How Does Competition Affect Exploration vs. Exploitation? A Tale of Two Recommendation Algorithms

H. Henry Cao, Liye Ma, Z. Eddie Ning, Baohong Sun*†

Dec, 2022

Abstract

Through repeated interactions, firms today refine their understanding of individual users' preferences adaptively for personalization. In this paper, we use a continuous-time bandit model to analyze firms that recommend content to multi-homing consumers, a representative setting for strategic learning of consumer preferences to maximize lifetime value. In both monopoly and duopoly settings, we compare a forward-looking recommendation algorithm that balances exploration and exploitation to a myopic algorithm that only maximizes the quality of the next recommendation. Our analysis shows that compared to a monopoly, firms competing for users' attention focus more on exploitation than exploration. When users are impatient, competition decreases the return from developing a forward-looking algorithm. In contrast, development of a forward-looking algorithm may hurt users under monopoly but always benefits users under competition. Competing firms' decisions to invest in a forward-looking algorithm can create a prisoner's dilemma. Our results have implications for AI adoption as well as for policy makers on the effect of market power on innovation and consumer welfare.

Keywords: AI, bandit, multi-homing, recommendation algorithm, customization, personalization, content, competition, experimentation, reinforcement learning.

*The authors thank J. Miguel Villas-Boas, T. Tony Ke, and participants at the SICS conference and the Bass FORMS conference for helpful suggestions. Comments are welcome.

†H. Henry Cao is Professor of Finance at Cheung Kong Graduate School of Business. Liye Ma is Associate Professor of Marketing at the Robert H. Smith School of Business, University of Maryland. Z. Eddie Ning is Assistant Professor of Marketing and Behavioural Science at the Sauder School of Business, University of British Columbia. Baohong Sun is Dean's Distinguished Chair Professor of Marketing at Cheung Kong Graduate School of Business. Emails: hncao@ckgsb.edu.cn, liyema@umd.edu, eddie.ning@sauder.ubc.ca, bhsun@ckgsb.edu.cn.

1 Introduction

Two trends have radically transformed the marketing landscape in the past two decades. First, the advent of e-commerce, social media, and mobile marketing has made firm-consumer interactions increasingly frequent and digitized (Godes and Mayzlin 2004, Fader and Winer 2012, Kannan and Li 2017). These interactions have produced a fine-grained digital consumer footprint that provides valuable information to firms. Second, the past decade has also witnessed exponential growth in leveraging data and computing power in the business world. The rapid development of cloud computing, big data, machine learning, and AI has provided powerful tools to assist in large-scale automated decision making. These tools have greatly increased firms' ability to understand and fulfill customers' needs in real time (Chintagunta, Hanssens, and Hauser 2016, Huang and Rust, 2018, Ma and Sun, 2020). Driven by these trends, firms now routinely analyze historical interactions with consumers to infer their preferences and generate customized offerings, often in real time. Prominent examples abound. Personalized product recommendation systems are now indispensable on e-commerce websites such as Amazon and Taobao. Digital advertisements are increasingly personalized based on a user's past behaviors. Even more prevalent are popular social media and content platforms such as Facebook, YouTube, Spotify, TikTok, and many news media sites that customize content feeds to individual users based on their historical interactions with the platform. Such personalized real-time customization is being conducted through increasingly sophisticated AI algorithms, which have become a major source of competitive advantage for many firms.

While the scale and scope may be new, the practice of learning about consumers and making customized recommendations dates back to the early days of marketing (Wedel and Kannan 2016). Conceptually, three paradigms exist for recommendations. First, using historical data, a firm can learn about consumer preferences at the group level in a static fashion, and make customized recommendations based on the inferred segmentation. A rich body of literature, e.g., dynamic choice models, incorporates consumer heterogeneity in an increasingly sophisticated manner, enabling effective segmentation and personalization (Kamakura and Russell 1989, Rossi, McCulloch, and Allenby 1996). These methods are now commonly used in industry to enhance sales, profit, customer satisfaction, and loyalty. Since such models are typically estimated only periodically using datasets containing large batches of historical observations, and decisions are updated infrequently (often non-machine-assisted human decisions), this paradigm is non-adaptive to users' real-time behaviors.

In the second paradigm, a firm can refine its learning adaptively using new information, potentially on a real-time basis (Zhang and Krishnamurthi 2004, Steckel et al. 2005, Sun,

Li, and Zhou 2006). In this paradigm, as time passes and new data become available, the firm continuously update its understanding of consumer preferences based on the new information. At each point in time, the firm makes personalized offerings based on its best understanding of a consumer's preferences. Many statistical techniques, such as machine learning, can help firms perform such adaptive learning and recommendations. For example, today's recommendation systems use methods such as content-based filtering and collaborative filtering to generate candidates to recommend. They then use a predictive model to rank them by objectives such as click-through rate or session watch time (Google Developers 2020). These automated algorithms are increasingly common to help firms effectively adapt to and act on a constant stream of incoming data in real time. We call this second paradigm *the myopic recommendation algorithm*. The word "myopic" highlights that these algorithms only aim to offer the "best" recommendation at the moment, without considering the long-term benefit of acquiring knowledge and improving personalized targeting.

Going one step further from the myopic algorithm, a third and more powerful recommendation paradigm is emerging. A firm not only learns adaptively from past information, but also takes a forward-looking perspective in its recommendations to proactively gather new information in a guided manner. For example, based on the current understanding, a consumer is most likely to enjoy a specific type of content, but the firm may instead find it useful to recommend something different. Such recommendation may lead to a reduction of service quality and a lower profit in the short term, but it speeds up the learning of consumer preferences, which can then improve future recommendations and enhance customer retention. Central to this paradigm is the exploitation-exploration trade-off, where the firm has to balance the conflict between maximizing the current payoff and acquiring new knowledge. The adoption of this third paradigm is partly driven by the recent success of reinforcement learning, which allows computers to better approximate human decision making (Sutton and Barto 2018). Major social media and content platforms, such as YouTube, are also developing reinforcement learning algorithms to maximize users' long-term satisfaction with the system (Chen et al. 2019). We call this third paradigm *the forward-looking recommendation algorithm*.

The proliferation of consumer data has attracted considerable attention from scholars in multiple fields. Research in computer science has developed a vast and powerful tool set to extract information from large volumes of data. Empirical research in marketing and economics has consistently confirmed the value of consumers' digital footprint on understanding their preferences and decisions (e.g., Winer and Neslin 2014). However the theoretical implications of firms' continuous personalization, especially under competition, are noticeably left out of both streams of research. Developing the capability to learn and recommend in

real time requires considerable investment. Adopting a forward-looking solution framework such as reinforcement learning is an even more demanding initiative. To optimize investment decisions, it is crucial for firms to understand the value of such continuous, forward-looking personalization in different competitive scenarios. It is also important for policy makers to understand how competition, or the lack thereof, affects the adoption of advanced algorithms and the subsequent consumer welfare implications.

In this paper, we study the competition between content platforms, such as YouTube and TikTok, which compete for users' attention through personalized content recommendations. This is a representative setting where firms offer adaptive personalized offerings to each user, and forward-looking, multi-homing users dynamically switch between different platforms. We examine myopic users as a special limiting case. While our analysis focuses on advertising-supported content recommendations, key intuitions and findings from this paper can potentially be generalized to other similar settings, such as product recommendations on e-commerce websites or targeted advertising.

In our model, users differ in their preferences for different types of content. Using a user's responses to past recommendations on the platform as noisy signals, a firm gradually learns about the user's preferences and adjusts the recommendations adaptively. We formulate firms' problem as a continuous-time, multi-armed bandit problem. This framework incorporates key factors such as firms' continuous learning of user preferences, adaptive responses to real-time information, and forward-looking optimization in a parsimonious manner. For simplicity, we focus on firms' learning from repeated interactions with a single user, thus ignoring potential learning across users.

Studying such a market allows us to focus on sequential learning and the optimal content personalization strategy by abstracting away from other strategic decisions such as advertising, pricing, and positioning. The setup also abstracts away from other issues related to the value of data, such as privacy (e.g., Acquisti, Taylor, and Wagman 2016, Ke and Sudhir 2022), sharing or selling data to third parties (e.g., Chen, Narasimhan, and Zhang 2001), or using data from existing customers to target similar customers (e.g., Schafer et al. 2007).

In light of rising concerns expressed by regulators over major tech firms' market power, our research investigates the effects of market power on firms' incentives to develop advanced AI algorithms and the implications on industry profit and consumer welfare. We address three questions. First, how does the presence of competition affect the optimal trade-off between exploration and exploitation in personalization? Many recommendation algorithms have been found to prioritize popular, mainstream content over less popular, niche content. Firms should understand how to balance between mainstream and niche content in their recommendation algorithms in different competitive scenarios. Second, does the forward-

looking algorithm provide additional value over the myopic algorithm to firms? And how does upgrading from the myopic algorithm to the forward-looking algorithm affect consumer welfare? Third, how does competition affect firms' decisions to invest in the forward-looking algorithm? These questions have theoretical and managerial implications for technology adoption as well as regulation.

We compare the myopic recommendation algorithm that only focuses on exploitation to the optimal forward-looking algorithm that balances exploitation and exploration. We analyze two market structures: (1) when a firm monopolizes a user's attention, and (2) when two firms compete for a user's attention. In the competitive scenario, the user strategically multi-homes by choosing which firm to consume content from at each moment. In this dynamic multi-homing problem, the two firms and the user each solves a bandit problem. Our study is among the first to study such a multi-agent bandit problem arising from multi-homing under competition.

In this paper, a recommendation algorithm is a mapping from information sets to content types. The optimal forward-looking algorithm in the monopoly case solves the monopoly's dynamic optimization problem, while the optimal forward-looking algorithm in the duopoly case is synonymous with equilibrium strategy. This approach allows us to abstract away from specific algorithms and focus instead on the value added to firms and consumers when algorithms gain forward-looking capability. This approach is in contrast to recent papers that examine outcomes generated by competition between agents that deploy Q-learning, a reinforcement learning algorithm that has recently received attention in Economics (Johnson et al. 2020, Asker et al. 2022, Banchio and Mantegazza 2022, Banchio and Skrzypacs 2022).

We derive closed-form solutions to the simultaneous bandit problem, and the solutions reveal several important insights. For a monopoly, the additional value of developing the forward-looking algorithm is non-monotonic in the firm's prior knowledge about a user's preferences. To expedite learning of consumer preferences, the forward-looking algorithm induces the firm to recommend more niche content, i.e., to customize more than the myopic algorithm does. The exploration-exploitation trade-off also means that the forward-looking algorithm leads to reduced profit in the near term, although the profit increases later to compensate for the near-term loss.

The situation changes substantially, however, when firms have to compete for a user's attention. The presence of competition pushes the optimal forward-looking algorithm to shift towards exploitation by recommending less niche content, due to less room for strategic experimentation. More importantly, the ability for the user to switch platforms forces firms to optimize with respect to the user's time preferences. This novel insight implies that the optimal forward-looking algorithm and the value it provides must depend on the user's

discount rate instead of the firms' discount rate. As users become more impatient, the forward-looking algorithm moves closer to the myopic algorithm and recommends less niche content. When users are fully myopic, the myopic algorithm becomes optimal.

