Personalized Algorithms and the Virtue of Learning Things the Hard Way

Omid Rafieian* Cornell University

Si Zuo* Cornell University

Abstract

Personalized recommendation systems are now an integral part of the digital ecosystem. However, consumers' increased dependence on these personalized algorithms has heightened concerns among consumer protection advocates and regulators. Past studies have documented various threats personalization algorithms pose to different aspects of consumer welfare, through violating consumer privacy, unfair allocation of resources, or creating filter bubbles that can lead to increased political polarization. In this work, we bring a consumer learning perspective to this problem and examine whether personalized recommendation systems hinder consumers' independent decision-making ability, an important construct given the growing fear of adversarial AI. We develop a utility framework where consumers learn their preference parameters through experience to examine the effect of personalized algorithms on the learning process. We establish regret bounds for different types of consumers based on their dependence on the personalized algorithm. We then run a series of calibrated simulations and show that although personalized algorithms increase consumer welfare for consumers who rely more on personalized recommendations by offering better recommendations, these consumers do not sufficiently learn their own preference parameters and make worse decisions in the absence of recommendation systems. Inspired by the trade-off between consumer learning and welfare, we introduce the notion of counterfactual regret, which is the regret incurred by the consumer when the personalized algorithm is unavailable. Finally, we examine a variety of consumer protection policies that aim to find a balance

*Please address all correspondence to: or83@cornell.edu and sz549@cornell.edu.

between these two outcomes and find policies that can achieve good welfare and learning outcomes.

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

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 consumers 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 consumers' choices easier by reducing their search costs. These systems are designed to effectively narrow down options in real time and guide consumers towards products or services that best align with their preferences and needs. By doing so, a personalized recommendation system ensures that consumers 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 is increasing scrutiny regarding their potential pitfalls in terms of privacy [Johnson, 2022], fairness [Lambrecht and Tucker, 2019, Barocas et al., 2019], and political polarization [Dandekar et al., 2013, Flaxman et al., 2016].

In this work, we bring a consumer learning perspective to this problem and examine how personalized recommendation systems affect consumer learning. We study digital contexts where consumers consume content and learn their preferences through experience. For example, an online news reader experientially learns how the features of an article presented on the homepage map into her utility from consuming that article. Personalized recommendations influence the process through which consumers 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 consumer learning.

Understanding the impact of personalized recommendation systems on consumer learning is important from a consumer protection standpoint because it provides insights into the primary tool digital consumers can use for decision-making. In particular, if consumers lack a comprehensive understanding of their preferences, they would make worse decisions in the

absence of recommendation systems. This results in a feedback loop wherein consumers rely overly on recommendation systems and learn less about their preferences. More importantly, consumers without sufficient learning of their preferences are more susceptible to digital manipulations by adversarial AI, a topic of growing interest among policy-makers [Kusnezov et al., 2023]. An adversarial player or platform can take advantage of consumers' insufficient learning and exploit them in various welfare-reducing ways. Together, we view consumer learning as an important yet understudied component of consumer welfare that allows us to design better policies around personalized algorithms.

In this paper, we study the interplay between personalized recommendation systems and consumer learning and aim to answer the following questions:

1. What is the value of personalized recommendation systems? How does the algorithm's information advantage translate into better recommendations?

2. How does the presence of a personalized algorithm affect consumer learning? How can we quantify this impact?

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

To answer these questions, we face two broad challenges. First, we need a theoretical framework that allows for dynamic preference learning through experience. In particular, our framework needs to capture learning through exploration by both the consumer and the recommendation system. Importantly, we want our framework to naturally reflect the algorithm's information advantage as it has access to the information from other consumers. Second, we need an empirical framework that obtains estimates of consumer preferences from sparse data and delivers counterfactual policy evaluation. Specifically, we want clear measures for consumer welfare and learning to evaluate outcomes under different counterfactual policies. Further, we want our framework to be adaptable enough that we can directly embed state-of-the-art personalized algorithms in our framework.

To address our first set of challenges, we build a general linear utility framework where consumers have some preference parameters over the space of item features. To characterize consumers' learning, we turn to the literature on Bayesian learning and model consumers who update the posterior distribution of their preference parameters. For consumer decision-making, we use the Thompson Sampling approach as it incorporates consumers' Bayesian learning and is behaviorally plausible [Schulz et al., 2019, Mauersberger, 2022]. Under Thompson Sampling algorithms, consumers 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 consumer. Lastly, to quantify the impact of recommendation systems on consumer learning, we use two measures that are commonly used in the literature on linear bandits: (1) Shannon entropy that accounts for the amount of learning by the consumer, and (2) expected regret, which is the overall welfare loss compared to the first-best optimal choice throughout.

For the second set of challenges, we develop an empirical strategy to learn consumer preferences from sparse data. We turn to the rich literature on collaborative filtering and matrix completion that often serve as the basis for personalized recommendation systems in practice. This strategy allows us to impute all the utility values for any pair of consumer-item. We can then project these imputed utilities on the item attributes to obtain the underlying consumer preferences. Once we have such consumer-specific preference parameters, we can simulate consumer learning for any sequence of movies consumed and evaluate consumer learning and expected regret measures. This allows us to quantify to what extent the presence of recommendation systems acts as a barrier to consumer learning [Lattimore and Szepesvári, 2020].

To illustrate the impact of recommendation systems on consumer learning, we focus on two types of consumers: (1) self-exploring consumers who make decisions on their own as though there is no recommendation system, and (2) recommendation-system-dependent (RS-dependent henceforth) consumers who follow the recommendations provided by the recommendation system. Theoretically, we use the information-theoretic regret bounds established in Russo and Van Roy [2016] to compare regret bounds for self-exploring and RS-dependent consumers. We find that if both consumer types start from the same prior distribution over the optimal action, the regret bound for the RS-dependent consumer is lower than that for the self-exploring consumer. Specifically, we show that the rank of the matrix plays an instrumental role in reducing the regret bound for RS-dependent consumers: the ratio of the regret bound for RS-dependent to self-exploring consumers is the square root of the ratio of the algorithm's rank to the dimensionality of the feature space. Therefore, we establish that the personalized algorithm increases consumer welfare by offering better recommendations.

Next, we examine our second research question on how the personalized algorithm affects consumer learning. Unlike finding regret bounds, evaluating learning outcomes is an empirical

question that depends on the distribution of preferences. We apply our empirical framework to the MovieLens data set, which is the main public data set used as a benchmark for recommendation systems. Our empirical exercise confirms the theoretical result that consumer welfare is higher for RS-dependent consumers compared to self-exploring consumers. However, we find that self-exploring consumers have a lower Shannon entropy of the distribution of optimal action, which indicates a higher degree of learning and lower uncertainty about preference parameters. Notably, we find that the Shannon entropy reaches a steady state at a higher value for RS-dependent consumers than self-exploring consumers, indicating a lower level of learning. Together, although personalized algorithms help consumers 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 consumers engage in.

Motivated by the trade-off between consumer 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 consumer in each period if they make decisions independently without the help of the personalized algorithm. As such, a consumer 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 consumers and show that while these consumers enjoy a low regret due to following personalized recommendations, they have a higher counterfactual regret than self-exploring consumers because they become worse independent decision-makers in the absence of personalized algorithms.

An immediate concern that emerges from our findings is the potential susceptibility of RS-dependent consumers to platform or third-party manipulations. That is, if the consumers have great uncertainty about their preference parameters, adversarial players can use this information to manipulate these consumers. This concern gives rise to the potential consumer protection policies to mitigate such issues. To that end, we evaluate a series of viable consumer protection policies in terms of both regret and counterfactual regret. In particular, we consider a class of policies where 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 find the right balance in the trade-off between consumer welfare and learning and achieve good outcomes in terms of both regret and counterfactual regret. Further, we focus on a self-regulatory policy wherein the personalized algorithm incorporates consumer learning as

part of the objective through some weights. Interestingly, we find that there are self-regulatory policies that achieve great counterfactual regret without sacrificing consumer welfare and the quality of personalized recommendations.

In summary, our paper provides several contributions to the literature. Substantively, we present a comprehensive study of the effect of personalized recommendation systems on consumer learning through a series of theoretical and empirical analyses. We document that even when personalized algorithms enhance consumer welfare by increasing the match value between consumers and recommended products, they can negatively affect consumers' 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 impact of these systems on consumer learning has been overlooked. The impact of personalized algorithms on consumer learning is particularly important as learning is the primary tool consumers have for independent decision-making. As such, our work extends the policy debate on the societal impact of personalized recommendation systems by bringing a consumer learning perspective, which helps us develop more effective policies to empower consumers. Methodologically, we build a framework that allows us to quantify the potential welfare loss due to underexploration in the context of personalized recommendation systems. A key innovation of our framework is in capturing the information advantage of the personalized algorithm through a low-rank assumption. Our framework is general and can be applied to a variety of domains that involve sequential decision-making. In particular, we develop a measure of counterfactual regret that can be used as a benchmark for studies related to the impact of adversarial AI. Finally, from a policy standpoint, we find that there are exploration-based policies that are simpler than the proposed algorithmic auditing policies and achieve desirable outcomes in both consumer welfare and learning. Additionally, we show that the platform can use more complicated self-regulatory algorithmic solutions to improve both outcomes further.

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 [Linden et al., 2003, Mazumder et al., 2010, Koren et al., 2021], models to estimate Conditional Average Treatment Effects [Athey and Imbens, 2016, Shalit et al., 2017, Wager and Athey, 2018, Nie and Wager, 2021], and personalized policy learning methods [Swaminathan and Joachims, 2015]. Related applied work in this domain has focused on different aspects of personalization, such as empirical

gains from personalization in a variety of domains [Hauser et al., 2009, Urban et al., 2014, Ascarza, 2018, Simester et al., 2020a,b, Rafieian and Yoganarasimhan, 2021, Liberali and Ferecatu, 2022, Yoganarasimhan et al., 2022, Rafieian, 2023, Dubé and Misra, 2023, Rafieian et al., 2023], the interplay between personalization and consumer protection policies [Goldfarb and Tucker, 2011, Johnson et al., 2020, 2023, Bondi et al., 2023], 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]. In particular, our paper is related to Aridor et al. [2020] who model consumer learning and use MovieLens data to test some theoretical predictions. Our paper differs from Aridor et al. [2020] as we allow for experiential learning of preference weights for item attributes, which makes our setting more consistent with the literature on sequential decision-making. Further, unlike Aridor et al. [2020] who explicitly impose a similarity structure on the item space, we organically capture similarities through item features. More broadly, our work adds to the literature on personalization by bringing a consumer learning view to this problem, which has been largely ignored in the prior work on personalized algorithms. We demonstrate how the negative impact of personalized algorithms on consumer learning can result in poor decision-making by consumers in the absence of algorithms.

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 study the product recommendation problem when consumers engage in sequential search as in Weitzman [1979] but learn their preference weights. Although our work is closely related to Dzyabura and Hauser [2019], it differs in several important ways. First, consumers in Dzyabura and Hauser [2019] update their preference weights for product attributes during costly search, whereas in our framework they realize their utility after consumption. Although preference weight learning during search is more plausible for the contexts they study where the consumer's search ends with a decision about purchase (e.g., buying a house, college applications), experiential learning is better suited for the content consumption problem we study as consumers consume hundreds or thousands of items over time (e.g., reading articles, watching movies). As such, we explicitly abstract away from the search process and

focus on consumer learning through repeated interaction.[1] Second, we focus on algorithmic recommendations and capture their information advantage by systematically pooling consumer data, whereas Dzyabura and Hauser [2019] focus more on expert recommendation in a search process, such as a real-estate agent helping home buyers. As such, the recommendation system in our setting more closely reflects the algorithms used by platforms and characterize the information advantage in a more data-driven manner.

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, Tehrani and Ching, 2023]. An important issue in this stream of work is computational complexity, which has made its application infeasible in high-dimensional domains [Ching et al., 2013, 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.

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

[1]It is worth emphasizing that our findings are robust to different search processes. Our choice to abstract away from search is only for simplicity and tractability.

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.

