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 learn their preferences to examine the effect of personalized algorithms on the learning process. 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 scales 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 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.

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

[*] We thank John Hauser, Yufeng Huang, Gui Liberali, Song Lin, John Lynch, Ted O'Donoghue, Marcel Preuss, and Hema Yoganarasimhan 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. 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 with the intent to make users' choices easier by reducing their search costs. 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 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 to even 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 [Korganbekova and Zuber, 2023]. 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 digital 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-theoretical 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 over 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]. 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]. 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 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, even absent the search costs and time-inconsistent preferences.

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 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, 2018b].

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, even absent search costs or time-inconsistent preferences.

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 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 consumer learning since they 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 and 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 the self-exploring user 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 provides several contributions to the literature. Substantively, we present a comprehensive study of the effect of personalized recommendation systems on important welfare and learning outcomes through a series of theoretical and empirical analyses. We document that even when personalized algorithms enhance welfare by offering better product recommendations to users, they can negatively affect users' learning by limiting the degree to which they explore their own preferences. While concerns related to privacy, fairness, and polarization are more thoroughly studied in the past literature on personalized recommendation systems, the topic of algorithmic dependence and its downstream impacts on users' learning has been less studied. Thus, our work extends this policy debate by deriving the underpinnings of algorithmic dependence and highlighting its negative consequences for users' independent decision-making ability, 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 the personalized algorithm. 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 find that there are exploration-based policies that are simple enough to be implemented and achieve desirable outcomes in terms of both 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 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], 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], 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 and 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] who 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 a channel separate from search cost that drives algorithmic dependence: 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 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, Erdem et al., 2005, Hitsch, 2006, Erdem et al., 2008, Ching et al., 2013]. An important issue in this stream of work is computational complexity, which has made its application infeasible in high-dimensional domains [Tehrani and Ching, 2023]. More recently, Lin et al. [2015] have shown that using heuristic-based index strategies for learning yields similar performance while having the advantage of computational and cognitive simplicity. 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]. 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]. 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 and some of their features are only realized after consumption [Nelson, 1974]. There are numerous examples of such contexts, including movie recommendations, news personalization, and content recommendation on social media 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 action A_j from the action set $\mathcal{A}$. In this context, each action corresponds to consuming a distinct

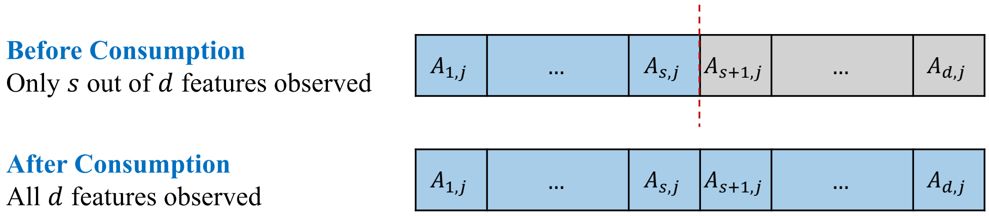


Figure 1: Illustration of search and experience features

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, 1974], 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

We extend our framework to sequential settings where users learn their preference parameters through sequential consumption of products. We take an information-theoretic perspective, recognizing that users may not have complete knowledge of their preference parameters and may experience uncertainty about them. This uncertainty is particularly likely in rich product spaces, such as content recommendation domains [Yao et al., 2022]. In such settings, users resolve their uncertainty sequentially by consuming products and learning their preferences over time. For example, a movie viewer may be uncertain about their preference for a specific sub-genre or an new visual style and thus learns by watching multiple films within that category.

Let t denote each time period and $A_{i,t}$ the product chosen by user i in period t. For notation brevity, we define $U_{i,t} = u_i(A_{i,t})$ and let $\mathcal{H}_{i,t}$ denote the prior sequence of products (actions) 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 consumer 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:

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

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

It is important to note 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. The sequential nature of learning indicates that consumers update their parameters in every time period in a Bayesian fashion. Following the literature on Bayesian learning [Ching et al., 2013, Peleg et al., 2022, Tehrani and Ching, 2023], for any $t \geq 0$, we present the consumer parameter updating from $\mu_{i,t}$ and $\Sigma_{i,t}$ to $\mu_{i,t+1}$ and $\Sigma_{i,t+1}$ as follows:

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

Algorithm 1 determines user learning given any consumption sequence $\mathcal{H}_{i,t}$. See Appendix A for the analytical derivation of the updating rules. Our goal is to examine how different consumption sequences can result in different levels of learning.

3.3 Consumer 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:

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

where one product from the set is recommended and rest of the products are not recommended.[1] We now characterize the user's decision-making process in the definition below:

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 as though there is no personalized RS, and (2) *RS-dependent user* who follows the personalized recommendation in every time period. Both types have the same learning process as described in Algorithm 1. 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 to characterize the consumption sequence for the RS-dependent users in §3.3.2.

3.3.1 Self-Exploring Consumer

In the absence of the personalized recommendation, users make decision 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:

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

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

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

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] and its excellent empirical performance in in terms of welfare [Chapelle and Li, 2011].[2]

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
    $A_{i,t} \leftarrow \text{argmax}_{A_j \in \mathcal{A}} \sum_{k=1}^{s} \tilde{\theta}_{i,k,t}^{T} A_{k,j}$    ▷ 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

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

light, making it advantageous in our later empirical analysis using the MovieLens data set.[3] Third, due to its simplicity, it is easy 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 Consumer

We now focus on user 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 assume that the recommendation system's objective is the same as that of user.[4] 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 consumers. 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 is represent 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 known 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

[3] As a robustness check, we show that our qualitative insights will not change if we use an approximate dynamic programming solution to this problem.