We analyze firms' incentives to invest in upgrading from the myopic algorithm to the forward-looking algorithm. First, our analysis shows that, when users are sufficiently impatient, firms under competitive pressure reap less benefit from the algorithmic upgrade than a monopoly. Second, in competition, upgrading to the forward-looking algorithm allows a firm to steal demand from the competitor. As a result, when the cost of developing the forward-looking algorithm is in an intermediate range, the equilibrium exhibits prisoner's dilemma: both firms invest in the forward-looking algorithm but the total industry profit is lower than no investment. User myopia can be both a curse and a blessing. A more impatient user decreases the value from the forward-looking algorithm but may help firms avoid the prisoner's dilemma. Similarly, a lower cost of developing the forward-looking algorithm can hurt firms by inducing the prisoner's dilemma.

There is also a trade-off between innovation incentives and consumer welfare. In particular, when users are sufficiently impatient relative to firms, a monopoly has a higher incentive than firms under competition to develop the forward-looking algorithm, but such technological advancement hurts users due to excessive customization. In contrast, under competition, the adoption of the forward-looking algorithm is always beneficial to users, but firms derive less value from such an algorithm than a monopoly does, hindering investment. These results have implications for policy makers who care about both consumer welfare and the adoption speed of AI technologies. In recent years, regulators around the world have increased scrutiny over the market power of algorithmic-driven tech firms (e.g., U.S. House 2020, Competition and Markets Authority 2020). At the same time, many governments have launched national AI strategies to push for faster adoption of AI technologies in the face of global competition (Berryhill et al. 2020). Our study shows the potential conflict between these two policy goals.

The rest of the paper is organized as follows. After reviewing the relevant literature in Section 2, we set up the monopoly model in Section 3. We then study competitive scenarios in Section 4. Section 5 explores extensions, including a game of algorithmic investment. Concluding remarks are offered in Section 6.

2 Literature Review

Dynamic Programming and Reinforcement Learning

The core idea of reinforcement learning (RL) is deriving solutions to stochastic dynamic programming problems under uncertainty. In our context, firms need to learn about consumer preferences and trade off instantaneous costs with future payoffs to maximize long-term profit. Facing the inter-temporal trade-off between exploration and exploitation, an agent solves a statistical decision model and learns about the payoff of different options over time through experimentation. A stream of marketing research derives and studies the properties of this problem in various applications of marketing decision support system. Applying to catalogs, Gonul and Shi (1998) show that the optimal mailing policy resulting from a dynamic programming model significantly outperforms its single-period counterpart. Applying a dynamic-programming-based approach to newspaper subscriber data, Lewis (2005) computes price paths that maximize profit over long-term relationships with customers. Li, Sun, and Montgomery (2006) derive an optimal multi-step, multi-segment, and multi-channel cross-selling campaign process that instructs firms when to target whom with what product using which channel. Sun and Li (2005) formulate firms' service allocation decisions as solutions to a dynamic programming problem and discuss how the experimental nature of interactive learning and acting on customer information improve customer experience and firm profit. Sun, Li, and Zhou (2006) present a conceptual framework of customer-centric marketing-mix decision making as a solution to dynamic programming problems with a two-step interactive procedure (adaptive learning and proactive marketing decisions). Ching (2010) present a dynamic oligopoly model of the drug market, in which both firms and patients learn about the quality of generic drugs over time through patients' experiences. Lin, Zhang, and Hauser (2015) consider a dynamic experiential learning problem in which consumers learn brand quality over time while facing random utility shocks. They show empirically that an index-based heuristic solution is nearly optimal and perform significantly better than myopic learning.

Recently, RL has been applied to marketing problems with the same idea of continuously following consumers to deliver the right intervention to the right consumer at the right time using the right channel. For example, formulating personalized news recommendations as a bandit problem, Li et al. (2010) propose an algorithm that generates a sequence of articles based on the historical activities of a user, and the article recommendation policy adapts based on users' real-time feedback with the goal of maximizing total user clicks in the long run. Theocharous, Thomas, and Ghavamzadeh (2015) formulate a personalized advertising recommendation system as an RL problem to maximize lifetime value (LTV) and show improvement over a myopic solution with supervised learning. Hybrid and concurrent RL

are proposed by Li et al. (2015) and Silver et al. (2013) to better incorporate lifetime value of customers and customer interactions. Other researchers have used the multi-armed bandit framework to improve adaptive online advertising (e.g., Urban et al. 2014, Schwartz, Bradlow, and Fader 2017), web content optimization (e.g., Agarwal, Chen, and Elango 2008, Hauser, Liberali, and Urban 2014), and pricing (e.g., Misra, Schwartz, and Abernethy 2019). Schwartz, Bradlow, and Fader (2017) propose the Thompson sampling algorithm (which assigns a treatment with a probability equal to the probability that the treatment is optimal) for the optimal allocation of advertisements. Misra, Schwartz, and Abernethy (2019) propose a dynamic price experimentation policy in online retailing by adaptively assigning users to the treatment with the highest potential. By adjusting adaptively, reinforcement learning improves over the static approach because successful treatments are rewarded by assigning more users to these treatments (Athey and Imbens 2019, Sutton and Barto 2018).

This fast developing literature demonstrates the potential of adaptive learning and customization using dynamic programming and RL approaches. The proliferation of empirical research and algorithm development highlights the need for additional studies to investigate the properties of continuous, forward-looking customization under competition.

Multi-Armed Bandits

From a modelling perspective, our research is related to the literature in economics that models learning and experimentation as a multi-armed bandit (MAB) problem (Rothchild 1974, Weitzman 1979, Keller and Rady 1999). Bolton and Harris (1999) and Keller, Rady, and Cripps (2005) study experimentation in teams, and show that members of a team underexperiment as they try to free-ride on information from others' experiments. In the contexts of experience goods market and labor market, respectively, Bergemann and Välimäki (1996) and Felli and Harris (1996) study a case where an agent pays for experiments that are owned by separate sellers who compete with each other on price. However, these papers do not consider a case where multiple bandits compete with each other to determine who gets the right to experiment, which is the focus of this paper.

This problem of competition and MAB has also received attention in computer science. The paper closest to ours is Aridor et al. (2021). Both papers study two multi-armed bandit algorithms that compete for users over time, and observe that competition pushes firms towards exploitation and disincentivizes firms from adopting better algorithms. However, there are a few key differences between our models. Most importantly, users in Aridor et al. (2020) are short-lived and cannot observe other users' experience. In contrast, firms in our model face a long-lived user who also solves its own bandit problem when choosing which firm to visit over time. Differences in our models also lead to very different conclusions.

Aridor et al. (2021) find that when facing utility-maximizing consumers, both firms adopt a myopic algorithm in equilibrium. In contrast, the equilibrium algorithm in the current paper is still forward-looking. By studying a long-lived user, we are also able to examine how the user's time preferences affect the equilibrium choices of algorithms, firms' incentives to invest in advanced algorithms, and the welfare impact of such investment.

Ke, Li, and Safronov (2021) find that competition pushes firms to behave more myopically with respect to assigning tasks to employees in a model of dynamic career concerns. Our paper differs both in the context and the mechanism that cause the effect. In Ke, Li, and Safronov (2021), the employee reaps all future surplus from reputation, causing firms to behave myopically. In our model, competing firms cater to the user's time preference, thus acting more myopically when the user is more myopic. Hansen, Misra, and Pai (2020) and Calvano et al. (2020) numerically study competition between pricing algorithms, but our paper instead focuses on competition between content recommendation algorithms without pricing decisions.

Micro Models of AI Technology

A recent stream of literature in economics and marketing has built theoretical models to study the general microeconomic impact of AI technology. Agrawal, Gans, and Goldfarb (2018a) argue that the current wave of AI technology can be thought of as an improved ability to predict future states. Agrawal, Gans, and Goldfarb (2019) split the decision-making process between machine prediction of states and human judgment of utility, and show that human judgment can be either complement or substitute for machine prediction. Agrawal, Gans, and Goldfarb (2018b) consider subscription pricing of such prediction technology. Miklos-Thal and Tucker (2019) show that algorithmic pricing can lead to collusive outcomes. Dogan, Jacquillat, and Yildirim (2019) and Athey, Bryan, and Gans (2020) study the effect of AI on delegation of decision authority in the presence of principal-agent problem. Berman and Katona (2020) investigate when curation algorithms do and do not create polarization in social networks. Liu, Yildirim, and Zhang (2019) consider price discrimination when consumers purchase from AI-enabled home devices. Xu and Dukes (2020) study personalized pricing when data analytics enables firms to have more information on consumer preferences than consumers have. These papers focus on documenting the general impact of machine-aided decision-making but they do not investigate the inter-temporal trade-off between exploitation and exploration.

Continuous-time Decision-Making

We study a continuous-time model with sequential arrival of information, which approximates the nature of real-time learning and acting on customer information. There is a related stream of literature on the continuous acquisition of information before an agent undertakes an irreversible action such as purchase or investment (e.g., Branco, Sun, and Villas-Boas 2012, Ke, Shen, and Villas-Boas 2016, Fudenberg, Strack, and Strzalecki 2018). Ke and Villas-Boas (2019) consider continuous learning of multiple alternatives before committing to a choice. These papers capture the continuous nature of learning and solve the optimal solution to the single decision-maker problem. Ning (2021) expands the single-agent problem into a continuous-time game by adding dynamic pricing while a buyer and a seller continuously receive information on their match value. Villas-Boas and Yao (2020) model a firm's optimal advertising retargeting policy to a consumer who continuously searches for product information. Deb, Öry, and Williams (2018) study a continuous-time crowdfunding game between a long-lived donor and short-lived buyers as information on the total donation arrives over time. In contrast to these papers, the current paper features competition between two firms, each deciding its own experimentation strategy and receiving private information, while factoring in competitive responses from the other firm.

Personalization

Our model also relates to the literature on personalization based on past behaviors. The literature on behavior-based price discrimination (e.g., Villas-Boas 1999, 2004, Fudenberg and Tirole 2000, Acquisti and Varian 2005, Pazgal and Soberman 2008) shows that personalized pricing based on past purchase behaviors generally hurts firms by intensifying price competition. Zhang (2011) expands the literature to allow for endogenous product designs that influence the information that firms collect. The current paper does not consider pricing. Instead, we focus our attention on personalized product offerings. We allow for rich dynamics where each firm makes personalized offerings over an infinite number of periods, where each decision affects both immediate profit and firms' future information about the customer. Our paper also relates to the extensive literature on targeting, but instead of deciding whether to target a consumer, in our context, firms decide what content to offer to that consumer.

Innovation

The paper contributes to the literature on competition and innovation. Dasgupta and Stiglitz (1980) and Spence (1984) argue that increasing the number of firms in the industry decreases firms' incentives to invest in cost reduction, whereas Aghion et al. (2005) and Vives (2008)

show that increasing the number of firms can foster innovation when the level of competition is low. While the aforementioned papers model innovation as a reduction in marginal cost or an increase in labor productivity, this paper considers a very different type of innovation. We consider the technological upgrade from a myopic algorithm to a forward-looking algorithm, and show that competition decreases the return from this upgrade when consumers are impatient.

3 Monopoly Model

Consider a firm that provides personalized content to each user over time. For example, platforms like YouTube, TikTok, Spotify, and Google News recommend personalized content to users based on their historical behaviors on the site. The firm's objective is to increase user engagement with the content on the site through increased views over a user's lifetime, which boosts monetization, such as advertising revenue.[1] Table 1 presents the notations for the monopoly case.

Each user has a unit demand for content at each time t. In this section, we consider the case of a true monopoly where the user visits the firm at each time t. In Section 4, we give the user the option of visiting other firms, so that firms have to compete for the user's time.