3 Theoretical Framework

3.1 Background

Suppose that there is a consumer who consumes content, where each piece of content is characterized with a set of features available to the consumer prior to consumption. Upon consumption, the consumer receives utility which is a function of content features and some idiosyncratic stochastic component. The consumer does not necessarily know the function but learns it through experience. The other key player in our context is a personalized algorithm whose main task is to learn individual-level consumer preferences and provide each consumer with effective recommendations. This setting reflects numerous digital content consumption contexts as follow:

- *News Consumption:* A news reader observes features of a set of articles (e.g., length, topic, headline) and decides whether or not to consume the content. Once she consumes the article, she realizes the utility from it and update her beliefs about the preference weights for different features. A personalized algorithm aims to learn the news reader's preferences and offer articles that the reader is more likely to consume.
- *Movies:* A viewer on a streaming platform wants to choose a movie to watch from a set of movies. Each movie is represented with a set of features or attributes over which the viewer has preferences. Given the quality differences between movies with similar features, there is inherent noise in the link between attributes and the overall utility. The viewer has uncertainty over these links but learn the preference weights through experience. The personalized algorithm aids the viewer by learning the viewer preferences and offering them movie recommendations.
- *Social Media:* A user on a social media platform (e.g., TikTok, Instagram) chooses content to consume from a set of contents whose some attributes are available to users prior to consumption. The user learns how these attributes are linked to her utility through consumption. The personalized algorithm used by the platform offers recommendations to make users more engaged.

As presented in the examples above, consumer learning is common in different content consumption settings and personalized algorithms play an instrumental role in shaping how

consumers learn. In particular, if the personalized algorithm reduces the exploration needed for consumer learning, relying on it likely leads to lower degrees of learning. On the other hand, these algorithms all need to explore to be able to offer good recommendations, which can increase learning for consumers relying on personalized algorithms. Our goal is to understand how the presence of a personalized recommendation system interferes with consumer learning. To do so, we develop a model with a consumer and a personalized recommendation system, where the consumer learns their own preference parameters by interacting with the system, and the personalized recommendation system also learns consumer preference based on their activity logs as well as those entered by other consumers in the database. To illustrate the impact of the recommendation system, we focus on two types of consumers: (1) self-exploring consumers who ignore the recommendation system and make their own decisions, and (2) RS-dependent consumers who follow the algorithmic recommendations. Both groups learn their preferences through experience, but only one relies on the algorithmic recommendations. In this section, we first describe how we incorporate consumer learning and choice in our framework and then present a welfare analysis.

3.2 Consumer Learning

We consider a generic utility framework wherein consumer i receives utility from taking action A_j from action set $\mathcal{A}$. Each action is characterized by a d-dimensional set of attributes, i.e., $\mathcal{A} \subset \mathbb{R}^d$. For example, an action can be a movie with d attributes (e.g., genre, runtime). The consumer-specific vector of preferences $\theta_i \in \mathbb{R}^d$ characterize the utility from action $A_j \in \mathcal{A}$ as follows:

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

where $\epsilon_{i,j}$ denotes the error term that comes from a mean-zero Normal distribution with known variance σ_ϵ^2, which implies that $E[u_i(a)] = \theta_i^T A_j$.

We extend our framework to sequential settings where consumers learn their preference parameters through experience. Let t denote each time period and $A_{i,t}$ the action chosen by consumer 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 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 update the preference parameters at the end of each time period t given the prior sequence $\mathcal{H}_{i,t}$ according to the

following rule:

$$\mu_{i,t} = \mathbb{E}[\theta_i \mid \mathcal{H}_{i,t}] \tag{2}$$

$$\Sigma_{i,t} = \mathbb{E}\left[(\theta_i - \mu_{i,t})(\theta_i - \mu_{i,t})^T \mid \mathcal{H}_{i,t}\right] \tag{3}$$

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 consumer 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 consumer'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 action 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 action from the set is recommended and rest of the actions are not recommended.[2] We now characterize the consumer's decision-making process in the definition below:

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

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

To isolate the impact of personalized algorithms on consumer learning, we consider two types of consumers: (1) self-exploring consumer who makes decisions on their own as though there is no personalized recommendation system, and (2) RS-dependent consumer who follows the recommendation system (RS) in every time period. Both types have the same learning process as described in Algorithm 1. In what follows, we first characterize consumer choice for self-exploring consumers in §3.3.1. We then present how the recommendation system provides recommendations to characterize the consumption sequence for the RS-dependent consumer in §3.3.2.

3.3.1 Self-Exploring Consumer

In the absence of the personalized recommendation, consumers make decision on their own. Given the utility framework in Equation (1), a forward-looking utility-maximizing consumer wants to optimize the overall utility over T periods. This naturally motivates consumers 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 consumer as maximizing the discounted expected utility stream as follows:

$$\underset{\pi}{\operatorname{argmax}}\, \mathbb{E}\left[\sum_{t=0}^{\infty} \delta^t U_{i,t} \mid \mu_{i,0}, \Sigma_{i,0}, \pi\right], \tag{4}$$

where δ is the discount factor and the expectation is taken over the randomness in actions $A_{i,t}$ and utilities $U_{i,t}$. Typical approaches to find the optimal sequence of choices by consumers 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].

We draw inspiration from this stream of literature and assume that consumers 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].[3]

Thompson Sampling aims to find the right balance between exploration and exploitation in the decision-making process. The algorithm starts by initializing the consumer's prior belief

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

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 for all possible actions by plugging in $\tilde{\theta}_{i,0}$ for $\theta_{i,0}$ in Equation (1). In the next step, the algorithm chooses the action 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 Consumer

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**
  $\tilde{\theta}_{i,t} \sim N(\mu_{i,t}, \Sigma_{i,t})$ ▷ Distribution Sampling
  $A_{i,t} \in \text{argmax}_{A \in \mathcal{A}} \tilde{\theta}_{i,t}^T A$ ▷ Action Selection
  $\Sigma_{i,t+1} \leftarrow \left(\Sigma_{i,t}^{-1} + \frac{1}{\sigma_\epsilon^2} A_{i,t} A_{i,t}^T\right)^{-1}$ ▷ Belief Updating (Same as Algorithm 1)
  $\mu_{i,t+1} \leftarrow \Sigma_{i,t+1}\left(\Sigma_{i,t}^{-1}\mu_{i,t} + \frac{1}{\sigma_\epsilon^2} A_{i,t} U_{i,t}\right)$ ▷ Belief Updating (Same as Algorithm 1)
end for

Assuming that the self-exploring consumer uses Thompson Sampling has several key advantages. First, it is a commonly used heuristic strategy for this dynamic problem, and early literature has shown it to be nearly optimal [Chapelle and Li, 2011]. Second, it is computationally light, making it advantageous in our later empirical analysis using the MovieLens data set.[4] 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 consumer choice in the presence of the recommendation system. To do so, we first introduce a personalized recommendation system that aims to simplify the consumer's decision-making problem. Since we want to quantify the impact of personalized recommendation systems on consumer learning, we assume that the recommendation system's objective is the same as that of consumer.[5] A natural difference between the personalized

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

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

recommendation system and a single consumer 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 consumers, 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 actions ($J = |\mathcal{A}|$) where each column is represent the vector of attributes for an action. We can define the matrix analog of the consumer utility in Equation 1 as follows:

$$U = \Theta^T A + E, \tag{5}$$

where $U_{N\times J}$ represent the utility for each pair of consumer and action 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 consumer reveals utility for a subset of items. The main question is how the data from other consumers $U^{\text{obs}}_{N\times J}$ help the personalized algorithm learn about the new consumer i with vector of preferences θ_i.

In principle, if the prior data from consumers do not inform us about the new consumer, the recommendation system's strategy will be the same as the self-exploring consumer. However, the prior empirical work on personalized recommendation systems suggests otherwise as it documents extensive similarities in consumer preferences [Koren et al., 2021, Rafieian and Yoganarasimhan, 2023]. The most common approach to characterize the similarities in consumer preferences is to use a factor model. That is, there is set of r independent d-dimensional factors $\{F_k\}_{k=1}^r$, where each factor $F_k \in \mathbb{R}^d$ presents a common type of preference for action attributes and each consumer's preferences is a linear combination of these r factors. We formally present this assumption as follows:

Assumption 1. *The matrix of consumers' preference weights $\Theta_{[d\times N]}$ can be decomposed as follows:*

$$\Theta_{[d\times N]} = F_{[d\times r]}\Gamma_{[r\times N]}, \tag{6}$$

where $F_{[d\times r]} = [F_1 \mid F_2 \mid \ldots \mid F_r]$ is the matrix containing all r factors, and $\Gamma_{[r\times N]} = [\gamma_1 \mid \gamma_2 \mid \ldots \mid \gamma_N]$ presents the 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 consumers' prior data, they have access to an accurate estimate of factor matrix F. As such, the recommendation system's task of learning

consumer i's preference parameters will turn into the task of learning consumer i's weights for r factors because we have: $\theta_i = F\gamma_i$. Hence, the recommendation system's data advantage translates into learning only r parameters, as compared to self-exploring consumer's learning of d parameters.

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

The recommendation system 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 item with the highest expected utility (lines 2 and 3). The RS-dependent consumer always follows the recommended action (line 4). Once the utility is realized, both the consumer and recommendation system update parameters of their posterior distribution $\mu_{i,t+1}^{(\theta)}$, $\Sigma_{i,t+1}^{(\theta)}$, $\mu_{i,t+1}^{(\gamma)}$, and $\Sigma_{i,t+1}^{(\gamma)}$. The algorithm repeats this process for $\mathcal{T}$ periods.

A few points are worth noting about the RS-dependent consumer. First, we assume that the consumer follows the recommended item every time period. It is easy to rationalize this choice in a variety of ways due to the search cost associated with exploration and the fact that the recommendation system learns faster than the consumer because of its information advantage. In the main analysis, we abstract away from this possibility and simply assume that the consumer always follow the recommendation system to illustrate the differences between the self-exploring and RS-dependent consumer.[6] Second, the consumer 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 consumer cannot learn from the recommendations beyond their own experience. This assumption is reasonable as recommendation systems

[6]If the RS-dependent consumer 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. In §4.5, we extend our analysis to include the rational RS-dependent consumer with a search cost.

Algorithm 3 Choice and Learning for the RS-Dependent Consumer

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

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

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

$\tilde{\gamma}_{i,t} \sim N(\mu_{i,t}^{(\gamma)}, \Sigma_{i,t}^{(\gamma)})$ ▷ RS: Distribution Sampling

$A_{i,t}^{RS} \in \text{argmax}_{A \in \mathcal{A}} \tilde{\gamma}_{i,t}^T F^T A$ ▷ RS: Recommendation Selection

$A_{i,t} \leftarrow A_{i,t}^{RS}$ ▷ Consumer: Action Selection

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

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

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

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

end for

are often very complex and it is not realistic to assume that consumers can learn further by observing that a product is recommended.

3.4 Main Outcomes

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

$$CW_i(T; \pi) = \sum_{t=0}^{T} u_i(A_{i,t}), \tag{7}$$

where CW_i is a consumer-specific function that depends on consumer 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 *regret*, which takes the difference between the utility from first-best in each period and consumer welfare. We formally define *regret* and *expected regret* as follows:

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

is given as follows:

$$R_i(T;\pi) = \sum_{t=0}^{T} \left(u_i(A_i^*) - u_i(A_{i,t})\right), \tag{8}$$

which is equal to the cumulative difference between the utility from always choosing optimal action and the utility from the sequence of actions taken till the end of period T. Based on the notion of regret in Equation (8)*, we can define the* ***expected regret*** *as follows:*

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

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.

One advantage of using regret instead of welfare is that we can directly use other established theoretical regret bounds from the literature. We later use these bounds to conduct a theoretical analysis of our problem.