[4] 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.

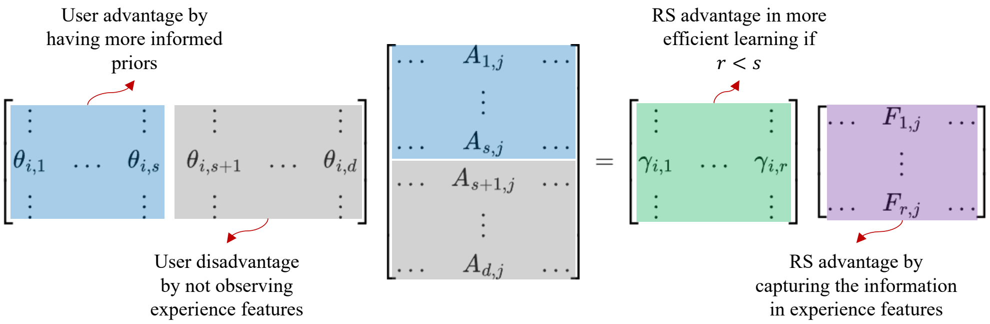

User advantage by having more informed priors
RS advantage in more efficient learning if $r < s$
$\theta_{i,1}$ … $\theta_{i,s}$
$\theta_{i,s+1}$ … $\theta_{i,d}$
$A_{1,j}$
$A_{s,j}$
$A_{s+1,j}$
$A_{d,j}$
=
$\gamma_{i,1}$ … $\gamma_{i,r}$
$F_{1,j}$
$F_{r,j}$
User disadvantage by not observing experience features
RS advantage by capturing the information in experience features

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

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, 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 consumers.

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 by the RS and users, by highlighting the potential for users to have more informed priors. These insights play a crucial role in establishing theoretical guarantees for different algorithms.

We now present the procedure that determines choice and learning 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: $\mu^{(\theta)}_{i,0}$, $\Sigma^{(\theta)}_{i,0}$, $\mu^{(\gamma)}_{i,0}$, and $\Sigma^{(\gamma)}_{i,0}$.

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 $\mu^{(\theta)}_{i,t+1}$, $\Sigma^{(\theta)}_{i,t+1}$, $\mu^{(\gamma)}_{i,t+1}$, and $\Sigma^{(\gamma)}_{i,t+1}$. The algorithm repeats this process for $\mathcal{T}$ periods.

Algorithm 3 Choice and Learning for the RS-Dependent User

```
    Input: $\mu^{(\theta)}_{i,0}, \Sigma^{(\theta)}_{i,0}, \mu^{(\gamma)}_{i,0}, \Sigma^{(\gamma)}_{i,0}, F, \mathcal{A}, \mathcal{T}$
    Output: $\mathcal{H}_{i,\mathcal{T}}, \{\mu^{(\theta)}_{i,t}, \Sigma^{(\theta)}_{i,t}\}_{t=1}^{\mathcal{T}}, \{\mu^{(\gamma)}_{i,t}, \Sigma^{(\gamma)}_{i,t}\}_{t=1}^{\mathcal{T}}$
1: for $t = 0 \to \mathcal{T}$ do
2:     $\tilde{\gamma}_{i,t} \sim N(\mu^{(\gamma)}_{i,t}, \Sigma^{(\gamma)}_{i,t})$                              ▷ RS: Sampling low-dimensional weights
3:     $A^{RS}_{i,t} \leftarrow \text{argmax}_{A_j \in \mathcal{A}} \sum_{k=1}^{r} \tilde{\gamma}^T_{i,k,t} F_k(A_j)$                 ▷ RS: Selecting recommendation
4:     $A_{i,t} \leftarrow A^{RS}_{i,t}$                                                ▷ User: Following recommendation
5:     $\Sigma^{(\theta)}_{i,t+1} \leftarrow \left( \left( \Sigma^{(\theta)}_{i,t} \right)^{-1} + \frac{1}{\sigma^2_\epsilon} A_{i,t} A^T_{i,t} \right)^{-1}$              ▷ User: Updating posterior variance
6:     $\mu^{(\theta)}_{i,t+1} \leftarrow \Sigma^{(\theta)}_{i,t+1} \left( \left( \Sigma^{(\theta)}_{i,t} \right)^{-1} \mu^{(\theta)}_{i,t} + \frac{1}{\sigma^2_\epsilon} A_{i,t} U_{i,t} \right)$        ▷ User: Updating posterior mean
7:     $\Sigma^{(\gamma)}_{i,t+1} \leftarrow \left( \left( \Sigma^{(\gamma)}_{i,t} \right)^{-1} + \frac{1}{\sigma^2_\epsilon} A_{i,t} A^T_{i,t} \right)^{-1}$              ▷ RS: Updating posterior variance
8:     $\mu^{(\gamma)}_{i,t+1} \leftarrow \Sigma^{(\gamma)}_{i,t+1} \left( \left( \Sigma^{(\gamma)}_{i,t} \right)^{-1} \mu^{(\gamma)}_{i,t} + \frac{1}{\sigma^2_\epsilon} A_{i,t} U_{i,t} \right)$        ▷ RS: Updating posterior mean
9: end for
```