For a given user at a given time, the firm can choose to recommend mass-market content (M) or niche-market content (N). For simplicity, we assume that there are two types of niche-market content, denoted as N_1 and N_2. All users equally enjoy mass-market content, but for niche-market content, they have different preferences. Some users enjoy N_1 more often than N_2, and vice versa. Let $T \in \{N_1, N_2\}$ denote the focal user's preferred type of niche-market content. This is drawn by Nature and is unknown to the firm.

Let $S_t \in \{M, N_1, N_2\}$ denote the type of content that the firm recommends to the focal user at time t. If a user receives niche-market content, the user likes the content with probability $\alpha > 0.5$ if the content matches the user's preferred content type, and with probability $1-\alpha$ if it is a mismatch. The parameter α captures the consistency of the user's preferences for niche-market content. An α close to 1 implies that the user always likes the same content type, whereas an α close to 0.5 implies that the user's preferences for content types are nearly random.

If the user receives mass-market content, the user likes the content with probability c. For

[1] There are various ways to display ads and generate revenue (pay-per-click/pay-per-impression), but they all share the common characteristic that revenue is proportional to user engagement, which is captured in our model. Note that we do not study the customization of advertising in this paper, but only customization of content.

the interesting case, we assume that $\frac{1}{2} < c < \alpha$, otherwise the firm either never recommends mass-market content or always recommends mass-market content. Note that a user is more likely to engage with mass-market content than randomly selected niche-market content. This reflects the general popularity of mass-market content.

So the probability that a user of type T likes the content recommended at time t, denoted as $y(T, S_t)$, can be written as:

$$y(T, S_t) = \begin{cases} \alpha & \text{if} \quad S_t = T \\ c & \text{if} \quad S_t = M \\ 1 - \alpha & \text{otherwise} \end{cases} \tag{1}$$

If the user likes the recommended content, the firm earns an advertising profit of size p and the user gets a utility of u, both of which can be normalized to 1.[2] So the expected flow profit generated from recommending content type S_t given that the user's preferred content type is T is simply $\pi_t = p * y(T, S_t) = y(T, S_t)$, and the standard deviation of the flow profit is

$$\sigma_t = \sqrt{\alpha(1-\alpha) \cdot \mathbb{1}\{S_t \neq M\} + c(1-c) \cdot \mathbb{1}\{S_t = M\}} \tag{2}$$

To capture the idea that these interactions occur at a high frequency and the firm can monitor a user's behavior continuously, we use a continuous-time model, where the firm's cumulative profit, Y_t, accrues as

$$dY_t = y(T, S_t)dt + \sigma_t dW_t, \tag{3}$$

where $y(T, S_t)$, as defined in equation (1), is the expected profit flow, and σ_t, as defined in equation (2), is the instantaneous standard deviation, and W_t is a standard Wiener process.[3]

[2]The normalization is only with loss of generality when considering social surplus, or how surplus is split between users and firms. Because our model focuses on one focal user, we do not discuss social surplus in this paper. In the Online Appendix, we study a case where the firm can control the degree of monetization that affects both the flow profit as well as the speed of learning.

[3]In discrete time, we have:

$$E[Y_t] = \sum_{s=1}^{t} y(T, S_s) \quad \text{and} \quad Var(Y_t) = \alpha(1-\alpha) \sum_{0}^{t} \mathbb{1}\{S_t \neq M\} + c(1-c) \sum_{0}^{t} \mathbb{1}\{S_t = M\}$$

The noise is independent across time, and by the central limit theorem, the distribution of Y_t can be approximated by the Gaussian distribution $\mathcal{N}(E[Y_t],\ Var(Y_t))$ for large t. The unique continuous-time process with independent noise in increments that satisfies $Y_t \sim \mathcal{N}(E[Y_t],\ Var(Y_t))$, where

$$E[Y_t] = \int_0^t y(T, S_s)ds \quad \text{and} \quad Var(Y_t) = \alpha(1-\alpha) \int_0^t \mathbb{1}\{S_t \neq M\}ds + c(1-c) \int_0^t \mathbb{1}\{S_t = M\}ds$$

is

$$dY_t = y(T, S_t)dt + \sigma_t dW_t,$$

3.1 Information and Learning Process

At $t = 0$, the firm receives a binary signal on the user's type with accuracy $\lambda_0 > 0.5$. That is, the firm observes the correct user type with probability λ_0, and observes the incorrect type with probability $1 - \lambda_0$. Thus, the firm either has a prior belief that the user prefers N_1 with probability λ_0, or has a prior belief that the user prefers N_2 with probability λ_0. This represents the prior knowledge the firm has about this user. We can always relabel the two content types without loss of generality, so we can simply assume that the firm has a prior belief that the user prefers N_1 with probability λ_0.

Let λ_t denote the firm's *posterior* belief that the user prefers N_1 over N_2. The history of realized profit from the user serves as the information source. We have

$$\lambda_t = Pr(T = N_1 | F_t)$$

where F_t is the filtration generated by the past observations of profit from the user.

The Exploration vs. Exploitation Trade-off

Note that the firm only gains information about the user's preferences when it recommends niche-market content. Consider a scenario where λ_t is close to 0.5. There is enough uncertainty about the user's preferences that the user has a higher probability of liking mass-market content than niche-market content. Thus, to maximize the immediate profit, the firm should recommend mass-market content. However, the firm may still want to offer niche-market content, because the user's response to it reveals information about her preferences, which allows the firm to make better recommendations in the future. Thus, in this model, the firm's decision on whether to recommend niche-market content captures the trade-off between exploration and exploitation in a parsimonious way.

Updating of Posterior Belief

From the firm's perspective, with a belief λ_t, the expected profit flow from recommending content type S_t to the user at time t can be written as:

$$y(\lambda_t, S_t) = \begin{cases} \lambda_t \alpha + (1 - \lambda_t)(1 - \alpha) & \text{if } S_t = N_1 \\ (1 - \lambda_t)\alpha + \lambda_t(1 - \alpha) & \text{if } S_t = N_2 \\ c & \text{if } S_t = M \end{cases} \tag{4}$$

Because the firm gains no information when it recommends mass-market content, the posterior belief, λ_t, is only updated when the firm recommends niche-market content. From Liptser and Shiryaev (1977), the updating process of λ_t, when the firm recommends niche-

market content, follows the process

$$d\lambda_t = [\alpha - (1-\alpha)]\frac{\lambda_t(1-\lambda_t)}{\sigma_t^2}[y(T,S_t) - y(\lambda_t,S_t)]dt + [\alpha - (1-\alpha)]\frac{\lambda_t(1-\lambda_t)}{\sigma_t}dW_t \quad (5)$$

where the term $[y(T,S_t) - y(\lambda_t,S_t)]$ represents the new information, which is the difference between the expected flow profit under the current belief and the true expected flow profit. The speed of updating is weighted by the difference in outcomes between the right and the wrong actions, which is captured by $\alpha - (1-\alpha) = 2\alpha - 1$. The term σ_t is the standard deviation of flow profit from equation (2).

Because the expected value of $[y(T,S_t) - y(\lambda_t,S_t)]$ is zero, the change to the posterior belief, λ_t, has zero drift. We can simplify equation (5) to

$$d\lambda_t = \frac{\sigma(\lambda_t)}{\sigma_t^2}[y(T,S_t) - y(\lambda_t,S_t)]dt + \sigma(\lambda_t)dW_t \quad (6)$$

where $\sigma(\lambda_t)$ is the instantaneous standard deviation of λ_t, i.e.,

$$\sigma(\lambda_t) \equiv \frac{\lambda_t(1-\lambda_t)(2\alpha-1)}{\sigma_t} = \frac{\lambda_t(1-\lambda_t)(2\alpha-1)}{\sqrt{\alpha(1-\alpha)}}$$

Note from the above equation that the instantaneous standard deviation of λ_t, $\sigma(\lambda_t)$, increases in α (α is assumed to be > 0.5), so the posterior belief is more responsive to user behaviors when α increases. When α is larger, different types of users exhibit more different behaviors. Thus, information inferred from their behavior is more precise. The belief λ_t is updated faster, so the instantaneous volatility of λ_t is higher. Notice that as λ_t goes to either 0 or 1, the standard deviation will go to zero. In the limit as time goes to infinity, the user's preferences will be fully revealed. But for any finite length of time, there will be some amount of uncertainty regarding the user's preferences.

3.2 Firm's Decisions

The firm is risk neutral and maximizes the present value of discounted expected profits with a discount rate of r_f by choosing recommendation S_t as a function of λ_t, which is the belief at time t that the user prefers type N_1 over N_2. The firm can only recommend one unit of content to a user at a time.

The lifetime value of the user given a path of S_t is

$$V(\{S_t\},\lambda_0) = E\int_0^\infty e^{-r_f t}y(\lambda_t,S_t)dt$$

where $y(\lambda_t, S_t)$ is the expected flow profit defined in equation (4).

The firm's problem is to find the optimal algorithm, $S_t = S(\lambda_t)$, that maximizes the user's lifetime value. We can then rewrite the lifetime value of the user with a prior λ_0 as

$$V(\lambda_0) \equiv \max_{S(\lambda_t)} V(\{S(\lambda_t)\}, \lambda_0),$$

In the Appendix, we derive and solve the Hamilton-Jacobi-Bellman equation. Under the optimal algorithm, the firm's value function must satisfy

$$V(\lambda_t) = \frac{y(\lambda_t, S_t)}{r_f} + b\lambda_t^{-(\gamma-1)/2}(1-\lambda_t)^{(\gamma+1)/2}, \text{ where } \gamma = \sqrt{1 + \frac{8r_f\alpha(1-\alpha)}{(2\alpha-1)^2}} \tag{7}$$

for some coefficient b.

We describe the myopic recommendation algorithm and the forward-looking algorithm separately, as discussed in the introduction.

Myopic Recommendation Algorithm

Consider a firm that employs the myopic recommendation algorithm, which only aims to maximize the instantaneous profit. This case resembles a firm with a supervised learning algorithm that continuously predicts the likelihood that a user will enjoy each type of content, and simply recommends the one with the highest ranking.

The firm recommends the content type that maximizes the instantaneous payoff $y(\lambda_t, S_t)$ from equation (4). The optimal myopic algorithm is the following: the firm recommends content type N_1 if $\lambda_t > \lambda^* = \frac{c-(1-\alpha)}{2\alpha-1}$, type N_2 if $\lambda_t < 1-\lambda^*$, and type M if $\lambda_t \in [1-\lambda^*, \lambda^*]$. Intuitively, the firm recommends whichever content type that the user has the highest probability of liking at each time point.

Forward-Looking Recommendation Algorithm

Now we solve the optimal forward-looking algorithm, which needs to balance the exploration vs. exploitation trade-off. Due to symmetry, we only need to focus on the case of $\lambda_t > 0.5$. In this case, if the firm recommends niche-market content, it must recommend type N_1. Also, as noted earlier, once the firm recommends mass-market content to the user, it stops learning, so it will always recommend mass-market content in the future. Thus, the firm's value function when it recommends mass-market content must be c/r_f. So to obtain the optimal forward-looking algorithm, we only need to know at what point the firm switches from recommending niche-market content to recommending mass-market content. Let $\hat{\lambda}$ denote the cutoff such

that the firm begins serving mass-market content to the user if $\lambda_t \leq \hat{\lambda}$. The threshold $\hat{\lambda}$ must satisfy the value-matching and the smooth-pasting conditions (see, e.g., Dixit 1993) for the value function:

$$r_f V(\hat{\lambda}) = c,\ V'(\hat{\lambda}) = 0$$

Plugging these boundary conditions back into the solution of the HJB equation (7) produces the solution for $\hat{\lambda}$ and b_2.