The second outcome we are interested in is consumer learning. Intuitively, the greater uncertainty the consumer has over her own preference parameters, the lower the degree of consumer learning. 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 given the prior sequence $\mathcal{H}_{i,t-1}$. We measure consumer 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 consumer is more uncertain as to what the optimal action is. For example, if the consumer deterministically chooses one action (maximum certainty and learning), Shannon entropy of the optimal action will be equal to zero, i.e., $H(A_{i,t}) = 0$. On the other hand, when the consumer is maximally uncertain between actions, each action has 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 consumer welfare. To overcome

this challenge, we develop the measure of *counterfactual regret*, which is the regret that would be incurred on the consumer at any time period, if the consumer makes the decision independently, absent the influence of the personalized algorithm. As such, if lower regret actions for a consumer come at the expense of consumer learning, we expect the counterfactual regret to be high for this consumer. We define *counterfactual regret* as follows:

Definition 4. *Let* $\mathrm{CA}^*_{i,t}$ *denote the counterfactual action, which is the random variable corresponding to the optimal action given the prior sequence* $\mathcal{H}_{i,t-1}$*. This is the optimal action the consumer would choose based on her past learning through experience, if the personalized algorithm was not available at period* t *only. For the sequence of actions* $\{A_{i,t}\}_{t=0}^{T}$ *chosen according to policy* π*, the* ***counterfactual regret*** *is given as follows:*

$$\mathrm{CR}_i(T;\pi) = \sum_{t=0}^{T} \left(u_i(A_i^*) - u_i(\mathrm{CA}_{i,t}^*)\right), \tag{11}$$

In this calculation, the counterfactual action $\mathrm{CA}^*_{i,t}$ *can be different from the chosen action* $A_{i,t}$ *under policy* π*. Based on the notion of counterfactual regret in Equation* (11)*, we can define the* ***expected counterfactual regret*** *as follows:*

$$\mathbb{E}\left[\mathrm{CR}_i(T;\pi)\right] = \mathbb{E}\left[\sum_{t=0}^{T} \left(u_i(A_i^*) - u_i(\mathrm{CA}_{i,t}^*)\right)\right], \tag{12}$$

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

The main benefit from using the notion of *counterfactual regret* is its direct comparability with the actual *regret*. In cases where consumers follow the personalized algorithms to choose the action, we expect the counterfactual action they would choose without the personalized algorithm to be different from the one offered by the algorithm. As such, there will be discrepancy between the *counterfactual regret* and the actual *regret*. The gap between the two highlights the potential loss due to insufficient learning.

3.5 Theoretical Analysis

We now conduct a welfare analysis of the two types of consumer we defined earlier: (1) self-exploring consumers who follow Algorithm 2, and (2) RS-dependent consumers who follow Algorithm 3. One advantage of using regret is that we can directly use other established regret bounds from the literature. The paper we heavily borrow from for our regret analysis

in this section is Russo and Van Roy [2016] that find the information-theoretic regret bounds for the Thompson Sampling algorithm that depends on the notion of Shannon entropy of the prior distribution of optimal action. We start with the self-exploring consumer and establish the following regret bounds:

Proposition 1. *Let π_{SE} denote the policy for the self-exploring consumer who follows the Thompson Sampling algorithm for choice and learning as in Algorithm 2. The regret bound for this consumer is as follows:*

$$\mathbb{E}\left[R_i(T;\pi_{\text{SE}})\right] \leq \sqrt{\frac{H(A^*_{i,0})dT}{2}}, \tag{13}$$

where $H(A^*_{i,0})$ is the Shannon entropy of the prior distribution of optimal action for self-exploring consumer defined at period 0 and d is the dimensionality of the action space.

Since the expected regret for the self-exploring consumer is the same as the expected regret for Thompson Sampling, the regret bound in Proposition 1 is identical to the one established in Russo and Van Roy [2016]. According to this result, a consumer with better prior knowledge about her preferences has a lower Shannon entropy of the prior distribution of optimal action, and therefore, a lower upper bound for expected regret. The other component in the expected regret for the self-exploring consumer is the dimensionality of the action space d, which implies that a higher dimensional preference distribution is harder to learn and induces higher expected regret.

The regret analysis for RS-dependent consumers is a bit more subtle. In this case, the consumer follows the recommendation from personalized recommendation system. As such, we need to find the regret bound for the personalized algorithm. An important difference is that the RS-dependent consumer operates in an r-dimensional environment as it has access to the factor information. On the other hand, it is reasonable to assume that the algorithm knows less about consumer preferences in the beginning. In the extreme case, we can 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 $\log(|\mathcal{A}|)$. The following proposition characterizes the following regret bound for the RS-dependent consumer:

Proposition 2. *Let π_{RS} denote the policy for the RS-dependent consumer who consistently follows the personalized recommendations. Further, let $H(A^{\text{RS}}_{i,0})$ denote the Shannon entropy of the prior distribution of optimal action for the personalized recommendation system. The*

regret bound for the personalized recommendation system is as follows:

$$\mathbb{E}\left[R_i(T;\pi_{\mathrm{RS}})\right] \leq \sqrt{\frac{H(A_{i,0}^{\mathrm{RS}})rT}{2}}, \tag{14}$$

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

The result in Proposition 2 links the information advantage and disadvantage of the personalized algorithm to the regret bounds. On one hand, the platform's algorithm has access to data for many other consumers, which helps decrease the regret bound and reduce the dimensionality of the problem from d to r. On the other hand, it is conceivable that in some contexts, consumers know more about their own parameters than the algorithm and have a lower Shannon entropy of the prior distribution of optimal action, i.e., $H(A_{i,0}^{*}) < H(A_{i,0}^{RS})$. Our regret bounds capture this algorithmic disadvantage.

In our analysis, we primarily focus on cases with identical prior for several reasons. First, because the platform's and consumer's objectives are aligned, it is reasonable to assume that the platform can use ways to elicit information from consumers through surveys or other means. Second, in many important settings that are new to consumers, the consumer prior is largely uninformative. For example, consider the content in a new media format introduced by a social media platform (e.g., TikTok). In these cases, it is reasonable to assume that consumers' preferences are largely unknown in the beginning and consumers learn them through experience. Third, since we are interested in the consumer protection aspect of the problem, our focus on consumers with uninformative prior is equivalent to focusing on consumers more susceptible to adversarial AI and targeted manipulations. Finally, although our regret bounds capture the prior information structure, assuming an uninformative case is theoretically and empirically convenient, which helps us deliver the main insights.[7] Now, if $H(A_{i,0}^{*}) = H(A_{i,0}^{RS})$, we can write the following corollary based on Propositions 1 and 2:

Corollary 1. *If $H(A_{i,0}^{*}) = H(A_{i,0}^{\mathrm{RS}})$, the regret bound for the RS-dependent consumer is $\sqrt{r/d}$-fraction of the regret bound for the self-exploring consumer.*

This result highlights a key benefit of following personalized algorithms. In many settings, the ratio r/d is pretty small as evident by the performance of low-rank methods in personalized recommendation systems [Koren et al., 2021]. In a lower dimensional setting, the algorithm needs to learn fewer parameters, so it identifies good recommendations more quickly than

[7]We later conduct a series of analysis that work with different priors for different types of consumers.

the consumers themselves. Thus, the consumer incurs lower regret by consistently following the algorithm.

Our result in Corollary 1 provides a theoretical answer to our first research question about the value of personalized algorithms and how their information advantage translates into better recommendations. We now turn to the second research question on how following personalized recommendations influences consumer learning. In principle, if the personalized recommendations create enough variation in actions to learn preference parameters, we do not expect it to affect consumer learning. We know that the algorithm needs to explore in the beginning, thereby providing some variation. However, since the learning rate is faster in light of Corollary 1, the algorithm gets to the exploitation stage quickly and the amount of exploration may be insufficient for consumer learning. In particular, if the exploitation stage of the algorithm recommends a narrow set of actions, the variation may be limited for consumer learning. Importantly, whether or not the exploitation stage of the algorithm generates enough variation for consumer learning largely depends on the distribution of consumer preferences, and therefore, is an empirical question. In our empirical framework, we examine this question.

4 Empirical Framework

As discussed in our theoretical analysis, the degree to which a personalized recommendation system can affect consumer learning and downstream welfare outcomes depends heavily on the distribution of consumer preferences. As such, we need an empirical framework to examine outcomes for different types of consumers. To do so, we need an empirical setting where (1) we observe real consumer preferences, (2) we have information on the item characteristics so we can simulate consumers' preference learning over the space of item characteristics, 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.[8] 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 item characteristics. Lastly, the MovieLens data set has been widely used as the benchmark dataset for research on personalized recommendation system, thereby ensuring that we can

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

obtain state-of-the-art personalized recommendation systems.

4.1 Problem Definition

The main purpose of our empirical framework is to complement our theoretical analysis by allowing us to evaluate welfare and learning outcomes using actual consumer preferences. As such, we can define our main problem as recovering consumer preferences from the observed data. To formally characterize our goal, we need to state a few assumptions on our empirical setting. The first assumption we make is about the relationship between ratings and consumer utility. Whereas our theoretical framework uses the concept of consumer utility, we only observe consumer ratings. Let $Y_{[N\times J]}$ denote the matrix of ratings for N consumers and J movies. We present our assumption that links ratings to utility as follows:

Assumption 2. *Consumer i's utility from watching movie j is well-approximated by consumer ratings, i.e., $U_{[N\times J]} \approx Y_{[N\times J]}$.*

Our next assumption is about the dimensionality of consumer preferences. Since we observe detailed movie tags, we use them as the main characteristics of the movie. These tags include attributes such as originality, great ending, and good soundtrack (please see Appendix B for the list of all movie tags). Normally, one could all tags to characterize the dimensionality of consumer preferences. However, it is reasonable to assume that only a subset of these tags are important for consumer utility. We assume that the top d most frequently used tags characterize the dimensionality of consumer preferences. We set $d = 100$ and present our formal assumption as follows:

Assumption 3. *The matrix of movie characteristics is defined as $A_{[100\times J]} = [A_1 \mid A_2 \mid \cdots \mid A_J]$, where A_j represents features of movie j in terms of top* 100 *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 consumer preferences. As before, we let $\Theta_{[d\times N]}$ denote the consumer preferences for all N consumers. Our assumption about stable preferences is presented as follows:

Assumption 4. *For any consumer i, the true consumer vector of preferences is represented as θ_i for all time periods, independent of the movies watched.*

It is worth noting that the consumer in our setting still learns these preferences through experience. That is, although the actual preference is stable, the consumer's belief about

it can change over time. This assumption is consistent with the economics and marketing literature on Bayesian learning [Ching et al., 2013]. This assumption is particularly suitable in our setting because 63% of consumer rated all movies on the same day. With the three assumptions presented above, we can now formally define our problem:

Problem 1. *Let $Y^{\text{obs}}_{[N\times J]}$ denote the movie ratings observed in our data. Given the matrix of movie features $A_{[d\times J]}$, our goal is to estimate $\Theta_{[d\times N]}$ such that $U_{[N\times J]} = \Theta^T_{[d\times N]} A_{[d\times J]}$.*

In next sections, we discuss our empirical strategy to address this problem and then use it for implementation and evaluation of policies we want to study in our paper.

4.2 Empirical Strategy

In this section, we present our solution to the problem presented in Problem 1. If the matrix of ratings $Y^{obs}_{[N\times J]}$ is complete such that all entries are observed, the solution to our problem becomes simple as we have both $U_{[N\times J]}$ ($\approx Y^{obs}_{[N\times J]}$) and $A_{[d\times J]}$. However, the main challenge is that the observed matrix of ratings $Y^{obs}_{[N\times J]}$ is generally sparse. That is, each consumer has only rated a subset of movies. As a result, for many entries in the matrix, we do not observe any rating in the data.

For our main empirical strategy, we use an intuitive impute-then-project solution, whereby we first impute the missing entries of the incomplete matrix, and then use that to identify the consumer preference matrix $\Theta_{[d\times N]}$. More specifically, we can present our two-step solution as follows:

- *Imputation Step:* We use state-of-the-art matrix completion techniques to impute the missing entries and complete the matrix. Let $\tilde{U}_{[N\times J]}$ denote the completed matrix of ratings. At a high-level, these methods use a low-rank decomposition such that $\tilde{U}_{[N\times J]} = B_{[N\times r']} C_{[r'\times J]}$, where $r' \ll \min\{N, J\}$.[9] We follow the package used in Cortes [2018] to ensure the state-of-the-art performance of our imputation step.
- *Projection Step:* Once we have the imputed matrix $\tilde{U}_{[N\times J]}$, we can use the pseudo-inverse of $A_{[d\times J]}$ to estimate for $\Theta_{[d\times N]}$. Let $A^+_{[J\times d]}$ denote the pseudo-inverse of matrix $A_{[d\times J]}$. We can then estimate consumer preference matrix as follows:

$$\widehat{\Theta}_{[d\times N]} = \left(\tilde{U}_{[N\times J]} A^+_{[J\times d]}\right)^T \tag{15}$$

It is worth noting that for the projection step, one could use each row i in $\tilde{U}_{[N\times J]}$ as the

[9] We use r' to denote the low-rank decomposition to not confuse it with the rank r that the recommendation system uses to identify the factors.

dependent variable and regress it on the movie features for all J movies and estimate $\widehat{\theta}_i$ for each consumer i separately. Our approach is advantageous as it yields better fit with the observed and imputed outcomes. However we use the regression-based approach to ensure the robustness of our main findings.