A few points are worth noting about the RS-dependent user. First, we assume that the user follows the recommended product every time period. It is easy to rationalize this choice in a variety of ways due to the search cost associated with exploration. In the main analysis, we abstract away from this possibility and simply assume that the user always follows the recommendation system to illustrate the differences between the self-exploring and RS-dependent users.[5] Second, the user learning procedure in Algorithm 3 is identical to Algorithm 2, and the difference in learning only comes from the prior consumption sequence in these two settings. Finally, an implicit assumption we make here is that the RS-dependent user cannot learn from the recommendations beyond their own experience. 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.

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. Next, in §4.2, we establish regret bounds for both user types and examine the rationality of user dependence on the RS. Finally, in §4.3, we provide a general discussion of our findings and identify key empirical questions that warrant further investigation.

4.1 Main Outcomes

As discussed earlier, we are interested in two key outcomes: user welfare and user 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. 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 utility from some notion of first-best in each period and user welfare. We formally define *expected regret* as follows:

Definition 2. *Suppose that $A_i^* \in \operatorname{argmax}_{a\in\mathcal{A}} \mathbb{E}[u_i(a) \mid \theta_i]$ 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***

[5] If the RS-dependent user is rational and incurs a search cost when searching independently but no search cost when relying on the RS, then she faces a choice between searching on her own and following the RS every period. This dynamic decision process is complex, so we abstract away in the theoretical discussion.

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. 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 user's preference learning based on the* ***Shannon entropy*** *of $A_{i,t}^*$, which is defined as follows:*

$$H(A_{i,t}) = -\sum_{k=1}^{|\mathcal{A}|} P(A_i^* = A_k \mid \mathcal{H}_{i,t-1}) \log_2\left(P(A_i^* = A_k \mid \mathcal{H}_{i,t-1})\right). \tag{10}$$

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 $H(A_{i,t}) = \log(|\mathcal{A}|)$.

One key challenge with Shannon entropy as the main measure of learning is that it is not directly comparable with regret as the main measure of welfare. To overcome this challenge, we develop the measure of *counterfactual regret*, which is the regret that the user would incur at any time period, if the user makes the decision independently, absent the influence of the personalized RS. As such, if lower regret actions for a user come at the expense of user learning, we expect the counterfactual regret to be high for this user. We 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 potential loss due to RS-dependence.

4.2 Regret Bounds

We now conduct a welfare analysis of the two types of users we defined earlier and establish theoretical regret bounds for their performance. We first present the regret bounds for self-exploring users who follow Algorithm 2 in §4.2.1. We then present the regret bounds for RS-dependent users who follow Algorithm 3 in §4.2.2. Finally, in §4.2.3, we theoretically compare the two regret bounds and examine the possibility of rational dependence.

4.2.1 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 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 a d-dimensional vector of preference weights, which has existing information-theoretic regret bounds [Russo and Van Roy, 2016]. 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 and the first-best 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:*

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

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

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 identical 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., $H(A_{i,0}^*) = 0$), which automatically makes the upper bound for the second term in Equation (12) equal to zero.

Next, we further 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 $U_{i,j}^{s} \sim N(\mu_{i,s}, \sigma_{i,s})$ and $U_{i,j}^{x} \sim N(\mu_{i,x}, \sigma_{i,x})$. 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:*

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

As shown in Proposition 2, the gap depends on the variance of the utility from experience features and the number of products available. Therefore, the gap grows as the experience features play a larger role in determining a user's utility.

4.2.2 RS-Dependent Users

We now turn to establishing regret bounds for the RS-dependent user. The regret analysis for RS-dependent users is a bit more subtle. In this case, the user follows the recommendation from 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 the factor information. Under Assumption 1, the first-best that the RS could obtain is the same as the first-best in Definition 2. Therefore, the following proposition 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 $H(A_{i,0}^{\mathrm{RS}})$ denote the Shannon entropy of the prior distribution of optimal action for the personalized RS. The regret bound for the personalized RS is as follows:*

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

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.

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].

4.2.3 Regret Comparison and the Possibility of Rational 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 by reducing the problem's dimensionality from d to r, 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 between regret under self-exploring and RS-dependent users has a lower bound as follows:*

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

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 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. Finally, the bound in Corollary 1 highlights that even in the absence of search costs and with time-consistent preferences, users will rationally develop a dependence on the recommendation system (RS). This is particularly notable because our dependence results stem solely from the algorithm's information advantage rather than time-inconsistent preferences or search costs—factors that have been central to prior explanations in the literature [Allcott et al., 2022, Korganbekova and Zuber, 2023].