We describe the optimal threshold and the firm's value function under the optimal threshold below:

Lemma 1 *Define*

$$\hat{\lambda} = \frac{[c-(1-\alpha)](\gamma-1)}{(2\alpha-1)(\gamma-1)+2(\alpha-c)}, \qquad \textit{where} \quad \gamma = \sqrt{1+\frac{8r_f\alpha(1-\alpha)}{(2\alpha-1)^2}}$$

1. *If $\hat{\lambda} > 0.5$, then the optimal forward-looking algorithm recommends content type N_1 if $\lambda_t > \hat{\lambda}$, type N_2 if $\lambda_t < 1-\hat{\lambda}$, and type M if $\lambda_t \in [1-\hat{\lambda}, \hat{\lambda}]$.*

   The firm's value function is symmetric around 0.5. For $\lambda_t > \hat{\lambda}$

   $$V(\lambda_t) = \frac{\lambda_t\alpha + (1-\lambda_t)(1-\alpha)}{r_f} + \frac{2(\alpha-c)}{r(\gamma-1)}\left(\frac{\hat{\lambda}}{1-\hat{\lambda}}\right)^{(\gamma+1)/2} \lambda_t^{-(\gamma-1)/2}(1-\lambda_t)^{(\gamma+1)/2}$$

   and $V(\lambda_t) = \frac{c}{r_f}$ for $0.5 \leq \lambda_t < \hat{\lambda}$.

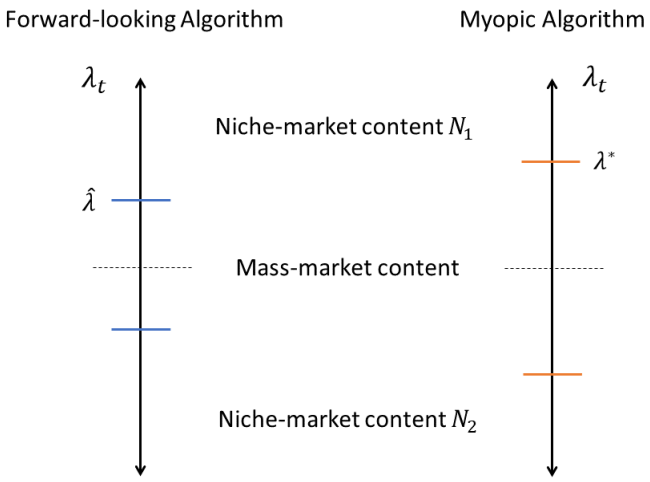
2. *If $\hat{\lambda} \leq 0.5$, then the optimal forward-looking algorithm recommends N_1 if $\lambda_t > 0.5$ and N_2 if $\lambda_t < 0.5$.*

   The firm's value function is symmetric around 0.5 where for $\lambda_t > 0.5$,

   $$V(\lambda_t) = \frac{\lambda_t\alpha + (1-\lambda_t)(1-\alpha)}{r_f} + \frac{2\alpha-1}{r_f\gamma}\left(\frac{\hat{\lambda}}{1-\hat{\lambda}}\right)^{(\gamma+1)/2} \lambda_t^{-(\gamma-1)/2}(1-\lambda_t)^{(\gamma+1)/2}$$

From Lemma 1 we have $\hat{\lambda} < \lambda^*$, which implies that the firm recommends more niche-market content under the forward-looking algorithm than under the myopic algorithm. Consider $\lambda_t \in (\hat{\lambda}, \lambda^*)$. The forward-looking algorithm recommends content type N_1, which is expected to generate lower instantaneous profit than type M, in order to gather more information about user i's preference. Figure 1 compares the decision under the forward-looking and the myopic algorithm.

Figure 1: Recommendations under the forward-looking vs. the myopic algorithm



Note that $\hat{\lambda}$ is an increasing function of γ, while γ is an increasing function of r_f and a decreasing function of α. Thus, $\hat{\lambda}$ increases with r_f and decreases with α. Intuitively, when r_f is smaller, the firm weights future profit more and thus it becomes more important for the firm to learn and adapt. Consequently, the firm is less likely to recommend mass-market content. When α is lower, it means that the user's preferences are less consistent and less correlated over time. It is more difficult to predict what a user likes at a given moment. Additionally, the firm also receives less precise information from the user's past behaviors. As a result, the firm recommends less niche-market content.

Proposition 1 *The optimal forward-looking threshold, $\hat{\lambda}$, is strictly lower than the myopic threshold, $\lambda^* = \frac{c-(1-\alpha)}{2\alpha-1}$. The optimal forward-looking threshold increases with discount rate, r_f, and decreases with preference consistency, α.*

3.3 Additional Value from the Forward-Looking Algorithm

Different recommendation algorithms require different levels of technology. An upgrade from the myopic algorithm to the forward-looking algorithm requires balancing the value from exploration and exploitation through techniques such as reinforcement learning. In this section, we examine the value of such a technological upgrade. This can also be interpreted as a monopoly's incentive to invest in such an upgrade if it is costly. We denote the firm's ex-ante expected profits under the myopic and forward-looking algorithms as $V^{MY}(\lambda_t)$ and $V^{FL}(\lambda_t)$, respectively.

Under the myopic algorithm, the firm switches from niche-market content to mass-market content when λ_t drops below λ^*, and makes a flow profit of c in perpetuity when serving mass-market content. To find the value function for $\lambda_t > \lambda^*$, we solve equation (7) with the boundary condition $r_f V(\lambda^*) = c$, from which we get $b = 0$. Thus, the firm's expected profit at $t = 0$ is

$$V^{MY}(\lambda_0) = \begin{cases} \frac{\lambda_0\alpha+(1-\lambda_0)(1-\alpha)}{r_f} & \text{for } \lambda_0 > \lambda^* \\ \frac{c}{r_f} & \text{for } \lambda_0 \in (1-\lambda^*, \lambda^*) \\ \frac{\lambda_0(1-\alpha)+(1-\lambda_0)\alpha}{r_f} & \text{for } \lambda_0 < 1-\lambda^* \end{cases} \tag{8}$$

The firm's expected profit at $t = 0$ under the forward-looking algorithm, $V^{FL}(\lambda_{it})$, is given in Lemma 1, from which we can get, for $\lambda_0 \geq \lambda^*$,

$$V^{FL}(\lambda_0) - V^{MY}(\lambda_0) = \frac{2(\alpha - c)}{r_f(\gamma - 1)} \left(\frac{\hat{\lambda}}{1-\hat{\lambda}} \right)^{(\gamma+1)/2} \lambda_0^{-(\gamma-1)/2} (1-\lambda_0)^{(\gamma+1)/2}$$

and for $\lambda_0 \in (\hat{\lambda}, \lambda^*)$,

$$V^{FL}(\lambda_0) - V^{MY}(\lambda_0) = \frac{\lambda_t\alpha + (1-\lambda_t)(1-\alpha) - c}{r_f} + \frac{2(\alpha - c)}{r_f(\gamma - 1)} \left(\frac{\hat{\lambda}}{1-\hat{\lambda}} \right)^{(\gamma+1)/2} \lambda_t^{-(\gamma-1)/2} (1-\lambda_t)^{(\gamma+1)/2}$$

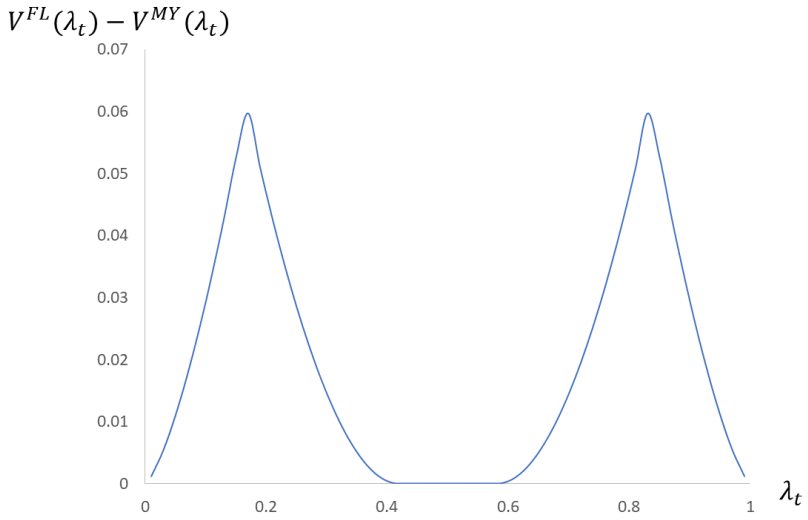
We call $V^{FL}(\lambda_0) - V^{MY}(\lambda_0)$ the additional value from the forward-looking algorithm.

Proposition 2 *The additional value from the forward-looking algorithm, $V^{FL}(\lambda_0) - V^{MY}(\lambda_0)$, is strictly positive for $\lambda_0 \notin [1-\hat{\lambda}, \hat{\lambda}]$. The value increases with λ_0 for $\lambda_0 < 1-\lambda^*$ and $0.5 < \lambda_0 < \lambda^*$, decreases with λ_0 for $1-\lambda^* < \lambda_0 < 0.5$ and $\lambda_0 > \lambda^*$, decreases with r_f, and increases with α.*

This result has implications for when firms should prioritize investing in a forward-looking algorithm. Proposition 2 shows that the benefit of the optimal algorithm is non-monotonic in the firm's knowledge about users. In Figure 2, we plot the additional value from the forward-looking algorithm as a function of λ_0. While it might be intuitive to think that learning is more important when the company knows less about customers' preferences, that is not always true. The less a company knows to begin with, the longer and more costly the learning process. Exploration requires tolerating a lower profit in the short term before enough information is gathered.

As λ increases, there are two effects. On the one hand, there will be less uncertainty, which decreases the additional value from the algorithm. On the other hand, the firm will suffer fewer losses in earlier periods, which increases the value for $\lambda < \lambda^*$. The additional value from the forward-looking algorithm peaks at $\lambda = \lambda^*$, which is the point at which the forward-looking algorithm and the myopic algorithm begin to diverge.

Figure 2: Additional value from the forward-looking algorithm



for $\alpha = 0.8$, $c = 0.7$ and $r = 0.6$

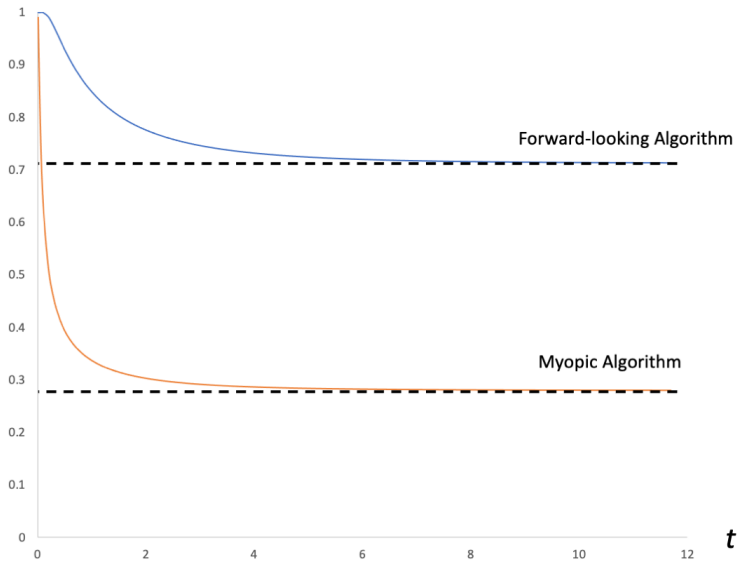
3.4 Evolution of Recommendations and Profit

As the firm learns about its users over time, how do the firm's beliefs evolve? What proportion of users are recommended mass-market versus niche-market content? In the Appendix, we solve for the population density of the firm's beliefs to characterize learning-induced user heterogeneity and describe the evolution of the firm's recommendations under the myopic and forward-looking algorithms. We present the results here.