Another alternative solution to Problem 1 is to only focus on the non-sparse parts of the matrix as in Bresler et al. [2014]. This approach has the benefit of only using observed outcomes, not the imputed ones. However, the main issue is that we need to drop many consumers as only a very small fraction of consumers rated a large number of movies. As such, we use the impute-then-project approach as our main empirical strategy in the paper, but use the approach without imputation as a robustness check.

4.3 Policy Simulation and Evaluation

We now present the details on policy simulation and evaluation. For each consumer i, we want to simulate the cold-start problem under two policies where the consumer is (1) self-exploring or (2) RS-dependent. Since we are interested in the cold-start setting, it is essential that neither policies use any information on consumer i. For example, if the factors used by the personalized algorithm are obtained by using the data of consumer i, the recommendations capture this consumer in a way that would not have been possible for the algorithm to generate. To address this issue, we use a train-test split procedure, with 5040 consumers in the training set and 1000 consumers in the test set. We use the training set to obtain the factors used by the personalized recommendation system. We simulate the policies and evaluate the outcomes only for the 1000 consumers in our test set. This train-test split ensures that the personalized recommendation system does not use consumer i's data to obtain the latent factors.

4.3.1 Policy Simulation

We now discuss policy simulation for each consumer in our test set under both self-exploring and RS-dependent scenarios. To do so, we closely follow Algorithms 2 and 3. Both algorithms require the consumer to know σ_ϵ, which is the standard deviation of the error term. In our empirical setting, we use the Root Mean Squared Error (RMSE) of our matrix completion algorithm on the training data set the known standard deviation and find $\widehat{\sigma}_\epsilon = 0.86$. As such, when generating the utility outcome for consumer i from watching movie j, we have:

$$\widehat{u}_i(A_j) = \widehat{\theta}_i^T A_j + \widehat{\epsilon}_i, \tag{16}$$

where $\widehat{\epsilon}_i \sim N(0, \widehat{\sigma}_\epsilon)$. The standard deviation $\widehat{\sigma}_\epsilon$ also appears in the equation for consumers' belief updating after realizing the utility of each movie.

To realistically simulate the process, we define the action set at period t as $\mathcal{A}_{i,t}$, which is a random subset of all actions. We set the size of this random subset at 100. This is because in most content consumption settings, the choice set changes over time, but the distribution of action attributes remain relatively fixed.[10] We now describe the policy simulation for two different types of consumers:

- *Self-Exploring Consumer:* For the self-exploring consumer, the process is exactly as described in Algorithm 2 with a minor difference that the action set updates in every round. For any consumer i with preference parameters $\widehat{\theta}_i$, we set fully uninformative priors, random subset of size 100 from the full action set at any time period $\mathcal{A}_{i,t}$, and $T = 1000$ as inputs. The output will be the following set:

$$\mathcal{D}_i^{SE} = \{\widehat{A}_{i,t}^*, \widehat{CA}_{i,t}^{SE}, \widehat{A}_{i,t}^{SE}, \widehat{U}_{i,t}^{SE}, \mu_{i,t}^{SE}, \Sigma_{i,t}^{SE}\}_{i=0}^{T}, \tag{17}$$

  where $\widehat{A}_{i,t}^*$ is the optimal action that should be chosen at each period given the true parameters $\widehat{\theta}_i$, $\widehat{A}_{i,t}^{SE}$ is the action chosen by the self-exploring consumer, and $\widehat{CA}_{i,t}^{SE}$ is the counterfactual action that would be chosen if the consumer follows the Thompson Sampling algorithm, and $\widehat{U}_{i,t}^{SE}$ is the estimated utility received by choosing the action $\widehat{A}_{i,t}^{RS}$ according to Equation (16). The prior distribution for the preference parameters is set as an uninformative prior with mean zero and standard deviation one that comes from a Normal distribution, and is updated in each period according to Algorithm 1. Since self-exploring consumer follows the Thompson Sampling algorithm throughout, we have $\widehat{CA}_{i,t}^{SE} = \widehat{A}_{i,t}^{SE}$. The reason we include $\widehat{CA}_{i,t}^{SE}$ as one of the outputs is for the calculation of *counterfactual regret*. For self-exploring consumers, we know that the expected regret and the expected counterfactual regret are always the same. However, collecting $\widehat{CA}_{i,t}^{SE}$ is useful when we calculate the counterfactual regret for RS-dependent consumers.
- *RS-Dependent Consumer:* For the RS-dependent consumer, we closely follow Algorithm 3. As illustrated in that algorithm, we need to specify the set of factors the personalized algorithm uses. To obtain these factors, we use our training data to first estimate $\widehat{\Theta}_{[d\times N_{tr}]}^{tr}$, and then decompose it into $\widehat{F}_{[d\times r]}^{tr}\widehat{\Gamma}_{r\times N_{tr}}$ using a low-rank decomposition. We use cross-

[10] Please note that this random sampling from the choice set induces some organic exploration for the RS-dependent consumer. As such, our simulation induces more consumer learning compared to the linear bandits problem with a fixed action set as in Russo and Van Roy [2014]. In that sense, our analysis is conservative in identifying the effect of personalized algorithms on consumer learning.

validation to find the optimal rank. We then give $\widehat{F}^{tr}_{[d\times r]}$ as the input to Algorithm 3. Like the self-exploring consumer, we start with fully uninformative priors for both the consumer and the personalized algorithm. We use the same $\mathcal{A}_{i,t}$ as the self-exploring consumer in each time period and set $T = 1000$. The output will be the following set:

$$\mathcal{D}_i^{RS} = \{\widehat{A}^*_{i,t}, \widehat{CA}^{RS}_{i,t}, \widehat{A}^{RS}_{i,t}, \widehat{U}^{RS}_{i,t}, \mu^{RS}_{i,t}, \Sigma^{RS}_{i,t}\}_{i=0}^{T}, \tag{18}$$

where $\widehat{A}^*_{i,t}$ is the optimal action that should be chosen at each period given the true parameters $\widehat{\theta}_i$ (same as the optimal action for the self-exploring consumer), $\widehat{A}^{RS}_{i,t}$ is the action recommended by the algorithm and chosen by the RS-dependent consumer, and $\widehat{CA}^{RS}_{i,t}$ is the counterfactual action that would be chosen if the consumer follows the Thompson Sampling algorithm at each time period, and $\widehat{U}^{RS}_{i,t}$ is the estimated utility received by choosing the action $\widehat{A}^{RS}_{i,t}$ according to Equation (16). The prior distribution for the preference parameters is set as an uninformative prior with mean zero and standard deviation one that comes from a Normal distribution, and is updated in each period according to Algorithm 1. Unlike the self-exploring consumer, the counterfactual action is different from the action chosen by the RS-dependent consumer, i.e., $\widehat{CA}^{RS}_{i,t} \neq \widehat{A}^{RS}_{i,t}$. In particular, if following personalized recommendations limits consumer learning, we expect the RS-dependent consumer to make poor decisions on their own and incur high counterfactual regret.

The procedure above details the step-by-step simulation of the events. For policy evaluation, we need to aggregate over all the these events and measure outcomes, such as expected regret and expected counterfactual regret.

4.3.2 Evaluation of Main Outcomes

As discussed earlier in §3.4, we are interested in six different outcomes: consumer welfare, regret, expected regret, Shannon entropy of optimal action, counterfactual regret, and expected counterfactual regret. In this section, we present how we evaluate these outcomes for both self-exploring and RS-dependent consumers. For the set of simulated outputs $\mathcal{D}_i^j$ corresponding to consumer i with type $j \in \{SE, RS\}$, we can calculate consumer welfare as follows:

$$\widehat{CW}_i^j = \sum_{t=0}^{T} \widehat{U}^j_{i,t}, \tag{19}$$

where $\widehat{U}^j_{i,t}$ is consumer i's utility from action chosen at period t, when consumer type is $j \in \{SE, RS\}$. Similarly, we can define the regret as the difference between the first-best

action given the true parameters and the action chosen as follows:

$$\widehat{R}_i^j = \sum_{t=0}^{T} \left(\widehat{u}_i \left(\widehat{A}_{i,t}^* \right) - \widehat{U}_{i,t}^j \right), \tag{20}$$

where $A_{i,t}^* = \text{argmax}_{A \in \mathcal{A}_{i,t}} \widehat{\theta}_i^T A$, which is the first-best action from the action set $\mathcal{A}_{i,t}$ in period t.[11] To measure the expected regret, we drop the error term as follows:

$$\widehat{ER}_i^j = \sum_{t=0}^{T} \widehat{\theta}_i^T \left(\widehat{A}_{i,t}^* - \widehat{A}_{i,t}^j \right). \tag{21}$$

We now discuss how we evaluate consumer learning through the Shannon entropy measure defined in 3. The calculation is based on the posterior distribution of preference parameters, which is characterized by $\mu_{i,t}^j$ and $\Sigma_{i,t}^j$. Let $H_{i,t}^j$ denote Shannon entropy of optimal action for consumer i of type j at time t. This value depends on the set of actions available $\mathcal{A}_{i,t}$ as follows:

$$\widehat{H}_{i,t}^j = - \sum_{A \in \mathcal{A}_{i,t}} P \left(\widehat{A}_{i,t}^* = A \mid \mu_{i,t}^j, \Sigma_{i,t}^j \right) \log_2 \left(P \left(\widehat{A}_{i,t}^* = A \mid \mu_{i,t}^j, \Sigma_{i,t}^j \right) \right), \tag{22}$$

where $P \left(\widehat{A}_{i,t}^* = A \mid \mu_{i,t}^j, \Sigma_{i,t}^j \right)$ calculates the probability of each action $A \in \mathcal{A}_{i,t}$ being the optimal action given the posterior distribution of preference parameters. If the consumer is more uncertain about her preferences, the distribution of optimal action given preferences is more evenly distributed, which corresponds to a higher Shannon entropy.

Next, we focus on our measures of counterfactual regret that allow us to compare consumer learning with our regret measures more easily. The definition is the same as the regret with the difference that we replace the chosen action in data $\mathcal{D}_i^j$ with the action that would be chosen if the consumer makes the decision independently. For the self-exploring consumer who makes decisions on her own, counterfactual regret is equal to regret. However, for the RS-dependent consumer, there can be a difference between these two values as the consumer chooses actions that may differ from her independent choice. We can define the counterfactual regret as follows:

$$\widehat{CR}_i^{RS} = \sum_{t=0}^{T} \left(\widehat{u}_i \left(\widehat{A}_{i,t}^* \right) - \widehat{u}_i \left(\widehat{CA}_{i,t}^j \right) \right). \tag{23}$$

[11] It is worth noting that the first-best action in each period is not different for self-exploring and RS-dependent consumers because we fix the action set.

We can now define the expected counterfactual regret by taking out the stochastic error component as follows:

$$\widehat{ECR}_i^j = \sum_{t=0}^{T} \widehat{\theta}_i^T \left(\widehat{A}_{i,t}^* - \widehat{CA}_{i,t}^j \right). \tag{24}$$

It is worth noting that the equations above calculate our outcomes of interest only for one instance of data. To account for the randomness in the process, we can easily run the simulation B times and take the average of the above values.

4.4 Main Results

In this section, we present our main results. In particular, we present answers to our main two research questions about the impact of the personalized algorithm on consumer welfare and consumer learning. In §4.4.1, we present our results about the impact of the personalized learning on the expected regret. Next, we examine whether following personalized recommendations leads to lower learning using our learning measures in §4.4.2.

4.4.1 Impact of Recommendation Systems on Consumer Welfare

We start with examining the impact of the personalized algorithm on consumer welfare. In §3.5, we theoretically show that the upper bound for expected regret is lower for RS-dependent consumers compared to self-exploring consumers. For each consumer i in our test data set ($N = 1000$), we calculated per-period expected regret and aggregate the values over all consumers. We present the results in Figure 1. The x-axis shows represents the period number, and the y-axis represents the average expected regret for all consumers in the test dataset. We focus on two different consumer types: self-exploring and RS-dependent.

