}\left[R\left(A^*\right)\right] - \mathbb{E}\left[R\left(A^{\text{TS}}\right)\right]
\end{aligned} \tag{33}$$

□

C.1.2 Rank of Matrix M

We now focus on the second element in Equation (31): rank of matrix M. The following lemma characterizes the upper bound for the rank of this matrix as follows:

Lemma 4. *For the matrix M defined in Equation* (26)*, the rank is bounded as follows:*

$$Rank(M) \leq d \tag{34}$$

Proof. We can rewrite $M_{j,k}$ as follows:

$$\begin{aligned}
M_{j,k} &= \sqrt{P(A^* = A_j)}\sqrt{P(A^* = A_k}\left(\mathbb{E}\left[R(A_j) \mid A^* = A_k\right] - \mathbb{E}\left[R(A_j)\right]\right) \\
&= \sqrt{P(A^* = A_j)}\sqrt{P(A^* = A_k}\left(\mathbb{E}\left[\beta^T \mid A^* = A_k\right] A_j - \mathbb{E}\left[\beta^T\right] A_j\right) \\
&= \sqrt{P(A^* = A_j)}\sqrt{P(A^* = A_k}\left(\mathbb{E}\left[\beta^T \mid A^* = A_k\right] - \mathbb{E}\left[\beta^T\right]\right) A_j
\end{aligned} \tag{35}$$

Since both $\mathbb{E}\left[\beta^T \mid A^* = A_k\right]$ and A_j are d-dimensional, we show that the rank of matrix M

is upper bounded by d. Therefore, $\text{Rank}(M) \leq d$. □

C.1.3 Frobenius Norm of Matrix M

We now focus on the third piece of the inequality in Equation (31): the Frobenius norm of matrix M. This is one of the pieces where the information-theoretic aspect of our problem appear. We first show a simple lemma about the mutual information:

Lemma 5. *For the three variables A^*, A^{TS}, and $R(A^{TS})$, the following relationship holds:*

$$I\left(A^*; \left(A^{TS}, R\left(A^{TS}\right)\right)\right) = \sum_{A_j, A_k \in \mathcal{A}} p_j^* p_k^* D_{KL}\left(P\left(R(A_j) \mid A^* = A_k\right) \| P(R(A_j)\right) \tag{36}$$

where $p_j^* = P(A^* = A_j)$ and $p_k^* = P(A^* = A_k)$.

Proof. We start by rewriting $I\left(A^*; \left(A^{\text{TS}}, R\left(A^{\text{TS}}\right)\right)\right)$ using a few simple facts about mutual information:

$$\begin{aligned}
I\left(A^*; \left(A^{\text{TS}}, R\left(A^{\text{TS}}\right)\right)\right) &= I(A^*; A^{\text{TS}}) + I(A^*; R(A^{\text{TS}}) \mid A^{\text{TS}}) \\
&= I(A^*; R(A^{\text{TS}}) \mid A^{\text{TS}}) \\
&= \sum_{A_j \in \mathcal{A}} P(A^{\text{TS}} = A_j) I(A^*; R(A^{\text{TS}}) \mid A^{\text{TS}} = A_j) \\
&= \sum_{A_j \in \mathcal{A}} P(A^{\text{TS}} = A_j) I(A^*; R(A_j)) \\
&= \sum_{A_j \in \mathcal{A}} p_j^* \left(\sum_{A_k \in \mathcal{A}} p_k^* D_{KL}\left(P\left(R(A_j) \mid A^* = A_k\right) \| P(R(A_j))\right) \right) \\
&= \sum_{A_j, A_k \in \mathcal{A}} p_j^* p_k^* D_{KL}\left(P\left(R(A_j) \mid A^* = A_k\right) \| P(R(A_j))\right),
\end{aligned} \tag{37}$$

where the first line is the application of chain rule, the second line uses the fact that A^* is independent of A^{TS}, the third line just expands, the fourth line uses the independence of A^{TS} from both A^* and $R(A^{\text{TS}})$, and the fifth line applies Equation (23). □

The right-hand side of Equation (37) presents the mutual information in terms of KL divergence, which can be further simplified using Pinsker's inequality through the following lemma:

Lemma 6. *Suppose that the error term in the reward function in Equation* (24) *comes from $N(0, \sigma_\nu^2)$. The following inequality holds:*

$$D_{KL}\left(P\left(R(A_j) \mid A^* = A_k\right) \| P(R(A_j)\right) \geq \frac{\left[\left(\mathbb{E}\left[R(A_j) \mid A^* = A_k\right] - \mathbb{E}\left[R(A_j)\right]\right)\right]^2}{2\sigma_\nu^2} \tag{38}$$

Proof. Please see the proof for Fact 9 and Lemma 3 in Russo and Van Roy [2016]. □