Proposition 3 *Assume* $\lambda_0 \notin [1-\hat{\lambda}, \hat{\lambda}]$. *As* t *approaches infinity, the firm recommends niche-market content to* $\frac{\lambda_0 - \hat{\lambda}}{1-\hat{\lambda}}$ *fraction of users under the forward-looking algorithm, and* $\frac{\lambda_0 - \lambda^*}{1-\lambda^*}$ *fraction of users under the myopic algorithm. Both fractions decrease with* r_f *and increase with* α.

We illustrate the evolution of the fraction of users who are recommended niche-market content in Figure 3. Note that under both the forward-looking and the myopic algorithm, the fraction of users who are recommended niche-market content decreases and converges to a constant in the long-term steady state. This fraction decreases with discount rate r_f and increases with preference consistency α. Intuitively, with a bigger α, the firm cares more about learning users' preferences, and with a smaller r_f, the firm cares more about the long-term profit, so the steady-state amount of niche-market content increases. The forward-looking algorithm recommends more niche-market content both in the short-term

Figure 3: Fraction of Users Receiving Niche-Market Content



for $\lambda_0 = 0.88$, $\alpha = 0.8$, $c = 0.7$ and $r = 0.6$

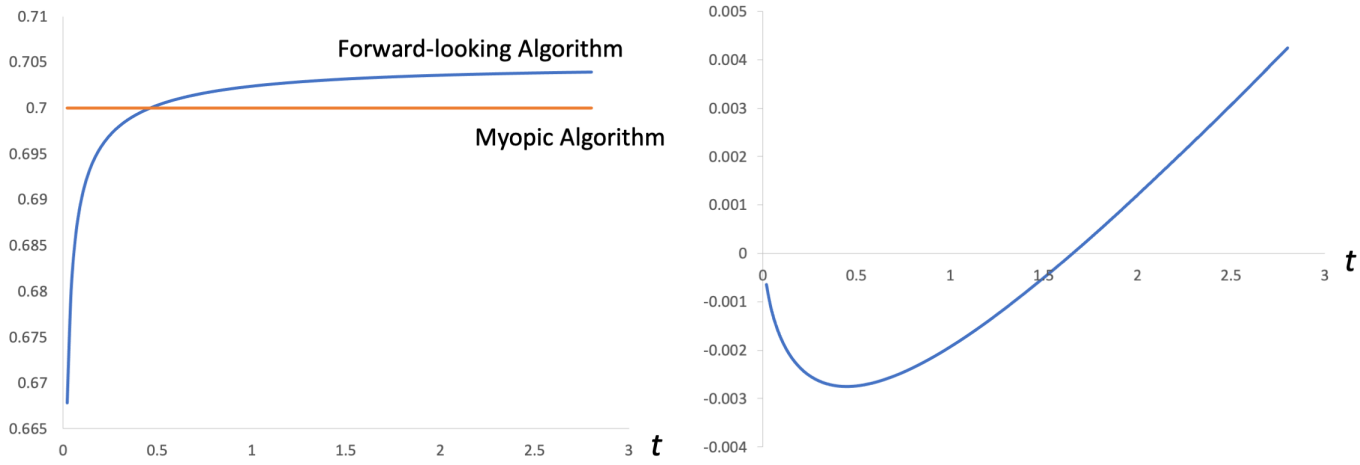
and in the long-term.

One can also compare the evolution of expected flow profit and expected cumulative profit over time under different algorithms. Compared to the myopic algorithm, the forward-looking algorithm may create lower profit in early periods. The flow profit under the forward-looking algorithm increases over time, which makes it more profitable than the myopic algorithm in the long run. When a company does not have much information about a customer's preference, e.g., when the customer is relatively new, the company should prioritize strategic experimentation to extract information from the customer's responses and expedite learning. However, doing so could lead to worse recommendations, lower user engagement, and lower profit in the near term. As companies upgrade their technology, it is important to recognize this implication and be prepared to tolerate worse performance in the near term. In Figure 4a, we plot the expected flow profit under different algorithms. In Figure 4b, we plot the the difference in expected cumulative profit between the forward-looking and myopic algorithms. The detailed derivation is presented in the Appendix.

4 Competition

In this section, we study firms' optimal recommendation algorithm under competition. We explore how competition affects the trade-off between exploration and exploitation, the ad-

Figure 4: Evolution of profit



(a) Expected profit flow (b) Difference in expected cumulative profit

for $\lambda_0 = 0.88$, $\alpha = 0.8$, $c = 0.7$ and $r = 0.6$

ditional value from the forward-looking algorithm, and the impact of different algorithms on consumer welfare. Table 2 presents the notations under competition.

We expand the monopoly model to allow for two firms, firm 1 and firm 2, competing for a multi-homing user's time. The user can consume content only from one of the two platforms at each t, but can switch back-and-forth over time without switching cost. At $t = 0$, both firms simultaneously choose their recommendation algorithms, which are functions mapping information sets to content types. Because the model is in continuous time and the user has no switching cost, we assume that the user can observe both firms' recommendations at each t for simplicity. The user optimally chooses to consume content from one of the two firms at each t. Both firms observe the user's choice, but cannot observe a user's flow utility when consuming content from the competitor. Thus, if the user does not visit firm j at time t, then firm j does not earn profit nor receive information about the user's preferences.

As in the monopoly model, each firm can recommend one of three types of content: mass-market content and two types of niche-market content. Niche-market content from the two firms is different, so that there are a total of four types of niche-market content. For simplicity, and to capture the fact that different platforms often carry different content, we assume the user's preferences for niche-market content types are independent between the two platforms.

We allow the user to have different values of α on the two platforms, so the user has different consistencies in her preferences for niche-market content on the two platforms. Let N_s^j denote niche-market content type s on firm j's platform. The user's expected flow utility

from seeing the content recommended by firm j at time t, $S_t^j \in \{M^j, N_1^j, N_2^j\}$, is

$$u(T^j, S_t^j) = \begin{cases} \alpha^j & \text{if} \quad S_t^j = T^j \\ c & \text{if} \quad S_t^j = M^j \\ 1 - \alpha^j & \text{otherwise} \end{cases} \tag{9}$$

where $T^j \in \{N_1^j, N_2^j\}$ represents the user's preferred type of niche-market content on firm j's platform.

As in the monopoly model, firm j's expected flow profit from serving content S_t^j to the user at time t (conditional on the user visiting firm j), denoted as $y^j(T^j, S_t^j)$, can be written as:

$$y^j(T^j, S_t^j) = \begin{cases} \alpha^j & \text{if} \quad S_t^j = T^j \\ c & \text{if} \quad S_t^j = M \\ 1 - \alpha^j & \text{otherwise} \end{cases} \tag{10}$$

The user has a discount rate of r_u, and both firms have a discount rate of r_f. We make no restrictions on the values of r_u and r_f, but our analysis focuses more on the case where users are less patient than the firms.[4]

Firm Learning

As in the monopoly model, at $t = 0$, firm j receives a binary signal on the user's preferred content type on its own platform with accuracy $\lambda_0^j > 0.5$. That is, firm j observes the correct user type with probability λ_0^j, and observes the incorrect type with probability $1 - \lambda_0^j$. We can always relabel the content types without loss of generality, so we can simply assume that firm j has a prior belief that the user prefers N_1^j with probability λ_0^j. Note that we allow $\lambda_0^1 \neq \lambda_0^2$, so firms can start with different amounts of information on the user's preferences. Each firm then updates its posterior belief λ_t^j in the same way as in the monopoly model.

User Learning

With competition, we need to model how the user learns, which then determines her platform choices. We assume that the user does not know her preferred type of niche-market content at firm j, T^j. She only observes her utility from the recommended content.

[4]This is motivated by the observation that online users often exhibit short attention spans and would quickly abandon sites or content that does not interest them, especially on mobile devices. A study by Google and Akamai finds that on e-commerce sites, a 100-millisecond delay in page load time decreases the conversion rate by 7% (Akamai 2017). A Facebook study finds that, on average, mobile users spend 1.7 seconds on each content, versus 2.5 seconds for desktop users (Facebook IQ 2016). Our model focuses on the case where users are less patient than firms, and examines how the two discount rates separately affect the equilibrium outcome.

Assumption 1 *The user does not know her preferred type of niche-market content,* T_t^j.

This assumption is needed to avoid unravelling of information. Consider the following case described in discrete time. The user visits firm 1 at time 0. If the user observes that firm 1 recommends the right niche-content type, then the user returns in the next period. If the user observes that firm 1 recommends the wrong type of niche-market content, then the user visits the competitor in the next period. Either way, the user's preferred content type is immediately revealed to firm 1, so there is no more learning. Such unravelling does not happen in reality for a few reasons. First, there are more than two types of content. Second, users may not know how much they like each type of content ex-ante. They also learn about their own preferences over time as they consume various types of content.

Under this assumption, the user has to update her belief on how well each firm's next recommendation will match her taste, given her past experiences with the firm. Intuitively, if a user enjoyed recent content recommended by Instagram Reels, then her expectation for the next recommendation from Instagram Reels increases. However, if Instagram Reels recommended multiple short videos that she did not enjoy, then she would have a lower expectation of the quality of the next recommendation, and may spend more time on TikTok instead.

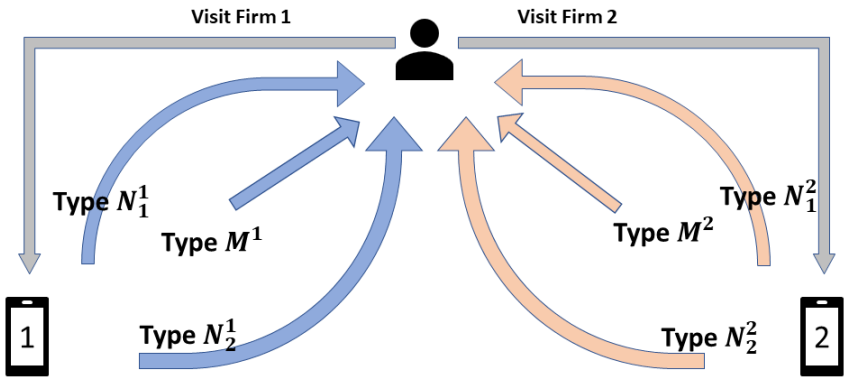
Given her knowledge of the game, the user has a prior belief of λ_0^j that firm j receives the correct signal on the user's preferences for niche-market content at time 0. The user updates her posterior belief that firm j infers her preferred type correctly. Because both the user and firm j receive the same feedback from each recommendation, the user's posterior belief is the same as firm j's posterior belief, λ_t^j.