A few important insights emerge from Figure 1. First, we observe that the regret experienced by both self-exploring consumers and RS-dependent consumers decreases over time. This suggests that both groups improve their decision-making as time progresses: self-exploring consumers do so by learning their preferences, while RS-dependent consumers benefit from increasingly accurate recommendations provided by the personalized algorithm. Second, we note that the expected regret in the first period is the same for both self-exploring and RS-dependent consumers, which is because both the self-exploring consumer and the personalized algorithm start with the same uninformative prior. However, the expected regret for RS-dependent consumers drops more quickly than self-exploring consumers. This is due to the fact that algorithm operates in a lower-rank environment and converges to good recommendations faster, as speculated in Corollary 1. Finally, we note that across all 1000 periods, the RS-dependent consumer experiences 70% (0.28/0.40) of the expected

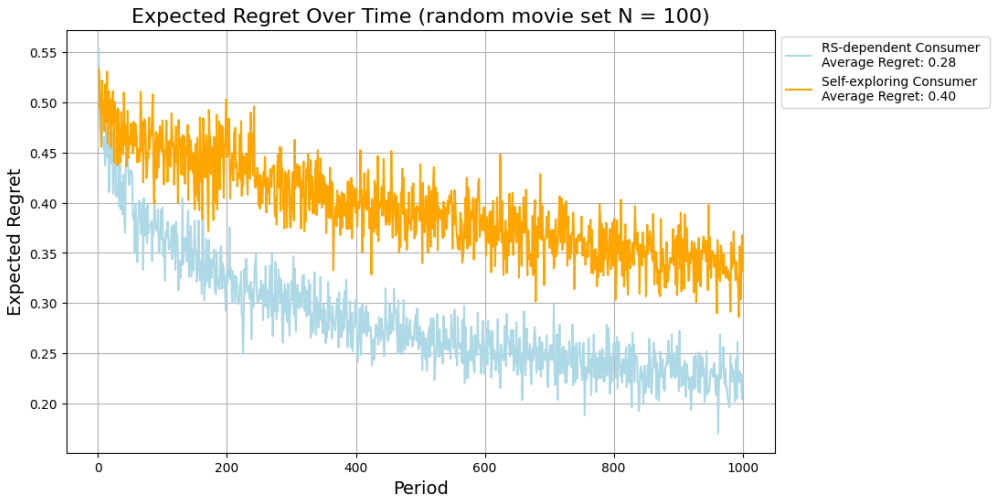

Figure 1: Expected Regret Over Time for Different Types of Consumers

regret compared to the self-exploring consumer. This 30% drop in the total expected regret highlights the value that personalized algorithms can create by helping consumers make better decisions.

4.4.2 Impact of Recommendation Systems on Consumer Learning

We now turn to our second research questions: how following the personalized algorithm's recommendations affects consumer learning. We start with our main learning measure: Shannon entropy of the optimal action. As shown in Equation (22), we can measure the per-period uncertainty the consumer has in decision-making given data $\mathcal{D}_i^{SE}$ and $\mathcal{D}_i^{RS}$. We aggregate this measure for all 1000 consumers in our test data and present the results in Figure 2. The x-axis in this figure is time period and the y-axis is the average Shannon entropy at each point for all consumers. At the first period, every action has the same probability of being the optimal action, which achieves a $\log_2(100) = 6.64$ Shannon entropy.

A few interesting patterns emerge from Figure 2. First, as illustrated in this figure, the entropy for both self-exploring and RS-dependent consumers decreases over time, reflecting learning and increased certainty about preferences. As consumers become more confident in their movie preferences, the uncertainty surrounding the optimal choice diminishes. Second, while the entropy levels for the two consumer types are relatively similar in the early periods, a divergence emerges over time. Specifically, the entropy for self-exploring consumers declines more rapidly, indicating a faster learning rate compared to RS-dependent consumers. Notably, the Shannon entropy for RS-dependent consumers stabilizes at a higher level than that of self-exploring consumers. This highlights a key finding: the incomplete learning caused

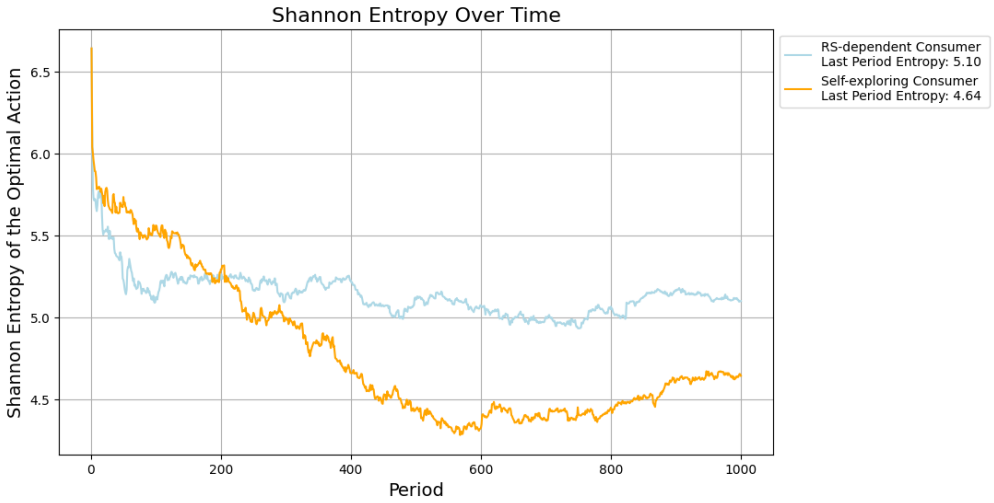

Figure 2: Shannon Entropy of Optimal Action Over Time for Different Types of Consumers

by the personalized algorithm persists over time. By the final period, the self-exploring consumer's entropy is approximately 90% (4.64/5.10) of the RS-dependent consumer's entropy, underscoring the long-term implications of the slower learning process facilitated by reliance on recommendations.[12]

The finding in Figure 2 suggests that following the personalized recommendations comes at the expense of consumer learning. However, as discussed earlier, it is hard to quantify the loss due to insufficient learning in terms of welfare. Our counterfactual regret measures helps overcome this challenge. This measure illustrates how the personalized algorithm affect consumers' independent decision-making ability. Following the procedure in Equation (24), we measure the expected counterfactual regret for each consumer in our test data and aggregate the values at each time period over all consumers. Figure 3 shows the result from this practice. As shown in this figure, the expected regret lines for self-exploring and RS-dependent consumers are identical to those in Figure 1. However, there is a new line that illustrates the expected counterfactual regret for the RS-dependent consumer, 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

[12]In addition to entropy measurement, we also use another measure for learning: the KL-divergence from the uninformative prior. The KL-divergence measures the difference between two distributions. Since both self-exploring and RS-dependent consumers start with the same uninformative prior, we can use the KL-divergence to measure how much the posterior belief has been updated (changes from the prior). We find that the self-exploring consumer's posterior distribution on the preference parameter has a larger KL-divergence from the prior compared with that of the RS-dependent consumer, implying higher learning. Details are in Appendix C.

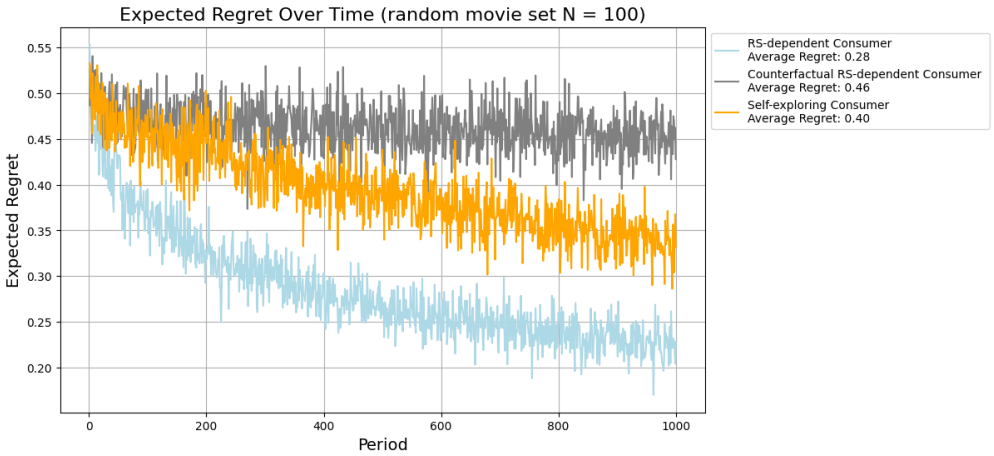

Figure 3: Expected Regret and Counterfactual Regret Over Time for Different Types of Consumers

regret for RS-dependent consumers. Importantly, the expected counterfactual regret for RS-dependent consumer is higher than the expected regret for self-exploring consumer. This finding suggests that although following personalized recommendations reduce expected regret, it comes at the expense of consumers' independent decision-making ability, an essential ability for safeguarding against potential adversarial AI attacks.

4.5 Robustness Check and Extensions

We consider several robustness checks and extensions considering how our results would change if we change (1) the regret measure to consumer welfare, (2) the imputed rating approach to actual ratings, (3) the consumer's dimensionality of preference parameter d, (4) the consumer's action set size, (5) the self-exploring consumer's informativeness of prior beliefs relative to the personalized recommendation system, (6) the personalized algorithm's rank used for matrix completion r. Our qualitative results remain unchanged under all these robustness checks: the RS-dependent consumer has a lower expected regret, but a higher Shannon entropy of optimal action and expected counterfactual regret.

When we switch from the expected regret measurement to the expected utility measurement, the RS-dependent consumer has a higher expected utility when RS is available but lower counterfactual utility from independent decision-making, compared with the self-exploring consumer, as shown in Figure A3 in Appendix D.1. For our second robustness check, we follow Bresler et al. [2014] and only consider the top 200 consumers and the top 400 movies for the simulation, so we could use the actual rating data not imputed rating. The resulting consumer-movie-rating matrix has 70% nonzero entries. As shown in Figure A4 in Appendix

D.2, the qualitative results remain largely the same as our analysis in the main text.

In our third robustness check, we change the dimensionality of the item (movie) space from 100 to lower and higher numbers. Our results persist when the consumer's preference space dimension d equals $20, 40, \ldots, 150$. When the dimensionality of the movie features increases, learning becomes more complex for consumers, leading to more sub-optimal choices. At the same time, the personalized algorithm faces a more complicated problem even though it has information advantage. We re-run our analysis for different values of d and present the results in for the average per-period regret in Figure A5 in Appendix D.3. As shown in this figure, the average regret unexpectedly increases with the dimensionality of the problem for all decision sequences. However, as d increases, the advantage of the personalized algorithm becomes more apparent because it identifies the optimal rank based on the data of existing consumers, which is substantially lower than the dimensionality of the problem.

In the main simulation, we use an action set $\mathcal{A}$ of cardinality 100 that contains 100 random movies in each time period. We demonstrate the robustness of our results to the reduced size of the action space in each period. We present our results in Figure A6 in Appendix D.4. Our primary results hold when the consumer's action set size (how many movies to check each time) is between 5 and 100. When the action set is small, the RS performs more similarly to the self-exploring consumer. However, as the complexity of the decision-making problem grows, the gap between groups widens.

Next, we consider an extension where consumers have a more informed prior than the personalized recommendation system, because they have watched some movies before and have a greater degree of certainty about their own preferences. We measure the information advantage of the consumer as she has already watched some movies before the first period (she starts to watch movies from period $-N$). Our findings indicate that even if the self-exploring consumer has watched movies for 1,000 periods prior to the first period, their average regret remains higher than that of the RS-dependent consumer. This result underscores the substantial information advantage provided by the recommendation system, demonstrating that this advantage cannot be easily offset by the consumer's prior knowledge.