4.3 Discussion

In this section, we provide a general discussion and highlight the remaining questions that are empirical. As shown in Corollary 1, the welfare gains from RS dominate the downsides as we increase the number of periods T. However, in many real settings, a single user is not available for infinitely many periods. Therefore, it is possible that the self-exploring user has better welfare outcomes in finite periods because (1) the gap g can be small, and/or (2) the RS algorithm takes a while to stabilize and offer bad recommendations in the beginning. The latter could happen if the effective rank r is high or the RS has a fully uninformative prior. Thus, a finite-sample evaluation of the welfare gain/loss from RS-dependence is an empirical question that depends on the actual values of the gap g and rank 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, thereby providing some variation. 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 gets to the exploitation stage quickly, and the amount of exploration may be insufficient for user learning. In particular, if the exploitation stage of the algorithm recommends a narrow set of products, the variation may be limited for users' preference learning. 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 evaluate the amount of learning and counterfactual regret for both self-exploring and RS-dependent users.

5 Empirical Framework

As discussed in §4.3, we turn to our empirical framework to examine the finite-sample properties of self-exploring and RS-dependent algorithms in terms of welfare and learning. In particular, we want to develop an empirical framework to quantify the RS algorithm's information advantage, which can be seen as the significant cause of RS-dependence. We further want to investigate the learning implications of this RS-dependence and assess the possibility of finding policies to balance the potential trade-off between welfare and learning.

To achieve these goals, we need an empirical setting where (1) we have preference data on user-product pairs, (2) we observe a rich set of search features for products to estimate consumer preferences for them and simulate different learning patterns under different algorithms, and (3) we can embed a state-of-the-art personalized recommendation system that learns complex consumer preferences and offers useful recommendations.

To satisfy these three requirements, we turn to the MovieLens 1M dataset.[6] The MovieLens data set contains over one million consumer ratings corresponding to a total of 6,040 distinct consumers and 3,706 unique movies, which we use as a measure of consumer preferences. In addition to ratings, the data provide information about the movies, such as genre and themes, as well as a large array of tags associated with each movie, which we use as the set of search features for movies. Lastly, the MovieLens data set has been widely used as the benchmark dataset for research on personalized recommendation systems, thereby ensuring that we can obtain state-of-the-art personalized recommendation systems.

In this section, we first present our empirical strategy in §5.1. We then present results on the welfare gains from personalized RS due to its information advantage in §5.2. Next, in §5.3, we examine the learning outcomes for self-exploring and RS-dependent users. Finally, in §5.4, we discuss the policy implications and search the space of policies for one that balances the trade-off between welfare and learning objectives.

5.1 Empirical Strategy

In order to satisfactorily answer the empirical questions, we face several challenges. We discuss these challenges along with their corresponding empirical strategy in the following sections. Before discussing the challenges, 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 consumers and J movies. We present our assumption that links ratings to utility as follows:

Assumption 2. *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 is about the set of available search features. Since we observe detailed movie tags, we use them as the main search features of the movie. These tags include attributes such as originality, great ending, and good soundtrack (please see Appendix B for

[6] This dataset is publicly accessible and can be obtained from `https://grouplens.org/datasets/movielens/1m/`.

the list of all movie tags). Normally, 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. We assume that the top d most frequently used tags characterize the dimensionality of user preferences. We set $d = 50$ and present our formal assumption as follows:

Assumption 3. *The matrix of 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.*

It is worth emphasizing that the choice of $d = 100$ is arbitrary and we ensure the robustness of our findings with different values of d.

The final assumption we make is about the stability of true user preferences. As before, we let $\Theta_{[d \times N]}$ denote the user preferences for all N consumers. Our assumption about stable preferences is presented as follows:

Assumption 4. *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 worth noting that the user in our setting still learns these preferences through experience. That is, although the actual preference is stable, the user's belief about it can change over time.

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 100 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 3, 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.[7]

5.1.2 Model-free Regret Measurement

Empirically measuring regret for algorithms different from the one used in the data is fundamentally challenging. On the one hand, one could impute the ratings for movies that

[7] 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. The qualitative insights remain unchanged.

are not rated, but this approach requires highly accurate model-based estimates and has the pitfall of information leakage between the algorithm used for imputation and the ones we want to evaluate. On the other hand, one could remain within the set of movies that each user has already rated, but this naturally limits the ability of algorithms that find strong products that are not chosen by the user in the observed data. Further, the context of movie recommendation is slightly different from the theoretical setting where the action set $\mathcal{A}$ is fixed. It is more conceivable to assume that once content is consumed, one does not receive the same utility from consuming it. This makes the set of available products for each policy different, making an apples-to-apples comparison between algorithms impossible.

To overcome these challenges, we use a random hold-out set of 20 movies that are rated by users in our test set to ensure we have the actual outcomes and the first-best among them. We then exclude these 20 movies from the set of movies that the user's ratings, so there is no information for the algorithm on the user's rating for these hold-out 20 movies. It is worth emphasizing that the sets of 20 hold-out movies is fixed within users but can vary across users in our test data. Therefore, we can train any algorithm on the set without the 20 hold-out movies and evaluate its regret and counterfactual regret on the hold-out set of movies in a model-free manner that directly uses observed ratings by the user.

5.1.3 Embedding the Personalized RS

To evaluate outcomes for the RS-dependent user, we need to use a personalized RS that updates its recommendations as it gathers 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 [2018a], we use a matrix factorization algorithm for our personalized recommendation system, which is widely used in the industry. We then train this state-of-the-art recommendation system and predict ratings for a hold-out set of 20 movies. The RS recommends the product with the highest predicted rating, and the RS-dependent user will consume it.