Multi-Homing with Strategic Firms

In this multi-homing problem, each firm decides what content to recommend, while the user decides which firm to visit. While their decisions have to balance the trade-off between exploration and exploitation, they also have to factor in decisions made by the other agents. For example, a firm's decision to "explore" with niche-market content does not yield any information if a user chooses to visit the other firm, and the user's choice set also depends on what types of content the user expects each firm to recommend. Figure 5 gives an intuitive illustration of the setup.

All three players face a bandit problem where the return from each "arm" depends on the other two players' strategies. A solution has to simultaneously solve all three players' bandit problem. To solve the problem, we first characterize the user's optimal choice rule when presented with a menu of content. Then taking user behavior as given, we look

Figure 5: Multi-Homing with Strategic Firms



for Nash equilibrium at $t = 0$ when the two firms simultaneously choose recommendation algorithms. Finally, we confirm that the user's choice rule is optimal under the equilibrium recommendation strategies. We can then confirm that the equilibrium we characterize indeed solves all three players' dynamic optimization problem simultaneously.

4.1 Against a Mass-Market Content Provider

We first analyze a simpler case, where a focal strategic firm competes with a traditional content provider that only serves mass-market content. This can also be seen as adding an outside option to the monopoly model. However, even though the outside option is non-strategic, the focal firm still has to compete with it for the user's attention. We assume that firm 1 is the focal firm and firm 2 is the outside option.

First, let us consider the user's preferences between niche-market content from firm 1 and mass-market content. Suppose that the user can always choose between niche-market content from firm 1 and mass-market content. Because in our model the user and the firm receive the same utility from each recommended content, the user's problem is identical to the monopoly's problem of choosing between recommending niche-market content and mass-market content, but with a different discount rate.

The user's optimal threshold to switch from niche-market content from firm 1 to mass-market content, $\widehat{\lambda_u^1}$, can be adapted from Proposition 1.

$$\widehat{\lambda_u^1} = \frac{[c-(1-\alpha^1)](\gamma_u^1-1)}{(2\alpha^1-1)(\gamma_u^1-1)+2(\alpha^1-c)} \qquad \text{where} \quad \gamma_u^1 = \sqrt{1+\frac{8r_u\alpha^1(1-\alpha^1)}{(2\alpha^1-1)^2}} \tag{11}$$

The user prefers niche-market content from firm 1 for $\lambda_t^1 > \widehat{\lambda_u^1}$ or $\lambda_t^1 < 1-\widehat{\lambda_u^1}$, and prefers

mass-market content for $\lambda_t^1 \in [1-\widehat{\lambda_u^1}, \widehat{\lambda_u^1}]$.

If both firms offer mass-market content, we assume that the user will split her time equally between the two firms. Let $D_t \in \{0, \frac{1}{2}, 1\}$ denote the user's demand for firm 1 at time t. Thus, we have $D_t = \frac{1}{2}$ when firm 1 recommends mass-market content, and $D_t = \mathbb{I}\{\lambda_t^1 > \widehat{\lambda_u^1}\}$ when firm 1 recommends niche-market content.

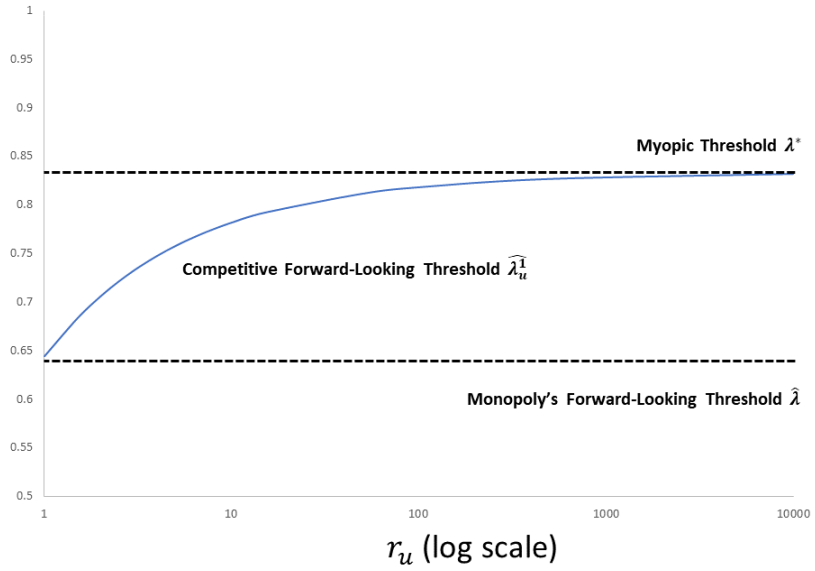
It is straightforward to show that firm 1's optimal forward-looking algorithm must follow the threshold $\widehat{\lambda_u^1}$. The firm switches from niche-market content to mass-market content as λ_t^1 drops below $\widehat{\lambda_u^1}$. To see this, note that the firm's profit from mass-market content is $\frac{1}{2}c$. Given our assumption that $c < 1$, we have $\lambda_t \alpha^1 + (1-\lambda_t)(1-\alpha^1) > \frac{1}{2}c$. That is, the expected flow profit from offering niche-market content is always higher than the flow profit from splitting demand with mass-market content. Thus, the firm has no incentive to offer mass-market content as long as the demand for niche-market content is positive. However, if $\lambda_t \in [1-\widehat{\lambda_u^1}, \widehat{\lambda_u^1}]$, then the user does not visit the firm offering niche-market content, so it has to recommend mass-market content.

Proposition 4 *When competing with a mass-market content provider, the firm recommends niche-market content if and only if* $\lambda_t \notin [1-\widehat{\lambda_u^1}, \widehat{\lambda_u^1}]$.

Note that the optimal threshold, $\widehat{\lambda}_u^1$, is optimal for the user. Giving the user an outside option forces the firm to adopt a recommendation algorithm that addresses the user's time preferences. If the user is less patient than firms, i.e. $r_u > r_f$, then $\widehat{\lambda}_u^1$ must be higher than the monopoly's optimal threshold, $\hat{\lambda}$. Thus, when there is an outside option, the firm recommends less niche-market content than a monopoly does, shifting away from exploration and toward exploitation. The firm has less room to experiment because users will switch to the competitor's platform if past recommendations do not sufficiently interest them. Conversely, if $r_u < r_f$, having an outside option forces the algorithm to shift toward exploration by recommending more niche-market content n.

Figure 6 shows firms' equilibrium threshold as a function of the user's discount rate r_u for the case where the user is less patient than firms, i.e., $r_u \geq r_f$, and compares them to the myopic threshold and the monopoly's forward-looking threshold derived from Section 3. Note that the optimal forward-looking algorithm when facing an outside option spans the continuum between the myopic algorithm and the monopoly's optimal algorithm.

Figure 6: The optimal threshold as a function of the user's discount rate



for $r_f = 1$, $c = 0.7$, and $\alpha^1 = 0.8$

4.2 Duopoly

Now we consider competition between two strategic firms that are both employing the forward-looking algorithm. We show that the same algorithm from Proposition 4 is optimal when the opponent is also strategic and forward-looking.

User's Behaviors

First, consider the user's dynamic preferences when her choices include niche-market content from firm j and mass-market content from the other firm. Because the user receives the same normalized utility from each recommended content as the firm, the user's problem is identical to the monopoly's problem of choosing between recommending niche-market content and mass-market content, but with a different discount rate. This is a bandit problem with a stopping option (mass-market content). Adapting Proposition 1, we can infer that the user's optimal content choice between niche-market content from firm j and mass-market content is marked by the threshold

$$\widehat{\lambda_u^j} = \frac{[c-(1-\alpha^j)](\gamma_u^j - 1)}{(2\alpha^j - 1)(\gamma_u^j - 1) + 2(\alpha^j - c)} \qquad \text{where} \quad \gamma_u^j = \sqrt{1 + \frac{8r_u\alpha^j(1-\alpha^j)}{(2\alpha^j-1)^2}} \tag{12}$$

The user prefers niche-market content from firm j for $\lambda_t^j > \widehat{\lambda_u^j}$ or $\lambda_t^j < 1 - \widehat{\lambda_u^j}$, and prefers mass-market content for $\lambda_t^j \in [1 - \widehat{\lambda_u^j}, \widehat{\lambda_u^j}]$.

Next, consider the user's dynamic preferences when her choices are comprised of niche-market content from two different firms. This is a two-armed bandit problem without a stopping option. The user's optimal policy is to consume from the firm with a higher Gittins index.[5] The Gittins index for niche-market content from firm j at a given λ_t^j, $G^j(\lambda_t^j)$, is equivalent to a fixed flow payoff such that she should switch from consuming firm j's niche-market content to consuming a fixed flow payoff of $G^j(\lambda_t^j)$ exactly at λ_t^j. We can thus use equation (12) to solve for $G^j(\lambda_t^j)$, by replacing c with $G^j(\lambda_t^j)$. We then have the following equation:

$$\lambda_t^j = \frac{[G^j(\lambda_t^j) - (1-\alpha^j)](\gamma_u^j - 1)}{(2\alpha^j - 1)(\gamma_u^j - 1) + 2[\alpha^j - G^j(\lambda_t^j)]} \tag{13}$$

which gives the Gittins index for firm j's niche-market content:

$$G^j(\lambda_t^j) = \frac{[\lambda_t^j(2\alpha^j - 1) + 1 - \alpha^j](\gamma_u^j - 1) + 2\alpha^j\lambda_t^j}{2\lambda_t^j + \gamma_u^j - 1} \tag{14}$$

Finally, the user must be indifferent between mass-market content from the two firms. If her choice set is only comprised of mass-market content from each firm, we assume that she evenly splits her time between the two firms.

For simpler notation, we let $S_t^j = M$ denote firm j recommending mass-market content to the user at time t, and let $S_t^j = N$ denote firm j recommending niche-market content to the user at time t.[6] Let $D_t^j(\lambda_t^1, \lambda_t^2 | S_t^1, S_t^2) \in \{0, \frac{1}{2}, 1\}$ denote the user's demand for firm j at time t. We can summarize D_t^1 as:

$$\begin{cases} D_t^1(\lambda_t^1, \lambda_t^2 | N, N) = \mathbb{I}\{G(\lambda_t^1) \geq G(\lambda_t^2)\} \\ D_t^1(\lambda_t^1, \lambda_t^2 | N, M) = \mathbb{I}\{\lambda_t^1 > \widehat{\lambda_u^j}\} \\ D_t^1(\lambda_t^1, \lambda_t^2 | M, N) = 1 - \mathbb{I}\{\lambda_t^2 > \widehat{\lambda_u^j}\} \\ D_t^1(\lambda_t^1, \lambda_t^2 | M, M) = \frac{1}{2} \end{cases} \tag{15}$$

whereas the user's demand for firm 2 at time t is $D_t^2 = 1 - D_t^1$.

Firms' Problem

Now we consider the firms' equilibrium recommendation algorithms at time 0, given that the user behaves as discussed above.

[5] See Bank and Küchler (2007) for a derivation of Gittins index theorem in continuous time.

[6] Because it is apparent which type of niche-market content firm j would choose given any λ_t^j, we drop the notation on the type of niche-market content.