We argue that RS has the information advantage so that it can decompose the consumer's preference weights into a lower-dimensional representation. However, it is not always the case that a smaller r leads to lower regret. When r is very small, the RS-dependent consumer explores only minimally. Conversely, when r is very large, the RS-dependent consumer may explore excessively, resulting in high expected regret for RS-dependent consumers. We find that $r = 10$ performs best for RS-dependent consumers compared with $r = 4, 20$. We present

the results from this practice in Figure A8 in Appendix D.6. As shown in this analysis, the RS-dependent consumer receives the lowest expected regret with the optimal rank 10. When the rank is lower, the process is a bit sub-optimal because the algorithm does not fully capture the complexity of the decision-making process. In the contrary, when the rank is higher than the optimal rank, the algorithm explores more leading to more sub-optimal choices by the RS-dependent consumer. In both sub-optimal rank scenarios, however, the RS-dependent consumer achieves a substantially lower expected regret than the self-exploring consumer, which highlights the value of the information advantage the algorithm has. Interestingly, the RS-dependent consumer's expected counterfactual regret decreases when r is higher than the optimal rank. This is likely because the personalized recommendation system induces more exploration when r is larger, thereby enhancing RS-dependent consumer's independent decision-making ability.

5 Policy Implications

Our results in §4.4 show that personalized recommendation systems can act as a barrier to consumer learning. In particular, we demonstrate that consumers who rely heavily on the recommendation system become worse independent decision-maker. This has important consumer protection policy implications as a lower decision-making ability makes consumers more susceptible to manipulations by adversarial AI. These concerns motivate us to consider different consumer protection policies. A typical candidate for such policies is a complete ban on the use of personalized algorithms. However, we note that the personalized algorithm can create substantial value for consumers, even without considering how much it reduces search costs. Our analysis of self-exploring consumers reflects the potential outcomes under a strict privacy regulation that completely bans consumer tracking and personalization: although such a ban can benefit consumer learning, consumers would struggle to choose actions with low regret. Thus, we want to examine if there are alternative policies that can achieve sufficient learning without completely losing the benefits of a recommendation system.

5.1 Random Availability Policies

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 consumers need to make their own decisions in some time periods. This likely results in an increase in consumer learning without fully eliminating the benefits of the personalized recommendation system. 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 belief. 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 for the self-exploring consumer. Please see Appendix E for the detailed description of the random availability algorithm.

We simulate the expected regret and expected counterfactual regret outcomes under each random availability policy and present the results in Figure 4. This figure clearly illustrates the trade-off between regret and counterfactual regret, which has a close connection 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 both lower expected regret and lower 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 consumer. For instance, the policy with $p = 0.2$ achieves what can be considered the best of both worlds: it results in lower expected counterfactual regret compared to the self-exploring consumer, while maintaining a level of expected regret comparable to that of the RS-dependent consumer. These findings suggest that even relatively simple policies, such as random availability, can effectively balance consumer welfare and learning, offering a compelling middle ground between personalization and consumer protection.

While random availability policies discussed in this section are not currently part of public policy conversations, we argue that they should be considered. Existing public policy proposals aimed at protecting consumers from personalized algorithms often emphasize algorithmic auditing. However, auditing approaches are frequently infeasible and may yield inconclusive results. The primary advantage of random availability policies lies in their simplicity and enforceability. These policies do not require complex oversight mechanisms and can provide a practical and effective means of balancing consumer protection with the benefits of personalized systems. Including such policies in the public policy discourse could broaden the range of feasible strategies to address the challenges posed by personalized algorithms.

5.2 Self-regulated Policies

Another class of policies we consider is the self-regulated recommendation system, where the platform wants to incorporate consumer learning as part of the objective. As shown in

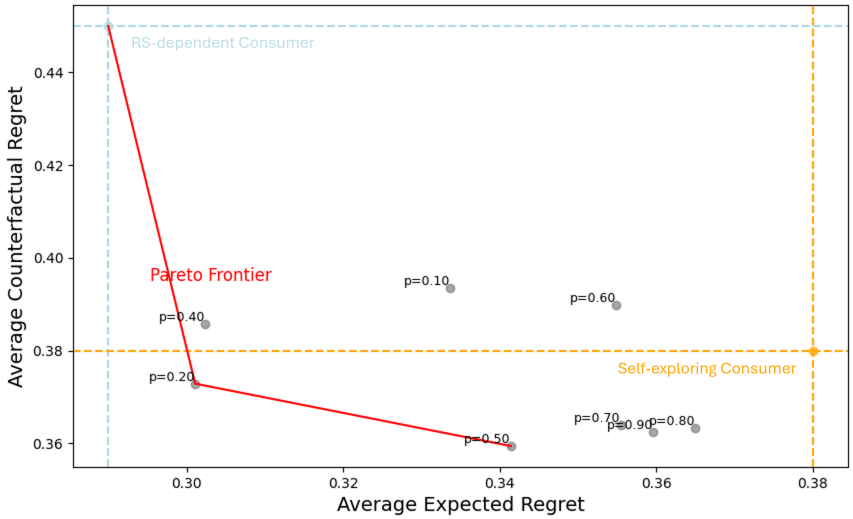

Figure 4: The performance of self-regulated policies in terms of regret and counterfactual regret.

Figure 2, RS-dependent consumer has insufficient learning compared with the self-exploring consumer. Hence, a reasonable policy is one that adds a regularization term on Shannon entropy. The idea is that different actions will decrease the Shannon entropy differently, so the recommendation system could recommend movies that would reduce the Shannon entropy more. In other words, these movies could reduce the uncertainty and aid consumer learning.

Similar to §4.4.2, in order to calculate consumer i's Shannon entropy at period t, we need to calculate the probability that movie j is the optimal movie given consumer i's belief $N(\mu_{i,t}, \Sigma_{i,t})$. Then, following the formula $H(A^*_{i,t}) = -\sum_{j \in \mathcal{A}} p_{i,j,t} \log_2(p_{i,j,t})$, we could calculate the change of Shannon entropy $\Delta H(A^*_{i,t,j}) = H(A^*_{i,t,j}) - H(A^*_{i,t-1})$ from choosing movie j at time t. Please see Appendix F for the detailed algorithm on how to calculate $\Delta H(A^*_{i,t,j})$. The regularization weight is represented by λ and λ is normalized from 0 to 1. A higher λ means a higher punishment for high entropy (high uncertainty), thereby resulting in more learning. Since the change of Shannon entropy $\Delta H(A^*_{i,j,t})$ is usually negative, by adding $-\lambda \times \Delta H(A^*_{i,j,t})$ in RS's selection function, we encourage the RS to pick the movies which could reduce the Shannon entropy more for consumers.

We simulate the expected regret and expected counterfactual regret based on different values of λ and present the results in Figure 5. Naturally, we expect higher values of λ to induce lower expected counterfactual regret, as consumers in these cases better learn their preferences. On the other extreme end, when $\lambda = 0$, our results basically simulate those

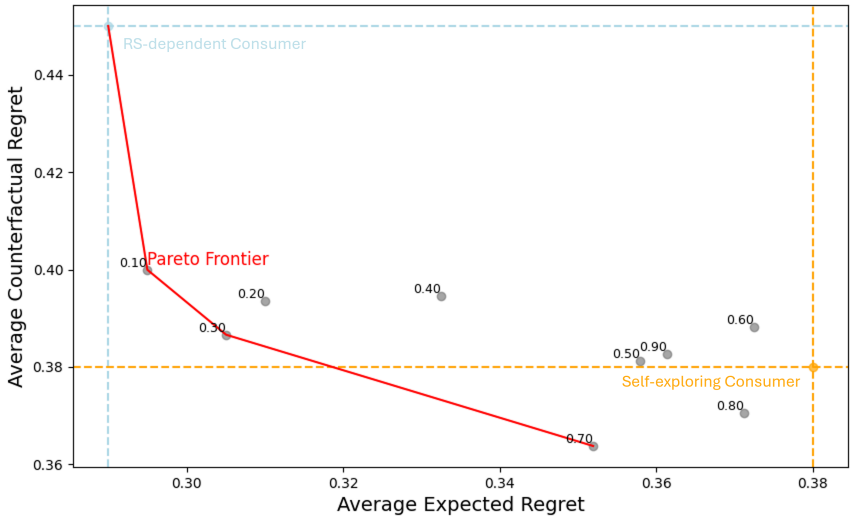

Figure 5: The performance of self-regulated policies in terms of regret and counterfactual regret (numbers in the graph represent values of λ).

for the RS-dependent consumer, because we use the exact same recommendation system as before, with zero weight for learning in the objective function. Similar to Figure 4, we find policies that Pareto dominate the self-exploring consumer. More importantly, Figure 5 shows that even small weights for learning in the objective ($\lambda = 0.10$) can substantially reduce the expected counterfactual regret, without sacrificing the expected regret. These results are promising, as they indicate that platforms can design algorithms that promote consumer learning without sacrificing the benefits of personalized recommendations.

6 Discussion and Conclusion

Consumers increasingly rely on personalized algorithms and recommendation systems to make decisions online. However, this increased reliance on algorithms has created concerns among consumer protection advocates and policymakers. Prior studies have focused on many downsides of personalized algorithms for consumer welfare in terms of privacy, fairness, and polarization. We examine personalized algorithms through the lens of consumers' learning and independent decision-making ability, an important construct given the growing use of advanced AI algorithms and the public's fear of adversarial AI. We build a theoretical framework that borrows from the literature on bandit algorithms and establishes regret bounds for consumer decision-making depending on their reliance on the personalized recommendation system. In our empirical framework, we find that although personalized algorithms increase

consumer welfare by quickly learning consumer preferences and offering good recommendations, following these algorithms can come at the expense of consumer learning and undermine their independent decision-making ability. Motivated by this finding, we run a series of policy simulations to examine whether a better policy design can mitigate these problems. Interestingly, we find that some very simple policies can substantially improve consumer learning without sacrificing the benefits of personalized recommendations.

Our paper makes several contributions to the literature. From a substantive point of view, we conduct a thorough investigation into how personalized recommendation systems influence consumer learning using a mix of theoretical and empirical analyses. We demonstrate that while personalized algorithms can enhance consumer welfare by improving product matches, they may also hinder consumer learning by restricting the exploration of personal preferences. Our work enriches the policy discourse on the societal implications of personalized recommendation systems by introducing a consumer learning perspective, which has been largely overlooked in the past literature studying the welfare implications of personalized recommendation systems. From a methodological standpoint, we build a theoretical framework that connects the literature on bandit algorithms to the consumer learning literature, which is of longstanding interest to marketing and economics scholars. A key innovation of our framework is in developing the counterfactual regret measure, which quantifies the potential welfare loss due to insufficient learning and can be applied to a variety of domains.

Our research holds significant implications for policymakers and platforms. From a policy standpoint, we emphasize the crucial role of consumer learning and the potential for personalized algorithms to undermine consumers' independent decision-making ability, thereby increasing their vulnerability to manipulations by adversarial players. Our findings suggest that exploration-based policies offer simpler alternatives to complex algorithmic auditing measures while still achieving favorable outcomes for both consumer welfare and learning. Furthermore, our research provides algorithmic solutions for platforms concerned about the negative consequences of their algorithms. We propose an algorithm that leverages the structure of consumer learning and incorporates it into the objective function. We stress that the adoption of these direct self-regulatory algorithmic solutions can lead to significant improvements in both consumer welfare and learning.

Nevertheless, our paper has certain limitations that open avenues for future research. First, our findings are largely based on the established theories on consumer learning, and we use empirically calibrated primitives to simulate outcomes. 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 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.

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.

E. Ascarza. Retention futility: Targeting high-risk customers might be ineffective. *Journal of Marketing Research*, 55(1):80–98, 2018.

S. Athey and G. Imbens. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360, 2016.

S. Barocas, M. Hardt, and A. Narayanan. *Fairness and Machine Learning*. fairmlbook.org, 2019. `http://www.fairmlbook.org`.

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

G. Bresler, G. H. Chen, and D. Shah. A latent source model for online collaborative filtering. *Advances in neural information processing systems*, 27, 2014.

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

A. T. Ching, T. Erdem, and M. P. Keane. Learning models: An assessment of progress, challenges, and new developments. *Marketing Science*, 32(6):913–938, 2013.

D. Cortes. Cold-start recommendations in collective matrix factorization. *arXiv preprint*

arXiv:1809.00366, 2018.

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

P. Dandekar, A. Goel, and D. T. Lee. Biased assimilation, homophily, and the dynamics of polarization. *Proceedings of the National Academy of Sciences*, 110(15):5791–5796, 2013.