5.2 Algorithm's Information Advantage and Regret Performance

We now want to examine the algorithm's information advantage and the welfare gains from following the RS. This analysis is complementary to the theoretical analysis provided in §4.2 as it uses real data to verify the finite-sample performance of different algorithms. In particular, we want to know (1) how quickly the RS learns good recommendations on the hold-out set of 20 movies and (2) whether the recommendations are better than the one users could find based on the search features. The former relates to the low-dimensional learning through the factor model, whereas the latter characterizes the gap g, which is an inherent gap in the first-best identified by the RS compared to the user.

For each user in our test set, we maintain the actual sequence of ratings in the data for the training set of movies that does not include the hold-out set. At each time period, our personalized RS updates its parameters and offers a new set of predictions on the hold-out set. Using those predictions, we can evaluate the RS algorithm's performance in terms of regret and other relevant metrics. Since this algorithm mimics what an RS-dependent user would consume on the hold-out set, we call it *RS-Dependent User* and compare it with 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. Since there is no learning in this benchmark, we expect the performance measures to be fixed.
- **Self-Exploring User with Known Preferences:** This algorithm predicts ratings using the user's known 51-dimensional preference vector, which comprises the linear preference on the top 50 frequently mentioned tags and the aggregated rating. This preference vector is obtained by regressing the user's actual ratings from all watched movies onto these 51 features. The user's preference is assumed to be known from the outset and remains constant over time. This serves as the best performance achievable by the self-exploring user. As such, any gap between this algorithm and the personalized RS shows the gap factor g characterized in our theoretical analysis.
- **Self-Exploring User:** This algorithm is the same as the previous one, with the difference that the user is learning according to Algorithm 1. Similar to the RS-Dependent User, we maintain the actual sequence of ratings in the data and update the preference parameters for search features over time. The algorithm then predicts ratings for all movies in the hold-out set based on their features. Since this algorithm mimics the self-exploring user's choice on the hold-out set, we call it Self-Exploring User.

Figure 3 shows the performances of all four algorithms in 400 periods aggregated over 100 simulations in total. A few important insights emerge from this figure. First, we find that RS outperforms all three benchmarks in the hold-out test set, showcasing its ability to capture and leverage salient patterns in user behavior. Second, we note a persistent gap between the regret under RS-Dependent User compared to the Self-Exploring User with

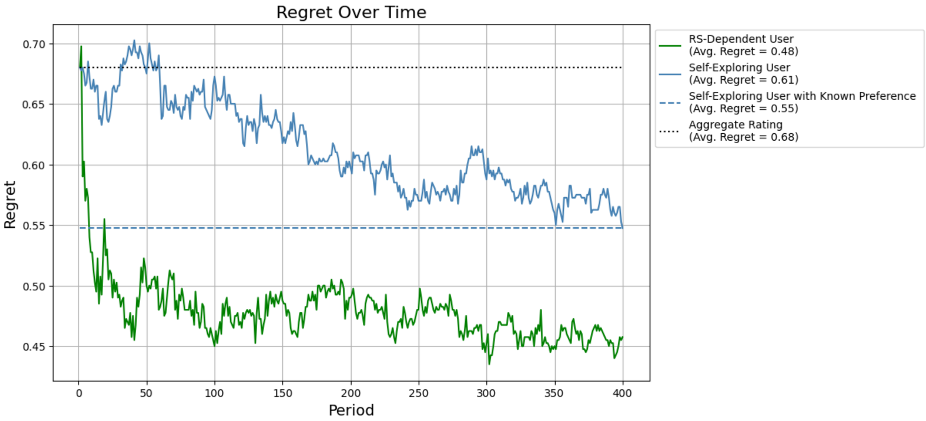


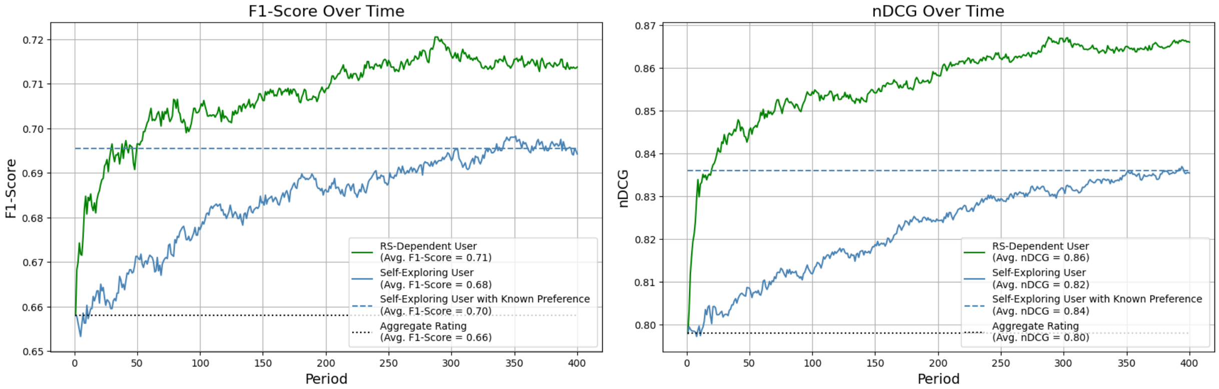
Figure 3: Comparing Learning Models- Regret (All Learning Models)

Known Preferences, which highlights the fundamental information advantage of the RS and the gap between the first-best and the first-best based on search features. Third, we note that the RS learns to perform better than the Self-Exploring User with Known Preferences quickly, which underscores its efficiency due to the low-rank learning. This is an important empirical finding as we do not impose any specific rank assumption that forces r to be lower than 51: the personalized RS algorithm finds the rank in a data-driven manner.