We can now combine Lemma 6 with Lemma 5 to write the following lemma about the Frenobius norm of matrix M:

Lemma 7. *The following inequality holds for the Frobenius norm of matrix M in Equation (26):*

$$\|M\|_F \leq 2\sigma_\nu^2 I\left(A^*; \left(A^{TS}, R\left(A^{TS}\right)\right)\right) \tag{39}$$

Proof. We can expand the Frobenius norm of matrix M as follows:

$$\begin{aligned}
\|M\|_F &= \sum_{A_j, A_k \in \mathcal{A}} P(A^* = A_j) P(A^* = A_k) \left(\mathbb{E}\left[R(A_j) \mid A^* = A_k\right] - \mathbb{E}\left[R(A_j)\right]\right)^2 \\
&\leq \sum_{A_j, A_k \in \mathcal{A}} P(A^* = A_j) P(A^* = A_k) (2\sigma_\nu^2) D_{KL}\left(P\left(R(A_j) \mid A^* = A_k\right) \| P(R(A_j))\right) \\
&= 2\sigma_\nu^2 I\left(A^*; \left(A^{\mathrm{TS}}, R\left(A^{\mathrm{TS}}\right)\right)\right),
\end{aligned} \tag{40}$$

where the inequality in the second line comes from Lemma 6. □

C.1.4 Revisiting the Inequality for Matrix M

We now revisit the inequality for matrix M in Equation (31), using all the lemmas above that help us simplify a relationship the mutual information and the per-period regret. In particular, Lemma 3 connects the per-period regret term to the trace of matrix M, and Lemma 7 connects the mutual information to the Frobenius norm of matrix M. We can write the following lemma for the relationship between the mutual information and the regret:

Lemma 8. *The ratio of the squared per-period regret to the mutual information between A^* and A^{TS} has the following upper bound:*

$$\frac{\left[\mathbb{E}\left[R\left(A^*\right)\right] - \mathbb{E}\left[R\left(A^{TS}\right)\right]\right]^2}{I\left(A^*; \left(A^{TS}, R\left(A^{TS}\right)\right)\right)} \leq 2\sigma_\nu^2 d \tag{41}$$

Proof. We can write:

$$\begin{aligned}
\frac{\left[\mathbb{E}\left[R\left(A^*\right)\right] - \mathbb{E}\left[R\left(A^{\mathrm{TS}}\right)\right]\right]^2}{I\left(A^*; \left(A^{\mathrm{TS}}, R\left(A^{\mathrm{TS}}\right)\right)\right)} &= \frac{\mathrm{tr}(M)}{I\left(A^*; \left(A^{\mathrm{TS}}, R\left(A^{\mathrm{TS}}\right)\right)\right)} \\
&\leq \frac{\mathrm{Rank}(M)\|M\|_F}{I\left(A^*; \left(A^{\mathrm{TS}}, R\left(A^{\mathrm{TS}}\right)\right)\right)} \\
&\leq \frac{d 2\sigma_\nu^2 I\left(A^*; \left(A^{\mathrm{TS}}, R\left(A^{\mathrm{TS}}\right)\right)\right)}{I\left(A^*; \left(A^{\mathrm{TS}}, R\left(A^{\mathrm{TS}}\right)\right)\right)} \\
&= 2\sigma_\nu^2 d,
\end{aligned} \tag{42}$$

where the first line applies Lemma 3, the second line applies the matrix inequality in Equation (31), and the third line applies Lemma 7. □

C.2 Regret Bounds for the General Case of Linear Reward

We now establish the regret bound for the general case with linear rewards and d-dimensional action space as follows:

Lemma 9. *Suppose that the reward function is linear in action features as in Equation* (24)*, where the error term comes from* $N(0, \sigma_\nu^2)$ *and the algorithm has access to the information about all features when making the decision. Further, assume that* $H(A_0^*)$ *is the Shannon entropy of the optimal action given the prior. The following regret bound holds for the Thompson Sampling algorithm with policy* π^{TS} *for any* T*:*

$$\texttt{Regret}(T, \pi^{TS}) = \sqrt{2\sigma_\nu^2 d H(A_0^*) T} \tag{43}$$

Proof. The proof uses the bound established in Lemma 8 and other inequalities as follows:

$$\begin{aligned}
\texttt{Regret}(T, \pi^{\mathrm{TS}}) &= \mathbb{E}\left[\sum_{t=0}^{T} \left(R\left(A^*\right) - R\left(A_t^{\mathrm{TS}}\right)\right)\right] \\
&= \mathbb{E}\left[\sum_{t=0}^{T} \mathbb{E}\left[\left(R\left(A^*\right) - R\left(A_t^{\mathrm{TS}}\right)\right) \mid \mathcal{H}_t\right]\right] \\
&= \mathbb{E}\left[\sum_{t=0}^{T} \frac{\mathbb{E}\left[\left(R\left(A^*\right) - R\left(A_t^{\mathrm{TS}}\right)\right) \mid \mathcal{H}_t\right]}{\sqrt{I\left(A^*; \left(A_t^{\mathrm{TS}}, R\left(A_t^{\mathrm{TS}}\right)\right) \mid \mathcal{H}_t\right)}} \sqrt{I\left(A^*; \left(A_t^{\mathrm{TS}}, R\left(A_t^{\mathrm{TS}}\right)\right) \mid \mathcal{H}_t\right)}\right] \\
&\leq \sqrt{2\sigma_\nu^2 d}\, \mathbb{E}\left[\sum_{t=0}^{T} \sqrt{I\left(A^*; \left(A_t^{\mathrm{TS}}, R\left(A_t^{\mathrm{TS}}\right)\right) \mid \mathcal{H}_t\right)}\right] \\
&\leq \sqrt{2\sigma_\nu^2 d} \sqrt{T\mathbb{E}\left[\sum_{t=0}^{T} I\left(A^*; \left(A_t^{\mathrm{TS}}, R\left(A_t^{\mathrm{TS}}\right)\right) \mid \mathcal{H}_t\right)\right]} \\
&= \sqrt{2\sigma_\nu^2 d} \sqrt{T\left[\sum_{t=0}^{T} I\left(A^*; \left(A_t^{\mathrm{TS}}, R\left(A_t^{\mathrm{TS}}\right)\right) \mid \{A_\tau^{\mathrm{TS}}, R\left(A_\tau^{\mathrm{TS}}\right)\}_{\tau=0}^{t-1}\right)\right]} \\
&= \sqrt{2\sigma_\nu^2 d}\, \sqrt{T I\left(A^*; \{A_\tau^{\mathrm{TS}}, R\left(A_\tau^{\mathrm{TS}}\right)\}_{\tau=0}^{t}\right)} \\
&= \sqrt{2\sigma_\nu^2 d}\, \sqrt{T\left[H\left(A^*\right) - H\left(A^* \mid \{A_\tau^{\mathrm{TS}}, R\left(A_\tau^{\mathrm{TS}}\right)\}_{\tau=0}^{t}\right)\right]} \\
&\leq \sqrt{2\sigma_\nu^2 d}\, \sqrt{T H\left(A^*\right)},
\end{aligned} \tag{44}$$

where the first line simply writes the equation for regret, the second line uses the law of iterated expectation to account for per-period information $\mathcal{H}_t$ that is available, the third line simply includes a mutual information term in the numerator and denominator that is

shown to be positive in Lemma 7, the fourth line uses Lemma 8, the fifth line applies the Cauchy-Schwarz inequality to the summation, the sixth line expands the mutual information term, the seventh line applies the chain rule for mutual information, the eighth line rewrites the mutual information in terms of Shannon entropy, and the ninth line uses the fact that the Shannon entropy is always non-negative. □

The lemma above, along with the ones earlier, all heavily borrow from Russo and Van Roy [2016]. We present them here in the appendix of our paper, so the reader can refer to the proofs here. In light of Lemma 9, we can prove the propositions in our main text.

C.3 Proof for Propositions

C.3.1 Proof for Proposition 1

Proof. The proof for Proposition 1 follows the decomposition presented in the main text of the paper. We can first expand that decomposition as follows:

$$\begin{aligned}
\texttt{Regret}_i(T;\pi) &= \mathbb{E}\left[\sum_{t=0}^{T}\left(u_i(A_i^*) - u_i(A_{i,t})\right)\right] \\
&= \mathbb{E}\left[\sum_{t=0}^{T}\left(u_i(A_i^*) - u_i(A_i^{*,s})\right)\right] + \mathbb{E}\left[\sum_{t=0}^{T}\left(u_i(A_i^{*,s}) - u_i(A_{i,t})\right)\right] \\
&= gT + \mathbb{E}\left[\sum_{t=0}^{T}\left(u_i(A_i^{*,s}) - u_i(A_{i,t})\right)\right],
\end{aligned} \tag{45}$$

where $A_i^{*,s}$ is the first-best product given the search features. We apply Lemma 9 to the second term. We use the fact that the utility function by the self-exploring user can is a reward function with s-dimensional parameters, where the variation of the utility from experience features goes into the error term. We denote the variance of this error term by σ^2_{SE}. As such, applying Lemma 9 gives us the upper bound of $\sqrt{2\sigma^2_{\text{SE}} s H(A_0^*)T}$. We further note that the variance σ^2_{SE} is upper bounded as follows:

$$\sigma^2_{\text{SE}} \leq \sigma^2_\epsilon + \sigma^2_x, \tag{46}$$

where the equality holds if there is no mutual information between the experience and search features. This completes the proof. □