Given the demand function, we can write firm j's expected flow profit from the user as $y^j(\lambda_t^j, S_t^j)D_t^j dt$. For a given pair of recommendation paths, $(\{S_t^1\}, \{S_t^2\})$, the expected lifetime value of the user for firm 1 is

$$V^1(\{S_t^1\} \mid \lambda_0^1, \lambda_0^2, \{S_t^2\}) = E\int_0^\infty e^{-r_f t} y^1(\lambda_t^1, S_t^1) D_t^1 dt$$

The firm's problem is to find the optimal path of content S_t^j to maximize the user's expected lifetime value. The expected lifetime value of the user to firm 1 with prior λ_0^1 is

$$V^1(\lambda_0^1) \equiv \max_{\{S_t^1\}} V^1(\{S_t^1\} \mid \lambda_0^1, \lambda_0^2, \{S_t^2\}),$$

Firm j's information about the user at time t can be written as $I_t^j = \{S_s^j, Y_s^j, D_s^j\}_{s<t}$, where S_s^j is firm j's recommendation at time $s < t$, Y_s^j is firm j's cumulative profit at time $s < t$, and D_s^j is an indicator function for whether the user visits firm j at time $s < t$. Firm j's recommendation algorithm is a function mapping each information set to a content type, denoted as $S^j(I_t^j) \in \{N_1^j, N_2^j, M\}$.

Equilibrium

In the monopoly case, the firm's optimal algorithm is characterized by a stationary function $S(\lambda_t)$. To facilitate direct comparisons with the monopoly model, we look for equilibrium in which firm j's recommendation algorithm is similarly characterized by a stationary function $S^j(I_t^j) = S^j(\lambda_t^j)$. Note that because the strategy does not depend on the user's state on the opponent's platform, in such an equilibrium, no matter what firm j's belief of the user's state on the competitor's platform is, there must be a weakly dominant action to take. Such an algorithm is appealing in practice because each firm only needs to infer a user's preferences for content on the firm's own platform.

Also, in practice, a firm can have different prior information on different users but applies the same recommendation algorithm to all users. Thus, we do not put assumption on the priors λ_0^j, and look for equilibrium strategy profiles that are robust to all possible priors.[7] Such an equilibrium exists and is unique. We describe the equilibrium strategy profile in the following proposition.

[7]Note that the set of equilibria could depend on the initial positions λ_0^1 and λ_0^2. There can be multiple equilibria that differ only on off-path nodes which have no impact on the equilibrium outcome. For example, suppose in equilibrium, both λ_0^j are low so both firms offer mass-market content immediately. Then one can construct an alternative equilibrium strategy that only differs for some higher value of λ_t^j. Because firms offer mass-market content so no learning occurs, we never reach such a state in equilibrium. We eliminate this trivial multiplicity by searching for equilibrium strategy $S^j(\lambda_t^j)$ that is invariant to λ_0^j's. That is, $S^1(\lambda_t^1)$ and $S^2(\lambda_t^2)$ constitute equilibrium regardless of what λ_0^1 and λ_0^2 are.

Proposition 5 *In a duopoly, firm j recommends niche-market content if and only if $\lambda_t^j \notin [1-\widehat{\lambda_u^j}, \widehat{\lambda_u^j}]$. This is the unique stationary algorithm, $S_t^j = S^j(\lambda_t^j)$, that constitutes equilibrium for all priors λ_0^1 and λ_0^2. Firms' equilibrium recommendations maximize the user's utility.*

Proposition 5 leads to several observations. First, comparing Proposition 5 to Proposition 4, we see that the equilibrium algorithm for each firm is the same as when it is competing against a non-strategic outside option. Second, note that the firms' equilibrium algorithms maximize the user's welfare, similar to when competing against an outside option.[8] Thus, when the user can go to multiple firms to consume content, regardless of each firm's strategic capability, the algorithm is now forced to solve the user's problem. The optimal exploration-exploitation trade-off under competition has to factor in the user's, instead of the firm's own, time preference.

Solving the user's problem means that the optimal trade-off between exploration and exploitation for firms facing competition is significantly different than for a monopoly. When r_u increases, users are more myopic in their content preferences. Consequently, a firm's choice of content has to be less forward-looking to prevent users from switching to the competitor. For example, consider a scenario where a user prefers mass-market content to niche-market content from either firm. A monopoly may still choose to offer niche-market content because the firm is more patient and values the information collected. However, if a competitor recommends mass-market content to the user, then the firm that recommends niche-market content will lose demand. If a user switches to the competitor's platform, then the firm can neither gather information nor profit from that user. This competitive pressure pushes firms to recommend content that caters to the user's time preferences. A monopoly can offer more niche-market content to extract future value from exploration, but when competing for users' attention, the value from exploration is muted.

Corollary 5.1 *In a duopoly, each firm recommends less (more) niche-market content than a monopoly would if the user has a higher (lower) discount rate than the firms. Firms offer less niche-market content as the user' discount rate increases. The optimal forward-looking algorithm under competition does not depend on firms' discount rate.*

[8] To see why this is true, consider an alternative problem where all three types of content (mass-market content, niche-market content from firm 1, and niche-market content from firm 2) are always available to the user. This is a classic multi-armed bandit problem for the user. Assuming both $\lambda_t^1 > 0.5$ and $\lambda_t^2 > 0.5$ WLOG, then by Gittins index theorem, the user's optimal content choice is niche-market from firm 1 if $G^1(\lambda_t^1) > \sup\{G^2(\lambda_t^2), c\}$, niche-market from firm 2 if $G^2(\lambda_t^2) > \sup\{G^1(\lambda_t^1), c\}$, and mass-market content if $c \geq \sup\{G^1(\lambda_t^1), G^2(\lambda_t^2)\}$. Note that this is exactly the user's content choice in equilibrium. In equilibrium, firm j offers niche-market content only when $\lambda_t^j > \widehat{\lambda_u^j}$ which is equivalent to $G^j(\lambda_t^j) > c$ by equations (12) and (13). Thus the user's most preferred content type is always available to her.

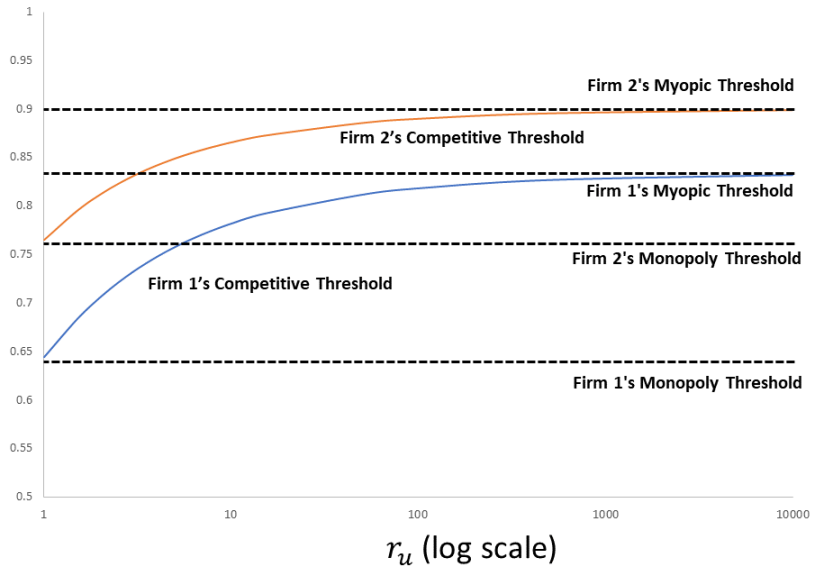
Research in computer science shows that many recommendation algorithms suffer from popularity bias, i.e., algorithms favor more popular, mainstream items over less popular, niche items (e.g., Abdollahpouri et al. 2017, Abdollahpouri et al. 2019). Martinez (2021) argues that such bias causes high quality but niche movies on Netflix to be under-recommended. Music streaming services, which heavily utilize recommendation algorithms, have not driven the audience "from the mainstream into the niche" as many have hoped (Blake, 2020). While such bias is mostly viewed as a problem to be solved, our model shows that the degree to which a recommendation algorithm should favor mainstream content over niche content depends on competitive pressure. Assuming users are less patient than firms, we predict that firms in a more competitive market, such as music streaming, would recommend more mainstream content than firms with more market power, such as YouTube. Similarly, we predict that when more competitors enter a market, such as when Facebook, Instagram, and YouTube entered the short video market to compete with TikTok, the incumbent's algorithm should move towards recommending more popular, mainstream content.

Second, the firm with a higher α^j has a greater ability to learn the user's preferences, because the noise in the user's response, σ_t (from equation 2), decreases in α. The fact that $\alpha^1 \geq \alpha^2$ implies $\widehat{\lambda_u^1} \leq \widehat{\lambda_u^2}$, so the firm with a higher α^j recommends more niche-market content. Intuitively, having a higher α^j means that the user's behavior is less noisy, which facilitates faster learning. A higher α^j also means higher profit from serving niche-market content. Both factors encourage the firm with a higher α^j to recommend more niche-market content even when the user dislikes previous recommendations.

Corollary 5.2 *A firm recommends more niche-market content than its competitor if user preferences are more consistent on its platform, i.e.,* $\widehat{\lambda_u^1} \leq \widehat{\lambda_u^2}$ *if and only if* $\alpha^1 \geq \alpha^2$.

Figure 7 shows both firms' equilibrium thresholds as functions of the user's discount rate r_u in the case of $r_u \geq r_f$. Each firm's optimal recommendation algorithm falls between its monopoly algorithm and its myopic algorithm. From equation (12), we can see that, as $r_u \to \infty$, $\widehat{\lambda_u^j}$ approaches the myopic threshold, λ^*. As $r_u \to r_f^+$, $\widehat{\lambda_u^j}$ approaches the monopoly's forward-looking threshold, $\widehat{\lambda}$. Thus, when users are myopic, the myopic algorithm itself is optimal. This implies that, when facing myopic users, the industry profit when firms use the forward-looking algorithm is the same as when firms only use the myopic algorithm. Referring to the difference between the industry profit under the forward-looking algorithm and that under the myopic algorithm as *the additional industry return from the forward-looking algorithm*, we can make the following statement.

Figure 7: Optimal thresholds as functions of the users' discount rate



for $r_f = 1$, $c = 0.7$, $\alpha^1 = 0.8$, and $\alpha^2 = 0.75$

Corollary 5.3 *As $r_u \to \infty$, the equilibrium forward-looking algorithm under competition converges to the myopic algorithm, and the additional industry return from the forward-looking algorithm converges to zero.*

Comparing Corollary 5.3 to Proposition 2, we see that the presence of a competitor decreases the value from the forward-looking algorithm if the user's discount rate is sufficiently high. This also implies that competition must lower firms' incentives to invest in the technological upgrade from the myopic algorithm to the forward-looking algorithm when users are sufficiently impatient. We examine this investment decision more closely in Section 5.

Because firms' equilibrium algorithms maximize the user's utility, the impact on user welfare of both firms upgrading their algorithms from myopic to forward-looking must be positive. However, note that the monopoly threshold, $\hat{\lambda}$, does not depend on the user's time preferences. For $r_u > r_f$, a monopoly's forward-looking algorithm recommends too much niche-market content with respect to user welfare, while a monopoly's myopic algorithm recommends too little. In the limit as $r_u \to \infty$, the myopic algorithm maximizes the user's utility. Thus there must exists a threshold on r_u such that, if the user's discount rate is above the threshold, the monopoly's technological upgrade from the myopic algorithm to the forward-looking algorithm actually decreases user welfare. On the other hand, when $r_u \to r_f$, the user's and the monopoly's time preferences align, so the monopoly's forward-looking algorithm maximizes the user's utility. Thus the development of the forward-looking

algorithm may lower user welfare under monopoly, but it always benefits the user under competition.