Next, we focus on other performance accuracy measures that capture how well each algorithm predicts the ranking. In particular, we use two commonly used accuracy measures: (1) *F1-Score* that compares the top 10 model recommendations with the optimal top 10 recommendations in the hold-out set [Chang et al., 2015], and (2) *Normalized Discounted Cumulative Gain (nDCG)*, which is an evaluation metric used to assess the quality of a ranked list of items by comparing the actual ranking with the ideal (perfect) ranking [Järvelin and Kekäläinen, 2002]. Both these measures are widely used in the literature on recommendation systems and collaborative filtering [Koren et al., 2021].

Figure 4 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.

In summary, our findings highlight the information advantage of the personalized RS algorithm that provides a rational account for the user to become dependent on algorithms. This is important because the prior literature has used accounts such as time-inconsistent preferences and search costs to justify the algorithmic dependence [Allcott et al., 2022, Korganbekova and Zuber, 2023]. However, as suggested in our analysis, there is a stark limitation for users in the context of experience goods, which rationally forces them to become



(a) F1-Score

(b) nDCG

Figure 4: Performance of different algorithms using F1-Score and nDCG metrics

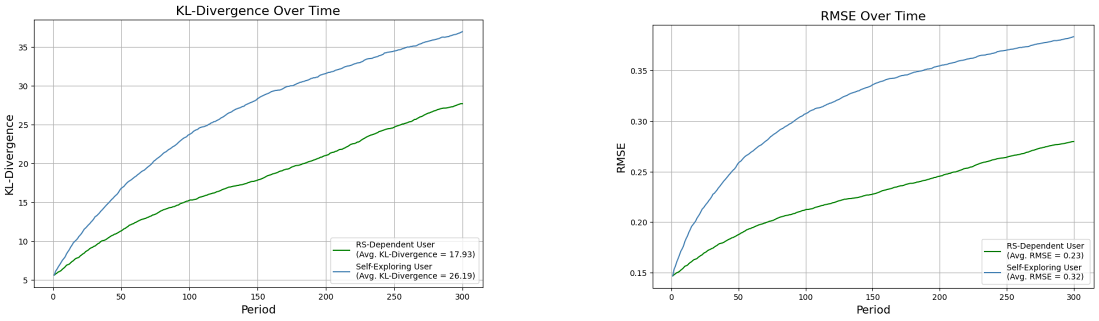
dependent on personalized RS, even with zero search cost and time-consistent preferences.

5.3 Implications for Users' Preference Learning

As shown in our theoretical and empirical analysis so far, RS-dependence has substantial welfare gains for users, making it a rational strategy for them. However, dependence on the RS has implications for users' preference learning, as it would change their prior experience from which they resolve their uncertainty. In this section, we examine the implications of RS-dependence for users' learning and their independent decision-making ability in the absence of RS.

We focus on the two user types studied in this paper: self-exploring users and RS-dependent users. Unlike §5.2 where we maintain the user's consumption sequence, in the current analysis, we change 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. That is, for each user, we only allow consuming the movies that are already watched and not in the hold-out set of 20 movies.

We start from a prior set of preferences that is consistent with the average taste. That is, the user's prior weight for the aggregated rating feature is one, and her weight is zero for all other features. Every round, each user type (self-exploring and RS-dependent) chooses a movie from the set, consumes it, realizes the utility from it, and updates their preference parameters. We can measure learning by comparing the updated preferences with the prior preferences. If the system is *learning*, we often expect a higher difference between the updated preferences and the prior. We use two measures to quantify this difference: (1) Kullback-



(a) KL Divergence

(b) RMSE

Figure 5: User learning under different algorithms

Leibler (KL) divergence, which is widely used to measure the difference between distributions, and (2) the Root Mean Square Error (RMSE) of the difference in mean parameters.

Figure 5 shows the results for both learning measures under self-exploring and RS-dependent users. As shown in both figures, the difference from prior grows faster under the self-exploring user compared to the RS-dependent user, suggesting that the self-exploring user learns more. It is worth emphasizing that in our analysis, both users start from the same set of priors, so the difference captures the relative difference.

The finding in Figure 5 suggests that following personalized recommendations comes at the expense of user learning. However, as discussed earlier, it is hard to quantify the loss due to insufficient learning in terms of welfare. Our *counterfactual regret* measure helps overcome this challenge. This measure illustrates how the personalized algorithm affects users' independent decision-making ability. We measure counterfactual regret using Equation (11) in Definition 4. We use the same idea of having a hold-out set of movies as the action set available at each time period. We consider three separate user types: (1) Self-Exploring User who chooses on their own and updates their preferences, (2) RS-Dependent User who relies on the RS to choose movies, and (3) RS-Dependent User Counterfactual who followed RS in all prior periods but chooses on their own at each period. If RS-dependence acts as a barrier to learning, we should see a higher regret for RS-Dependent User Counterfactual.