C.3.2 Proof for Proposition 2

Proof. This proof uses a simple lemma that finds upper and lower bounds for the expected maximum of k draws from a Normally distributed variable. We state the lemma as follows:

Lemma 10. *Let $X_1, X_2, \ldots, X_k \sim \mathcal{N}(\mu, \sigma^2)$ be i.i.d. normal random variables with mean μ*

and standard deviation σ. Define the maximum:

$$M_k = \max(X_1, X_2, \ldots, X_k).$$

Then, the expectation $\mathbb{E}[M_k]$ satisfies the following bounds for any $k \geq 2$:

$$\mu + \sigma\left(\sqrt{2\log k} - \frac{\log\log k + \log(4\pi)}{\sqrt{8\log k}}\right) \leq \mathbb{E}[M_k] \leq \mu + \sigma\left(\sqrt{2\log k}\right) \tag{47}$$

Proof. Let $X_1, X_2, \ldots, X_k \sim \mathcal{N}(\mu, \sigma^2)$ be i.i.d. normal random variables. Define the standardized variables:

$$Z_i = \frac{X_i - \mu}{\sigma}, \quad \text{so that } Z_i \sim \mathcal{N}(0,1).$$

Let $M_k = \max(X_1, X_2, \ldots, X_k)$ and define the corresponding standardized maximum:

$$Z_{(k)} = \max(Z_1, Z_2, \ldots, Z_k).$$

To find the upper bound, we apply a series of inequalities and definitions for some $t > 0$ as follows:

$$\begin{aligned}
\exp\left(t\mathbb{E}[Z_{(k)}]\right) &\leq \mathbb{E}\left[\exp\left(tZ_{(k)}\right)\right] \\
&= \mathbb{E}\left[\max_i \exp\left(tZ_i\right)\right] \\
&\leq \sum_{i=1}^{k} \mathbb{E}\left[\exp\left(tZ_i\right)\right] \\
&= k\exp\left(\frac{t^2}{2}\right),
\end{aligned} \tag{48}$$

where the first line applies the Jensen's inequality, the second line uses the definition of $Z_{(k)}$, the third line applies the union bound, and the fourth line applies the moment generating function. Taking logs from both sides of Equation (48), we have $t\mathbb{E}[Z_{(k)}] \leq \log(k) + t^2/2$, which we can simplify as follows:

$$\mathbb{E}[Z_{(k)}] \leq \frac{\log(k)}{t} + \frac{t}{2} \tag{49}$$

To find a tight upper bound, we choose t that minimizes the RHS of Equation (49), which is $t = \sqrt{2\log(k)}$. This gives us the following upper bound:

$$\mathbb{E}[Z_{(k)}] \leq \sqrt{2\log(k)}. \tag{50}$$

To obtain the lower bound, we seek to approximate $\mathbb{E}[Z_{(k)}]$, which can be related to extreme

value approximations. The cumulative distribution function (CDF) of $Z_{(k)}$ is given by:

$$F_{Z_{(k)}}(z) = P(Z_{(k)} \leq z) = P(Z_1 \leq z, \ldots, Z_k \leq z) = (\Phi(z))^k .$$

The probability density function (PDF) is then:

$$f_{Z_{(k)}}(z) = k\,(\Phi(z))^{k-1}\,\phi(z),$$

where $\Phi(z)$ and $\phi(z)$ denote the standard normal CDF and PDF, respectively. A well-known result in extreme value theory provides an approximation for $\mathbb{E}[Z_{(k)}]$:

$$\mathbb{E}[Z_{(k)}] \geq \sqrt{2\log(k)} - \frac{\log(\log(k)) + \log(4\pi)}{\sqrt{8\log(k)}}. \tag{51}$$

Combining Equations (50) and (51), we get the following result:

$$\sqrt{2\log(k)} - \frac{\log(\log(k)) + \log(4\pi)}{\sqrt{8\log(k)}} \leq \mathbb{E}[Z_{(k)}] \leq \sqrt{2\log k}, \quad \text{for } k \geq 2.$$

By scaling back to the original normal variables:

$$\mathbb{E}[M_k] = \mu + \sigma\mathbb{E}[Z_{(k)}],$$

we obtain the desired bounds:

$$\mu + \sigma\left(\sqrt{2\log(k)} - \frac{\log(\log(k)) + \log(4\pi)}{\sqrt{8\log(k)}}\right) \leq \mathbb{E}[M_k] \leq \mu + \sigma\left(\sqrt{2\log k}\right).$$