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

J.-P. Dubé and S. Misra. Personalized pricing and consumer welfare. *Journal of Political Economy*, 131(1):131–189, 2023.

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.

S. Flaxman, S. Goel, and J. M. Rao. Filter bubbles, echo chambers, and online news consumption. *Public opinion quarterly*, 80(S1):298–320, 2016.

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.

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. Johnson. Economic research on privacy regulation: Lessons from the gdpr and beyond. 2022.

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`.

A. Lambrecht and C. Tucker. Algorithmic bias? an empirical study of apparent gender-based discrimination in the display of stem career ads. *Management science*, 65(7):2966–2981, 2019.

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.

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.

X. Nie and S. Wager. Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*, 108(2):299–319, 2021.

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.

U. Shalit, F. D. Johansson, and D. Sontag. Estimating individual treatment effect: generalization bounds and algorithms. In *International Conference on Machine Learning*, pages 3076–3085. PMLR, 2017.

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

D. Simester, A. Timoshenko, and S. I. Zoumpoulis. Efficiently evaluating targeting policies: Improving on champion vs. challenger experiments. *Management Science*, 66(8):3412–3424, 2020a.

D. Simester, A. Timoshenko, and S. I. Zoumpoulis. Targeting prospective customers: Robustness of machine-learning methods to typical data challenges. *Management Science*, 66

(6):2495–2522, 2020b.

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.

A. Swaminathan and T. Joachims. Counterfactual risk minimization: Learning from logged bandit feedback. In *International Conference on Machine Learning*, pages 814–823. PMLR, 2015.

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.

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.

S. Wager and S. Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 0(0):1–15, 2018. doi: 10.1080/01621459.2017.1319839.

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

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

H. Yoganarasimhan, E. Barzegary, and A. Pani. Design and evaluation of optimal free trials. *Management Science*, 2022.

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:

$$\theta_i \sim \mathcal{N}(\mu_{i,t}, \Sigma_{i,t}) \tag{25}$$

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:

$$\Sigma_{i,t+1}^{-1} = \Sigma_{i,t}^{-1} + \frac{1}{\sigma_\epsilon^2} A_{i,t} A_{i,t}^T \tag{26}$$

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 $\frac{1}{\sigma_\epsilon^2}$:

$$\mu_{i,t+1} = \Sigma_{i,t+1} \left(\Sigma_{i,t}^{-1} \mu_{i,t} + \frac{1}{\sigma_\epsilon^2} A_{i,t} U_{i,t} \right) \tag{27}$$

□

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	

C Results with KL Divergence as the Learning Measure

In addition to the Shannon entropy defined in Equation (22), we use the KL-divergence of the consumer's posterior distribution at any point t from the consumer's prior distribution of preference parameters to measure learning. The prior belief on the preference vector $\theta_0 \sim N(\mu_0, \Sigma_0)$. $\mu_0 = 0$ and Σ_0 is an identity matrix. Suppose the posterior belief $\tilde{\theta}$ at

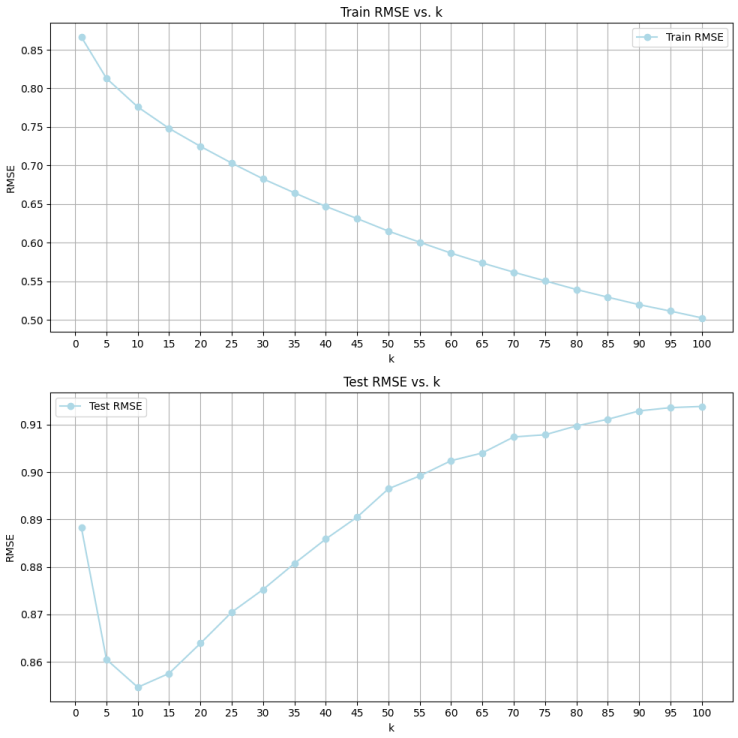

Figure A1: Training and Test Sample RMSE for Matrix Completion with Different Ranks

period t follows $N(\mu_{t+1}, \Sigma_{t+1})$. Then the KL-divergence between $\tilde{\theta}$ and θ_0 is calculated as below:

$$KL(\mathcal{N}(\mu_{t+1}, \Sigma_{t+1}) \mid\mid \mathcal{N}(0, I)) = \frac{1}{2}\left(\text{tr}(\Sigma_{t+1}) + \mu_{t+1}^T \mu_{t+1} - k - \log(\det(\Sigma_{t+1}))\right) \quad (28)$$

where $\text{tr}(\Sigma_{t+1})$ is the trace of the posterior covariance matrix Σ_{t+1}, $\mu_{t+1}^T \mu_{t+1}$ is the dot product of the posterior mean vector μ_{t+1} with itself, k is the number of dimensions, $\log(\det(\Sigma_{t+1}))$ is the natural logarithm of the determinant of the posterior covariance matrix Σ_{t+1}. Note that the determinant of the identity matrix is 1, and the logarithm of 1 is 0, which simplifies the last term. We present the results in Figure A2. The results show a higher KL divergence for the self-exploring consumers compared to RS-dependent consumers, confirming the main pattern identified in Figure 2.

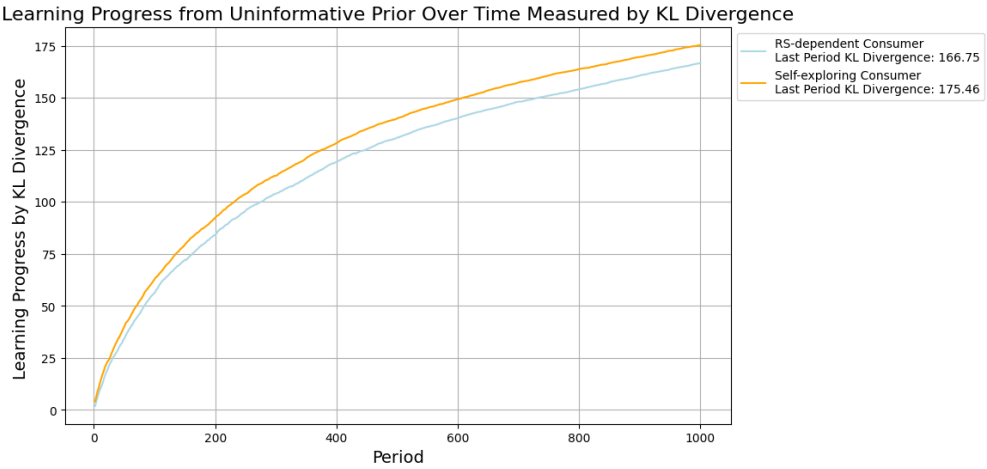

Figure A2: KL Divergence from the Uninformative Prior Over Time for Different Types of Consumers

D Main Results, Robustness Check and Extensions

D.1 Welfare Measurement Instead of Regret

We use the analysis presented in Figure 3 and create the plot for expected utility instead of expected regret. The pattern is the qualitatively the same: although expected utility is higher for the RS-dependent consumer when the recommendation system is available, this consumer has a lower expected utility when she makes decisions on her own, which is due to insufficient learning. One advantage of the expected utility measure is its interpretability in a 1-5 rating scale. The average movie chosen by the RS-dependent consumer has 3.80 rating, which is higher than the 3.69 average rating by the self-exploring consumer. However, if the RS-dependent makes decisions independent of the algorithmic recommendation, the average rating of the chosen movie would be 3.62, indicating worse decision-making ability compared to the self-exploring consumer.

D.2 Using Actual Rating Instead of Imputed Rating

Following Bresler et al. [2014], we only consider the top 200 consumers and the top 400 movies for the simulation, so we could use the real rating data not imputed rating. The resulting consumer-movie-rating matrix has 70% nonzero entries. The mean rating of this subsample is 3.73, which is slightly higher than the full sample (3.56). We spilt the 200 consumers' sample into training sample (70%, 140 consumers) and test sample (30%, 60 consumers) and use the test sample for the cold-start problem simulation. One advantage of this approach is that we use the actual ratings observed in the data, not the imputed ones. Figure A4 shows the

Figure A3: Expected Utility Over Time for Decision Sequences

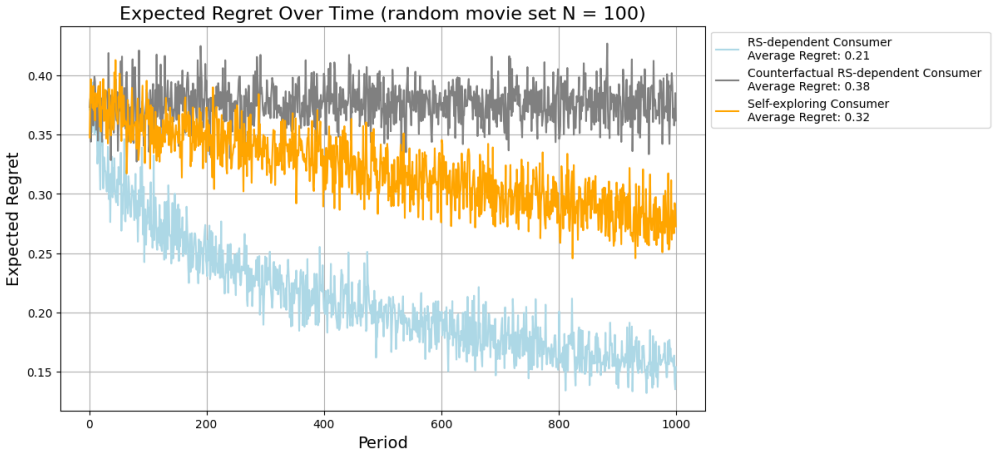

Figure A4: Expected Regret Over Time Using Actual Ratings

results from this practice and confirm the same qualitative insights as our analysis in the main text: Figure A4 shares similar patterns as Figure 3.

D.3 Consumer's Preference Parameter Dimension

One of the main remarks in our theoretical framework is that the expected regret would increase with respect to d. When the dimensionality of the movie features increases, learning becomes more complex for consumers, leading to more mistakes, thereby leading to higher regrets. Figure A5 shows how the preference parameter dimension d affects consumers' expected regret measures. The x-axis represents the consumer's preference dimension $d = 20, 40, \ldots, 150$, and the y-axis represents the average expected regret across 1000 periods.