We present the results of this practice in Figure 6. The two lines for RS-Dependent User and Self-Exploring User are similar to those shown in Figure 3, with the difference that the sequence of movies consumed has changed. The red line shows the counterfactual regret, which measures how much regret the RS-dependent incurs as an independent decision-maker at any point. This figure shows a persistent gap between the expected regret and expected counterfactual regret for RS-dependent users. Importantly, the expected counterfactual regret for the RS-dependent user is higher than the expected regret for the self-exploring user. This finding suggests that although following personalized recommendations reduces expected regret, it comes at the expense of users' independent decision-making ability, an essential ability for safeguarding against potential adversarial AI attacks.

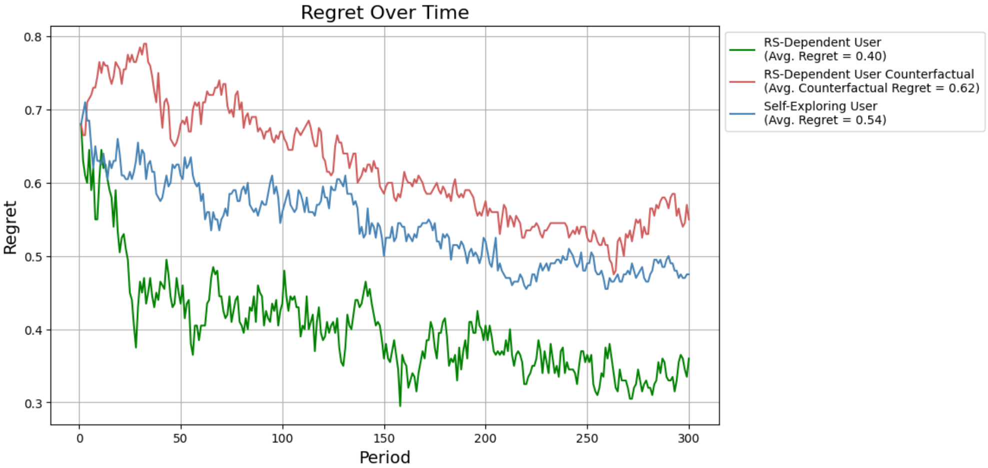


Figure 6: Comparison of regret and counterfactual regret for different algorithms

5.4 Policy Implications

Our findings from §5.2 and §5.3 highlight an important trade-off between welfare and learning. This trade-off gives rise to the following question: is there a policy that finds the right balance between these two objectives?

We consider a specific class of policies that add randomness to the availability of the recommendation system. As a result of this random availability, RS-dependent users need to make their own decisions in some time periods. This likely results in an increase in users' preference learning without fully eliminating the benefits of the personalized RS. Specifically, we operationalize these policies using a single parameter p that controls the probability by which the recommendation system will be unavailable. When the personalized algorithm is available, the RS-dependent consumer follows the algorithm; otherwise, she chooses by herself, given her updated preferences. It is worth noting that when the personalized recommendation system is always available as in $p = 0$, the outcomes correspond to those for the RS-dependent consumer. On the contrary, when the recommendation system is not available, as in $p = 1$, our outcomes correspond to those of the self-exploring consumer.

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. That is, for each random availability policy, we change the

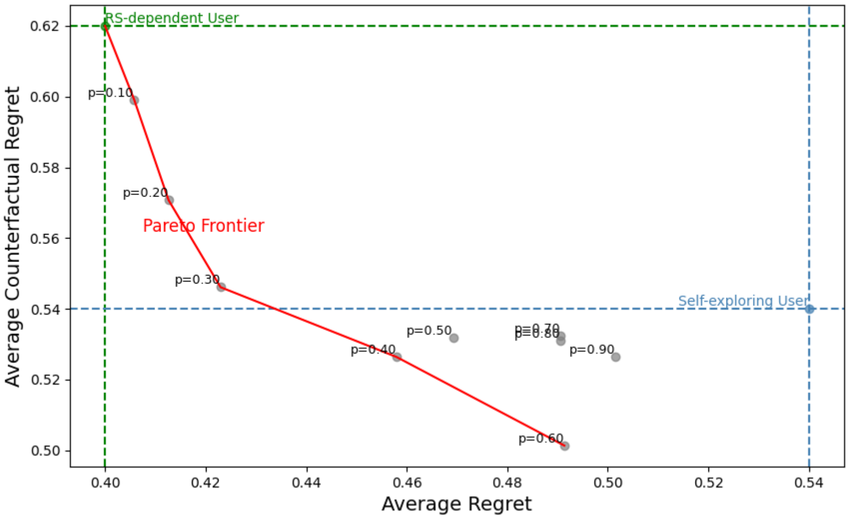


Figure 7: The performance of random availability policies in terms of regret and counterfactual regret

order of consumption according to the policy and evaluate its regret and counterfactual regret performance on the hold-out set. Figure 7 presents the results from this practice. This figure clearly illustrates the trade-off between regret and counterfactual regret, which is closely connected with the well-known exploration-exploitation trade-off. We show the Pareto Frontier of different random availability policies. Notably, some random availability policies Pareto dominate the self-exploring policy, achieving lower expected and counterfactual regret. This finding suggests that random availability policies can offer superior alternatives to strict data protection policies that completely ban personalized recommendation systems.

Importantly, our results reveal 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: it results in a counterfactual regret comparable to that of the self-exploring user while maintaining a level of expected regret comparable to that of the RS-dependent user. These findings suggest that even relatively simple policies, such as random availability, can effectively balance the trade-off between welfare and learning. Thus, this finding offers important implications for platforms and users who want to self-regulate.