□

We now use this lemma for the expected maximum when it comes from $N(\mu_{i,s} + \mu_{i,x}, \sigma_{i,s}^2 + \sigma_{i,x}^2)$ (actual first-best) and when it comes from $N(\mu_{i,s}, \sigma_{i,s}^2)$ (search first-best). We first apply the lower bound in Lemma 10 to get the following result:

$$\begin{aligned}\mathbb{E}\left[u_i(A_i^*)\right] &\geq (\mu_{i,x} + \mu_{i,s}) + \sqrt{\sigma_{i,x}^2 + \sigma_{i,s}^2}\left(\sqrt{2\log(|\mathcal{A}|)} - \frac{\log(\log(|\mathcal{A}|)) + \log(4\pi)}{\sqrt{8\log(|\mathcal{A}|)}}\right) \\ &= (\mu_{i,x} + \mu_{i,s}) + \sqrt{2\log(|\mathcal{A}|)}\sqrt{\sigma_{i,x}^2 + \sigma_{i,s}^2}\left(1 - \frac{\log(\log(|\mathcal{A}|)) + \log(4\pi)}{4\log(|\mathcal{A}|)}\right)\end{aligned} \tag{52}$$

We now apply the upper bound in Lemma 10 to the utility from the search first-best, which

results in the following result:

$$\begin{aligned}\mathbb{E}\left[u_i(A_i^*)\right] &= \mathbb{E}\left[\sum_{k=1}^{s}\theta_{i,k}A_{k,i}^*\right] + \mathbb{E}\left[\sum_{l=s+1}^{d}\theta_{i,l}A_{l,i}^*\right] \\ &= \mathbb{E}[\max_j U_{i,j}^s] + \mu_{i,x} \\ &\leq \mu_{i,s} + \sigma_{i,s}\left(\sqrt{2\log(|\mathcal{A}|)}\right) + \mu_{i,x},\end{aligned} \tag{53}$$

where the first equality only splits the search and experience features, the second line uses the independence between search and experience utility across products, and the third line applies Lemma 10.

Combining Equations (52) and (53), we get the following result and complete the proof:

$$\mathbb{E}\left[u_i(A_i^*) - u_i(A_i^{*,s})\right] \geq \sqrt{2\log(|\mathcal{A}|)}\left(\left(\sqrt{\sigma_{i,s}^2+\sigma_{i,x}^2} - \sigma_{i,s}\right) - O\left(\frac{\sqrt{\sigma_{i,s}^2+\sigma_{i,x}^2}}{\log(|\mathcal{A}|)}\right)\right) \tag{54}$$

□

C.3.3 Proof for Proposition 3

Proof. The proof for Proposition 3 directly applies Lemma 9 to the case of an r-dimensional action space. In this case, the variance of the error term for the reward function is σ_ϵ^2. Therefore, the upper bound for regret can be written as follows:

$$\texttt{Regret}(T;\pi_{\text{RS}}) \leq \sqrt{2\sigma_\epsilon^2 r H(A_{i,0}^{\text{RS}})T},$$

□

D Details of the Empirical Framework

D.1 Details on the Movie Attributes

Table A1 presents the top tags in the MovieLens dataset.

Table A1: Top 100 Movie Lens Data Tags

Top 1-34	Top 35-68	Top 69-100
original	pg-13	unlikely friendships
mentor	cinematography	passionate
great ending	redemption	very interesting
catastrophe	light	dramatic
dialogue	intense	relationships
good	family	so bad it's funny
great	corruption	independent film
chase	not funny	murder
runaway	unusual plot structure	sexy
good soundtrack	twists & turns	drinking
storytelling	entirely dialogue	childhood
vengeance	suprisingly clever	complex
story	pornography	creativity
weird	transformation	lone hero
drama	cult film	atmospheric
greed	adapted from:book	based on book
great acting	happy ending	first contact
imdb top 250	very funny	entertaining
culture clash	death	narrated
brutality	life & death	friendship
fun movie	social commentary	obsession
adaptation	stylized	based on a book
criterion	interesting	loneliness
life philosophy	enigmatic	sexualized violence
suspense	fight scenes	oscar (best supporting actress)
melancholic	harsh	very good
predictable	police investigation	gunfight
visually appealing	revenge	stereotypes
talky	justice	underrated
great movie	quirky	secrets
oscar (best directing)	excellent script	nudity (full frontal - brief)
clever	feel-good	
destiny	gangsters	
fantasy world	violence	

D.2 Robustness Checks and Extensions

D.2.1 User's Preferences on Principal Component Attributes

In the paper, we define the user's learning based on the top 50 most frequently mentioned tags and aggregated ratings. Here, we conduct a robustness check where users learn based on the top 20 Principal Component Attributes and aggregated ratings. The first 20 PCAs are calculated using all movie tags and capture 80% of total variance in features. Figure A1

replicates Figure 4 but replaces tag-based learning with PCA attribute-based learning. We observe a very similar pattern in which the RS model outperforms the User with Learning model. Additionally, PCA-based learning performs slightly better than tag-based learning, as indicated by the lower final-period regret compared to the User with Learning model in Figure 4. This suggests that principal component attributes may contain more movie feature information than tags alone.