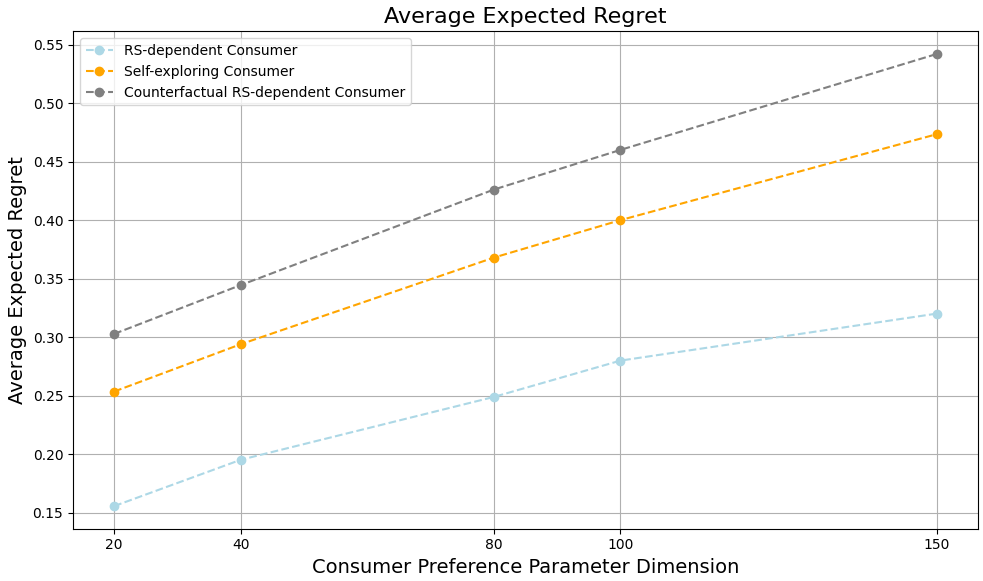

Figure A5: Expected Regret and Counterfactual Regret Over Time for Different Preference Dimensionality

We set $d = 100$ in the main body of this paper.[13]

We have a few observations from Figure A5. First, the basic pattern in Figure 3 persists: the self-exploring consumer's expected regret is higher than the RS-dependent consumer and lower than the counterfactual regret for the RS-dependent consumer, regardless of the values of d. Second, as the preference dimensionality increases, the expected regret measures increase. This result is consistent with the theoretical prediction that expected regret would increase with respect to d. Finally, the gap between the self-exploring consumer and the RS-dependent consumer's average expected regret expands when d is large, indicating that the RS's information advantage is larger when the learning system is more complex. In addition, the gap between the expected counterfactual regret for the RS-dependent consumer and the self-exploring consumer's expected regret is also larger when d is large, showing that the learning loss is greater when consumers have a complicated preference system.

D.4 Consumer's Action Set Size

In the main simulation, we use an action set $\mathcal{A}$ of cardinality 100 that contains 100 random movies in each time period. However, in reality, it is possible that consumers would consider fewer movies, implying that the actual action set $\mathcal{A}$ would contain fewer movies than 100. In

[13]With every preference parameter dimension d, the RS would choose the optimal rank again. The RS's optimal rank r for $d = 20, 40, 80, 100, 150$ are $r = 4, 5, 10, 10, 20$ respectively.

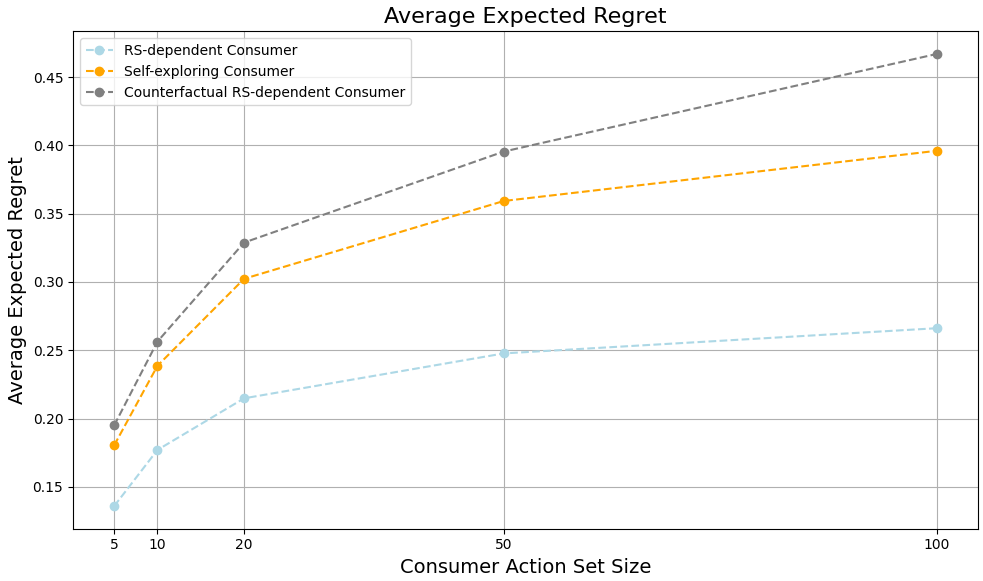

Figure A6: Expected Regret and Counterfactual Regret Over Time for Different Sizes of the Action Set

this extension, we want to see how the regret patterns would change with respect to the size of the action set. Figure A6 shows how the action set size affects consumers' expected regret measures. The x-axis represents the size of the consumer's action set $\mathcal{A}$ (5 movies, 10 movies, ..., 100 movies), and the y-axis represents the average expected regret across 1000 periods.

Our primary results hold when the consumer's action set size (how many movies to check each time) is between 5 and 100. When the action set is small, the personalized algorithm performs more similarly to the self-exploring consumer. However, as the complexity of the decision-making problem grows, the gap between groups widens.

D.5 Self-exploring Consumer's Information Advantage

Next, we consider an extension where consumers have a more informed prior than the personalized recommendation system, because they have watched some movies before and have a greater degree of certainty about their own preferences. We measure the information advantage of the consumer as she has already watched some movies before the first period (she starts to watch movies from period $-N$). Figure A7 shows how the self-exploring consumer's prior information (prior periods' length) affects consumers' expected regret measures. The x-axis represents the number of movies the self-exploring consumer has already watched before the first period (50, 100, ..., 1000), and the y-axis represents the average expected regret across 1000 periods. For example, 500 on the x-axis represents the scenario where

Figure A7: Expected Regret and Counterfactual Regret Over Time with Consumer's Information Advantage

the self-exploring consumer has started to watch movies from period -500. 0 on the x-axis represents the case in the main body of the paper, where the self-exploring consumer has no prior knowledge. Please note that the personalized algorithm does not have any prior knowledge.

There are two insights from Figure A7. First, the more prior information the self-exploring consumer has, the lower the regret she experiences throughout the 1000 periods in the game. In addition, the algorithm's information advantage is significant. Even if the self-exploring consumer has watched movies for 1000 periods before the game, her average regret is still higher than the RS-dependent consumer's average regret. This shows that the personalized algorithm's information advantage is substantial and not easily overcome by the consumer's prior knowledge.

D.6 Personalized Algorithm's Rank

We argue that persoanlized algorithm has the information advantage so that it can decompose the consumer's preference weights into a lower-dimensional representation. However, it is not always the case that a smaller r leads to lower regret. When r is very small, the RS-dependent consumer explores only minimally. Conversely, when r is very large, the RS-dependent consumer may explore excessively, resulting in high expected regret for RS-dependent consumers. Figure A8 shows how the personalized recommendation system's low

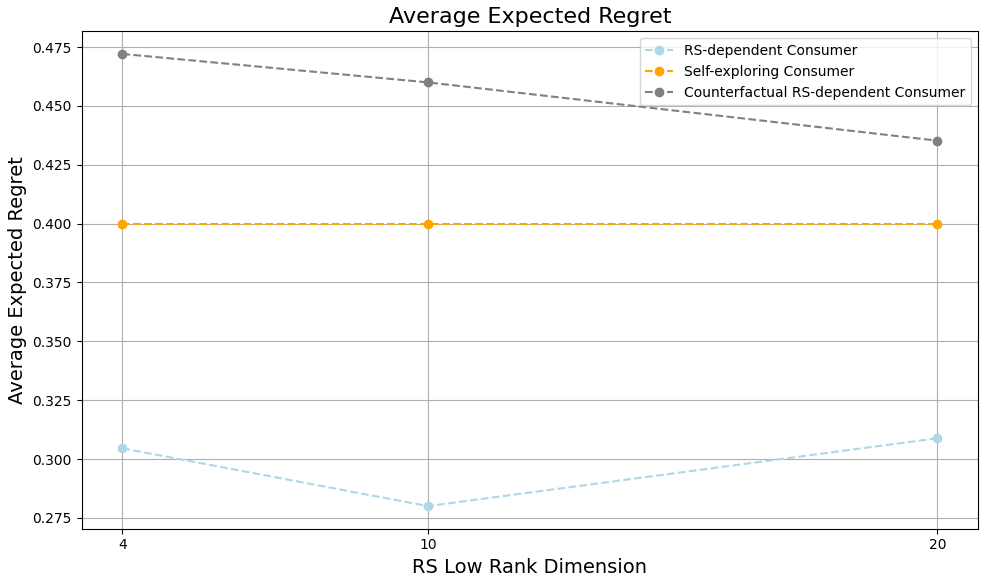

Figure A8: Expected Regret and Counterfactual Regret Over Time for Different Algorithm's Rank Specifications

rank representation r affects consumers' expected regret measures. The x-axis represents the personalized recommendation system's low rank representation $r = 4, 10, 20$, and the y-axis represents the average expected regret across 1000 periods.

As shown in Figure A8, the RS-dependent consumer receives the lowest expected regret with the optimal rank 10. When the rank is lower than optimal, the process becomes somewhat sub-optimal because the algorithm fails to fully capture the complexity of the decision-making process. Conversely, when the rank exceeds the optimal level, the algorithm engages in more exploration, leading to more sub-optimal choices for the RS-dependent consumer.Despite these sub-optimal rank scenarios, the RS-dependent consumer still achieves significantly lower expected regret compared to the self-exploring consumer, underscoring the value of the algorithm's information advantage. Interestingly, the expected counterfactual regret for the RS-dependent consumer decreases as r exceeds the optimal rank. This is likely because the recommendation system promotes greater exploration at higher ranks, thereby improving the RS-dependent consumer's ability to make independent decisions.

E Algorithm for the Random Availability Policies

In Section 5.1, we propose the policies with random availability. We show the detailed pseudo-code in Algorithm 4. Most parts of this algorithm are largely similar to Algorithm 3. The main difference is that the algorithmic recommendation is only available with a

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

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

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

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

probability p.

F Algorithm for the Self-Regulated Policies

In Section 4.4.2, we propose the policies with self-regulation. We show the detailed pseudo-code in Algorithm 5. Most parts of this algorithm are largely similar to Algorithm 3. The main difference is that the algorithm chooses an action considering the change in Shannon Entropy.

Algorithm 5 Choice and Learning for the RS-Dependent Consumer under RS with Regulation

Input: $\mu_{i,0}^{(\theta)}, \Sigma_{i,0}^{(\theta)}, \mu_{i,0}^{(\gamma)}, \Sigma_{i,0}^{(\gamma)}, F, \mathcal{A}, \mathcal{T}, \lambda\Delta(H(A_{i,j,t}^*))$

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

for $t = 0 \rightarrow \mathcal{T}$ **do**
  $\tilde{\gamma}_{i,t} \sim N(\mu_{i,t}^{(\gamma)}, \Sigma_{i,t}^{(\gamma)})$ ▷ RS: Distribution Sampling
  $A_{i,t}^{RS} \in \text{argmax}_{A \in \mathcal{A}} \tilde{\gamma}_{i,t}^T F^T A - \lambda \times \Delta(H(A_{i,j,t}^*))$ ▷ RS: Recommendation Selection
  $A_{i,t} \leftarrow A_{i,t}^{RS}$ ▷ Consumer: Action Selection
  Refer to Algorithm 3 for belief updating steps.
end for

The way we get the regulation term follows Algorithm 6:

Algorithm 6 Change in Entropy Calculation Upon Movie Selection

Input: $\mu_{i,0}^{(\theta)}, \Sigma_{i,0}^{(\theta)}, \mathcal{A}, \mathcal{T}, n$
Output: $\Delta H(A_{i,j,t}^*)$ for each period t and movie j

for $t = 0 \to \mathcal{T} - 1$ **do**
    for each movie $j \in \mathcal{A}$ **do**
        Calculate $H(A_{i,t}^*)$ using Equation (10) with $\mu_{i,t}^{(\theta)}$ and $\Sigma_{i,t}^{(\theta)}$
        Simulate the selection of movie j at period t
        Update $\mu_{i,t+1}^{(\theta)}$ and $\Sigma_{i,t+1}^{(\theta)}$ based on the selection of movie j
        Calculate $H(A_{i,t+1,j}^*)$ using updated beliefs $\mu_{i,t+1}^{(\theta)}$ and $\Sigma_{i,t+1}^{(\theta)}$
        $\Delta H(A_{i,j,t}^*) = H(A_{i,t+1,j}^*) - H(A_{i,t}^*)$ ▷ Change in entropy from selecting movie j
    end for
    Record $\Delta H(A_{i,j,t}^*)$ for each movie j
end for