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 valuable 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. Unlike prior research, which has primarily focused on issues such as privacy, fairness, and polarization, our work shifts attention to algorithmic dependence and its implications for independent decision-making and autonomy. Therefore, our work contributes to the policy debate 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.

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.

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*, 2018a.

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

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

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.

T. Erdem, M. P. Keane, T. S. Öncü, and J. Strebel. Learning about computers: An analysis of information search and technology choice. *Quantitative Marketing and Economics*, 3: 207–247, 2005.

T. Erdem, M. P. Keane, and B. Sun. A dynamic model of brand choice when price and advertising signal product quality. *Marketing Science*, 27(6):1111–1125, 2008.

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.

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.

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.

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.

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.

A. Peleg, N. Pearl, and R. Meir. Metalearning linear bandits by prior update. In *International Conference on Artificial Intelligence and Statistics*, pages 2885–2926. PMLR, 2022.

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.

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

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.

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.

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. At any given time t, we have an updated belief about θ_i which is normally distributed as:

$$\theta_i \sim \mathcal{N}(\mu_{i,t}, \Sigma_{i,t}) \tag{17}$$

New data comes in the form of $A_{i,t}$ (a $d \times 1$ vector of action attributes at time t) and $U_{i,t}$ (the utility observed from taking action $A_{i,t}$). Given the new observation $(A_{i,t}, U_{i,t})$, the likelihood function (probability of observing $U_{i,t}$ given $A_{i,t}$ and θ_i) is normal with mean $A_{i,t}^T\theta_i$ and variance σ_ϵ^2. The precision (inverse of the covariance matrix) of the prior distribution for θ_i is $\Sigma_{i,t}^{-1}$, and the precision of the new data is $\frac{1}{\sigma_\epsilon^2}A_{i,t}A_{i,t}^T$.

Therefore, the updated precision matrix (posterior precision) is the sum of the prior precision and the precision of the likelihood:

$$\Sigma_{i,t+1}^{-1} = \Sigma_{i,t}^{-1} + \frac{1}{\sigma_\epsilon^2}A_{i,t}A_{i,t}^T \tag{18}$$

The updated mean combines the prior mean and the new data, weighted by their respective precisions. The weight for the prior mean is its precision $\Sigma_{i,t}^{-1}$, and the weight for the new data $U_{i,t}$ is $\frac{1}{\sigma_\epsilon^2}$:

$$\mu_{i,t+1} = \Sigma_{i,t+1}\left(\Sigma_{i,t}^{-1}\mu_{i,t} + \frac{1}{\sigma_\epsilon^2}A_{i,t}U_{i,t}\right) \tag{19}$$

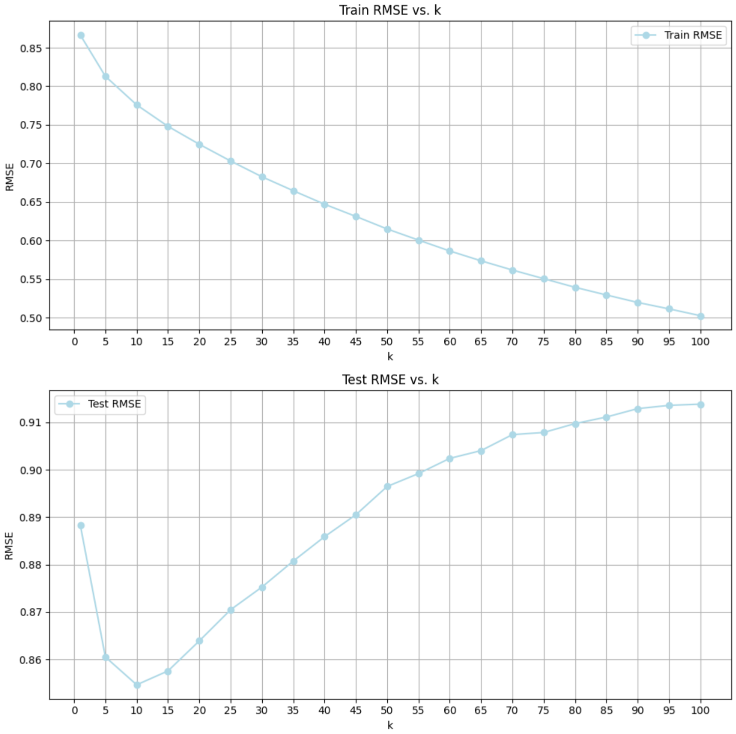
□

B Details of the Empirical Framework

In this section, we discuss two key parameters that we use in our empirical analysis: d and r. We set $d = 100$, corresponding to the 100 most frequently mentioned tags in the Movie Lens data, as shown in Table A1. We construct the action matrix $A_{[d\times J]}$ for all $J = 3080$ movies. For r, we use a data-driven approach and select the best-performing low-rank matrix completion on the held-out set. As shown in Figure A1, $r = 10$ achieves the lowest RMSE on the held-out test set.

Table A1: Top 100 Movie Lens Data Tags

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



Train RMSE vs. k

Test RMSE vs. k

Figure A1: Training and Test Sample RMSE for Matrix Completion with Different Ranks