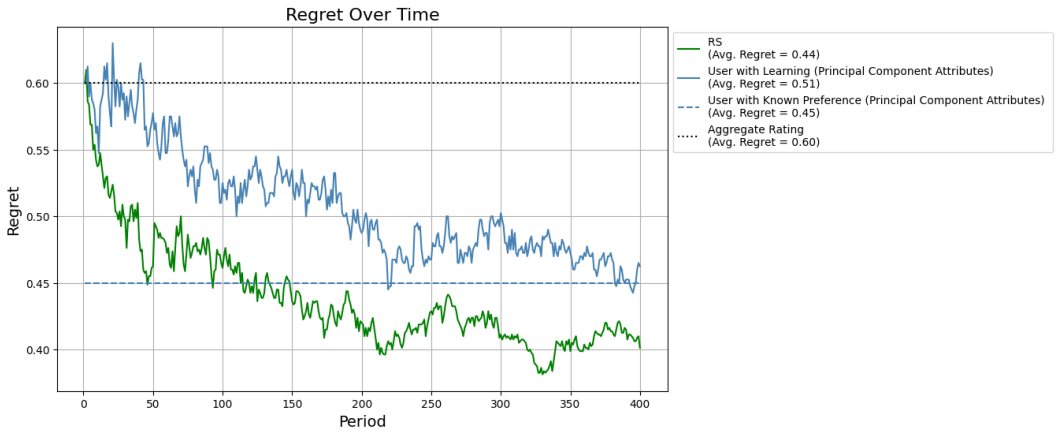
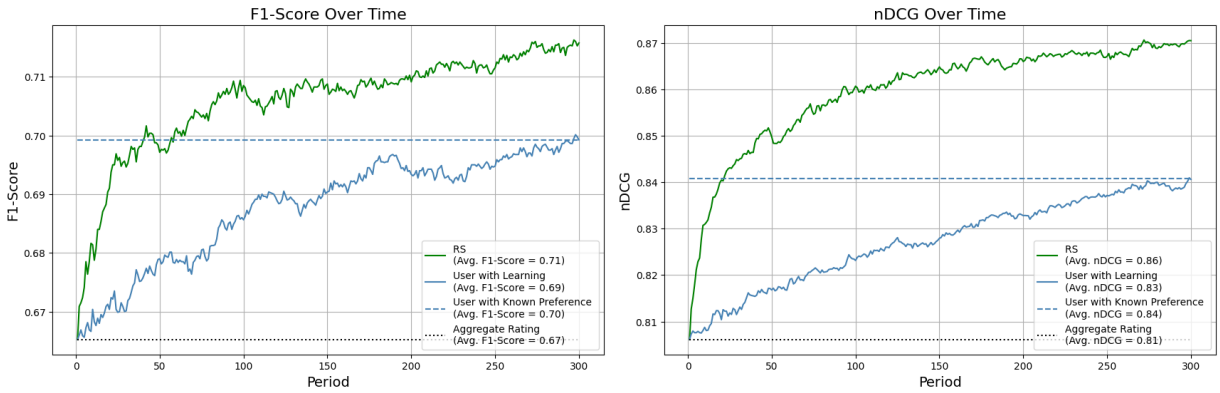


Figure A1: Regret performance of different algorithms when users rely to PCA features

D.2.2 Alternative Performance Metrics

In §5.2, we use regret as our main performance metric. In this section, we focus on additional performance accuracy measures that evaluate how well each algorithm predicts the ranking of items. Specifically, we use two commonly employed accuracy metrics: (1) *F1-Score*, which compares the top 10 recommendations from the model with the actual top 10 recommendations in the hold-out set [Chang et al., 2015], and (2) *Normalized Discounted Cumulative Gain (nDCG)*, an evaluation metric that assesses the quality of a ranked list by comparing the actual ranking with the ideal (perfect) ranking [Järvelin and Kekäläinen, 2002]. Both of these metrics are widely adopted in the literature on recommendation systems and collaborative filtering [Koren et al., 2021]. Figure A2 illustrates the performance of different algorithms in terms of F1-Score and nDCG. As shown in both figures, the personalized RS quickly outperforms all other benchmarks, highlighting its efficiency and inherent information advantage.



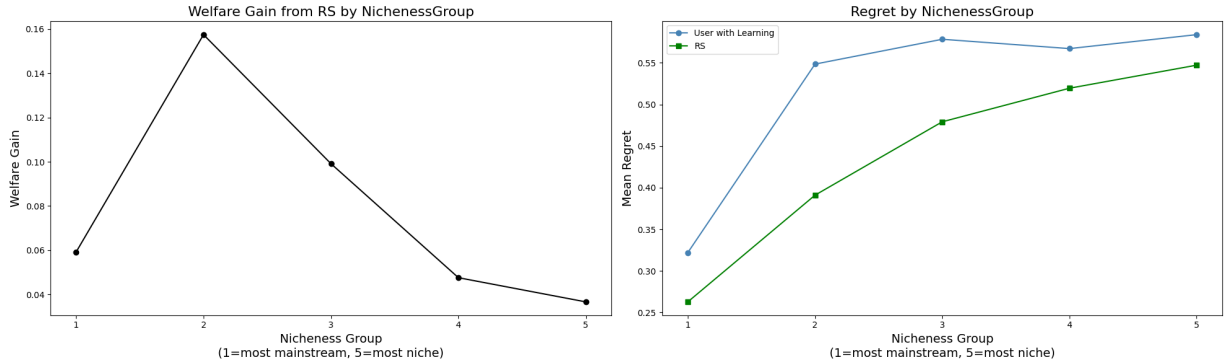
(a) F1-Score (b) nDCG

Figure A2: Performance of different algorithms using F1-Score and nDCG metrics

D.2.3 Heterogeneity in Welfare Gains from RS Based on Nicheness of User Preferences

We want to explore how the nicheness of user preference could affect RS performance. We define the nicheness of a user using the correlation between the aggregated rating of the movie and the user's actual rating of the movie in the training sample. That is, if the user has a niche preference, then the correlation should be low, as the aggregated rating would not predict the user's actual rating well. The lower the correlation, the more niche the user is. We categorize users into 5 groups based on their nicheness numbers using the 20% quantiles, where 1 represents the most mainstream users and 5 represents the most niche users.

Figure A3 presents the results corresponding to Figure 4 for user groups with different levels of nicheness. As observed in the graphs, RS's welfare gain is inverse-U shaped. When users are highly mainstream, the aggregated rating alone already provides accurate predictions of their choices. Conversely, for highly niche users, both RS and the users themselves struggle to accurately learn preferences. As a result, RS's relative prediction power is maximized when users are neither too mainstream nor too niche.

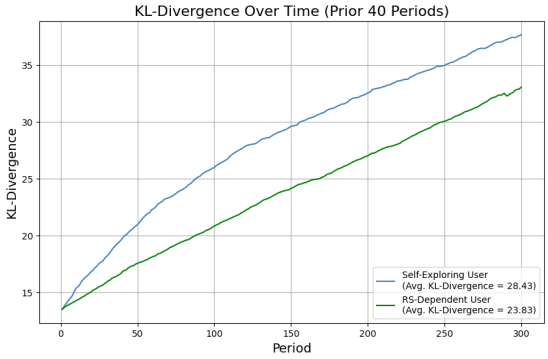


(a) Welfare Gain (b) Regret

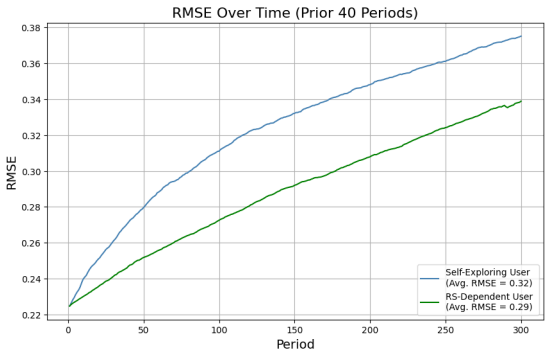
Figure A3: Welfare gain and regret from RS and *User with Learning* across nicheness of user preferences

D.2.4 User's Prior Information Compared to RS

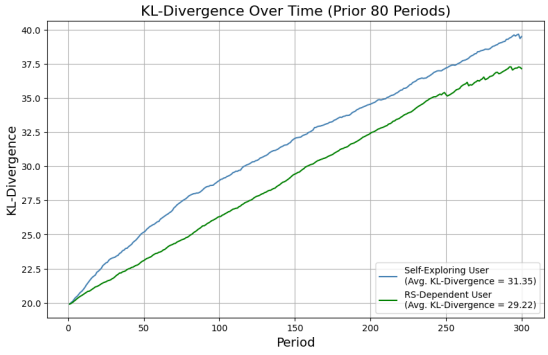
In §5.3, we assume both user types start from a prior that assigns weight one to the aggregate learning and zero weights to other features. Although aggregate rating is a relatively informed prior, users can have more informed priors. We manipulate the level of informativeness of prior by taking the posterior distribution of users in the analysis of §5.3 after watching a certain number of movies. In particular, we use the posterior distribution of self-exploring user at period 40, 80, and 120 as the prior for both types of users and measure learning under for self-exploring and RS-dependent users. Figure A4 shows the KL divergence and RMSE when users have more informed priors. These measures are calculated using the difference between the posterior at each point and the prior used in the analysis of §5.3: prior mean one for the aggregate rating, prior mean zero for other parameters, and prior variance one for all parameters. Across all prior settings, we replicate our main finding that the self-exploring user learns more than the RS-dependent user. As expected, the more informed the prior gets, the total amount of learning is lower for both groups, since they start from an informed prior that does not leave much room for additional learning.



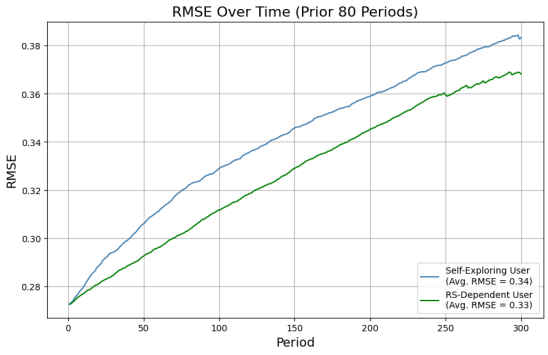
(a) KL Divergence (Prior 40 Periods)



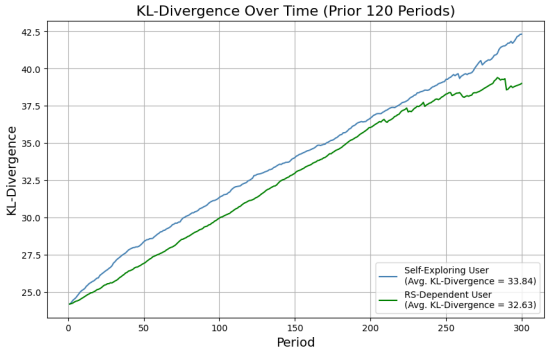
(b) RMSE (Prior 40 Periods)



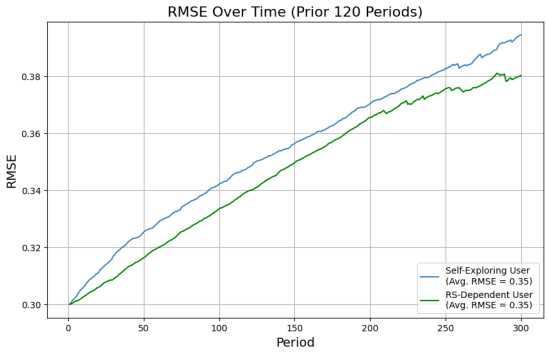
(c) KL Divergence (Prior 80 Periods)



(d) RMSE (Prior 80 Periods)



(e) KL Divergence (Prior 120 Periods)



(f) RMSE (Prior 120 Periods)

Figure A4: User Learning under Different Algorithms (with User's Prior Information)

E Algorithm for the Random Availability Policies

In Section 5.4, we propose the policies with random availability. We show the detailed pseudo-code in Algorithm 4. Most parts of this algorithm are similar to Algorithm 3. The main difference is that the algorithmic recommendation is only available with a probability p.

Algorithm 4 Choice and Learning for the RS-Dependent User with Stochastic RS Availability

Input: $\mu_{i,0}^{(\theta)}, \Sigma_{i,0}^{(\theta)}, \mu_{i,0}^{(\gamma)}, \Sigma_{i,0}^{(\gamma)}, F, \mathcal{A}, \mathcal{T}, p$

Output: $\mathcal{H}_{i,\mathcal{T}}, \{\mu_{i,t}^{(\theta)}, \Sigma_{i,t}^{(\theta)}\}_{t=1}^{\mathcal{T}}, \{\mu_{i,t}^{(\gamma)}, \Sigma_{i,t}^{(\gamma)}\}_{t=1}^{\mathcal{T}}$

for $t = 0 \rightarrow \mathcal{T}$ **do**
  $\xi_{i,t} \sim \text{Bernoulli}(p)$ ▷ Determine RS Availability
  if $\xi_{i,t} = 1$ **then**
    $\tilde{\gamma}_{i,t} \sim N(\mu_{i,t}^{(\gamma)}, \Sigma_{i,t}^{(\gamma)})$ ▷ RS: Distribution Sampling
    $A_{i,t} \leftarrow \text{argmax}_{A_j \in \mathcal{A}} \sum_{k=1}^{r} \tilde{\gamma}_{i,k,t} F_{k,j}$ ▷ RS: Recommendation Selection
  else
    $\tilde{\theta}_{i,t} \sim N(\mu_{i,t}, \Sigma_{i,t})$ ▷ User: Distribution Sampling
    $A_{i,t} \leftarrow \text{argmax}_{A_j \in \mathcal{A}} \sum_{k=1}^{s} \tilde{\theta}_{i,k,t} A_{k,j}$ ▷ User: Action Selection
  end if
  Refer to Algorithm 3 for belief updating steps.
end for