Corollary 5.4 *Under monopoly, for r_u sufficiently high, the user's welfare is lower when the firm uses the forward-looking algorithm than when the firm uses the myopic algorithm. Under duopoly, the user's welfare is always higher when firms use the forward-looking algorithm.*

We summarize our main findings conceptually in Table 3.

Our discussion suggests that there may exist a trade-off between technology adoption and consumer welfare. When users are sufficiently impatient compared to firms, a firm with monopolistic power derives more value from the forward-looking algorithm, and thus has a higher incentive to develop the forward-looking algorithm, but such technology hurts users; in contrast, firms under competition derive less value from the forward-looking algorithm, thus may have less incentives to develop the forward-looking algorithm, even though such technology is beneficial to users. However, whether this trade-off exists depends on the user's discount rate. If $r_u = r_f$, then a monopoly's adoption of the forward-looking algorithm must also increase user welfare. The following section formally investigates firms' incentives to invest in the forward-looking algorithm.

Table 3: Impact of Competition When Users Are Sufficiently Impatient

	Optimal level of exploration	Industry return from algorithmic upgrade	Effect of upgrade on user welfare
Monopoly	High	High	Negative
Duopoly	Low	Low	Positive

5 Investment in Algorithmic Upgrades

In the previous section, firms' algorithmic capability is exogenous. Should firms invest in a technological upgrade from the myopic algorithm to the forward-looking algorithm? We consider an extended game with investment decisions. At the beginning of the game, both firms first decide whether to develop the forward-looking algorithm for a fixed cost K. If a firm invests, it proceeds with the forward-looking algorithm. If a firm does not invest, it proceeds with the myopic algorithm. The two firms compete for a user in the same fashion as in Section 4.

Optimal Strategy Against the Myopic Algorithm

To analyze this investment game, we first have to complete our analysis by considering what happens if firms have asymmetric technologies. We label the forward-looking firm as firm 1 and the myopic firm as firm 2.

As before, we focus on the optimal stationary algorithm that can be characterized by a stationary function $S^1(I_t^1) = S(\lambda_t^1)$. We look for stationary strategy profiles that are robust to all possible priors. One can show that the equilibrium strategy under symmetric duopoly from Proposition 5 is also the optimal strategy against the myopic algorithm.

Proposition 6 *When competing against the myopic algorithm, firm* 1 *recommends niche-market content if and only if* $\lambda_t^1 \notin [1 - \widehat{\lambda_u^1}, \widehat{\lambda_u^1}]$. *This is the unique stationary algorithm that is optimal for all priors* λ_0^1 *and* λ_0^2.

The full proof is presented in the Appendix. The proof involves three steps. First, we solve for user behavior. We show that when both firms offer niche-market content, the user visits the firm with the higher λ. When the myopic firm offers mass-market content, the user prefers niche-market content from the forward-looking firm 1 if and only if $\lambda_t^1 \notin [1 - \widehat{\lambda_u^1}, \widehat{\lambda_u^1}]$. In the second step, we show that if firm 1 can observe the user's state at firm 2, λ_t^2, then the algorithm described in Proposition 4 and 5 is optimal. Because the algorithm does not depend on λ_t^2, it must also be optimal when firm 1 cannot observe λ_t^2. Finally, we prove that any other stationary algorithm must be sub-optimal for some priors.

While the optimal algorithm is the same as when both firms are forward-looking, the user's switching behavior is different. For $\lambda_t^1 > \widehat{\lambda_u^1}$ and $\lambda_t^2 \in (\widehat{\lambda_u^2}, \lambda^*)$, the user always prefers firm 1 when firm 2 is myopic, whereas the user prefers firm 1 if and only if $G^1(\lambda_t^1) > G^2(\lambda_t^2)$ when firm 2 is forward-looking. Thus the firm with the forward-looking algorithm is expected to receive more demand.

Similar to Corollary 5.3, Proposition 6 also implies that as $r_u \to \infty$, the optimal forward-looking algorithm approaches the myopic algorithm. So the additional value of having the forward-looking algorithm approaches 0 as the user becomes myopic.

Investment Equilibrium

To analyze the firms' decisions to invest in an upgrade from the myopic algorithm to the forward-looking algorithm, we also need to know their expected payoffs from the symmetric and asymmetric equilibrium. However, it is difficult to calculate expected payoffs analytically in our model. Instead, we analyze their investment decisions numerically by simulating firms' payoffs from the subsequent game.

We simulate the game in discrete time with steps of size $dt = 0.05$ for a total length of $T = 2000$. The parameters used are $\lambda_0^1 = \lambda_0^2 = 0.7$, $\alpha^1 = \alpha^2 = 0.75$, $c = 0.6$, and $r_f = 0.01$. At different levels of r_u, we simulate the game 1000 times each and compute the average payoff for each firm over the 1000 simulations. We present the results in Table 4. Let $V^j(TN_1, TN_2)$ denote firm j's expected payoff when firm 1's technology is $TN_1 \in \{FL, MY\}$ and firm 2's technology is $TN_2 \in \{FL, MY\}$, with FL standing for forward-looking and MY standing for myopic. As expected, when both firms are forward-looking, their payoffs decrease as the user becomes more myopic. When technology is asymmetric, there is an advantage to having the forward-looking algorithm, but the size of the advantage decreases as r_u increases.

Table 4: Simulated Firm Payoffs

	$r_u = 1$	$r_u = 3$	$r_u = 10$	$r_u = 30$	$r_u = 100$	$r_u = 300$	$r_u = 1000$
$V^j(FL, FL)$	3.570	3.526	3.448	3.363	3.292	3.249	3.222
$V^1(FL, MY)$	5.874	5.375	4.795	4.364	4.052	3.885	3.786
$V^2(FL, MY)$	0.968	1.341	1.753	2.053	2.268	2.383	2.450
$V^j(MY, MY)$	2.994	2.994	2.994	2.994	2.994	2.994	2.994

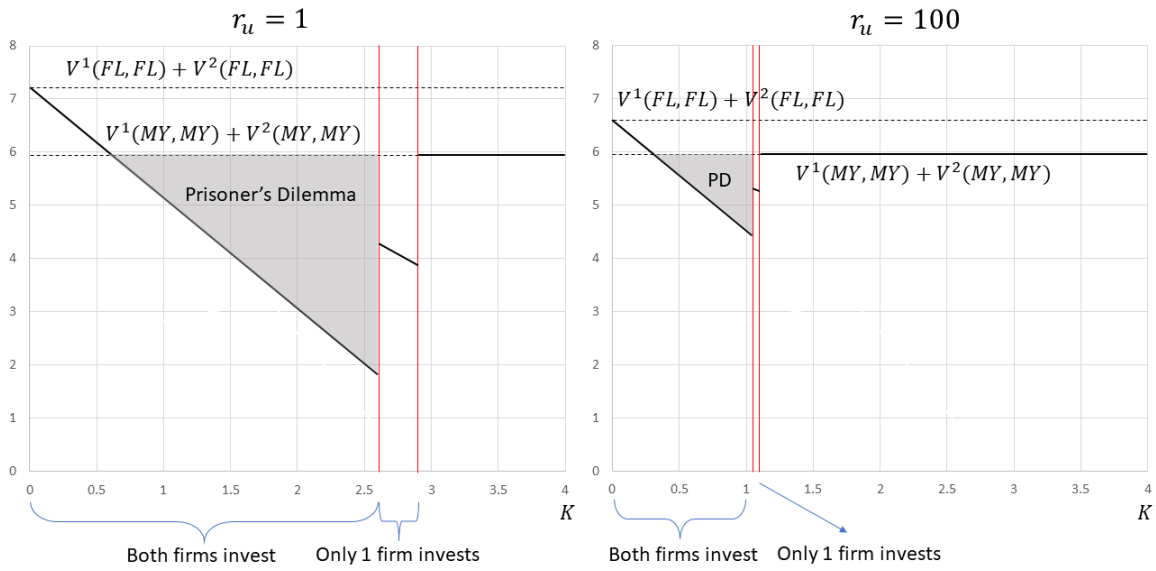
Depending on the cost, K, there can be 0, 1, or 2 firms investing in equilibrium. If $V^j(MY, MY) \geq V^1(FL, MY) - K$, then there exists an equilibrium where neither firm invests in the forward-looking algorithm. If $V^j(FL, FL) - K \geq V^2(FL, MY)$, then there exists an equilibrium where both firms invest. Otherwise, there exists an equilibrium where only one firm invests.

Figure 8 depicts equilibrium investment decisions and the firms' payoffs net of investment cost for different values of K and r_u. The equilibrium is described by two thresholds on K. When K is high, neither firm invests in equilibrium, and the industry profit is characterized by $V^1(MY, MY) + V^2(MY, MY)$. For an intermediate range of K, only one firm invests, and the industry profit is $V^1(FL, MY) + V^2(FL, MY) - K$. Finally, for lower values of K, both firms invest, and the industry profit is $V^1(FL, FL) + V^2(FL, FL) - 2K$.

When r_u is higher, the additional value of the forward-looking algorithm decreases. As a result, both thresholds move toward 0, so that firms require lower costs in order to invest. A higher degree of consumer myopia discourages competing firms from investing in the algorithmic upgrade.

Note that as K decreases and firms begin to invest in the upgrade, each investment leads to a drop in total industry profit. There exists an intermediate region of K that exhibits prisoner's dilemma, where both firms invest in the forward-looking algorithm and receive net payoffs lower than what they would receive if both used the myopic algorithm. Having a higher investment cost, K, can benefit the firms if it moves the equilibrium out of the

Figure 8: Industry Profit Net of Investment Costs



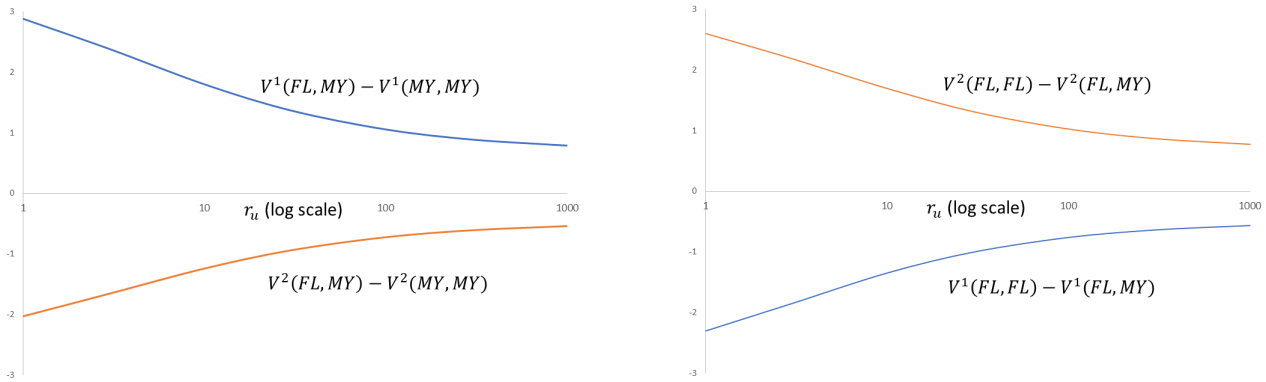
for $\lambda_0^1 = \lambda_0^2 = 0.7$, $\alpha^1 = \alpha^2 = 0.75$, $c = 0.6$, and $r_f = 0.01$

prisoner's dilemma region.

To get intuition on why the prisoner's dilemma arises, we examine sequentially the effects of each firm's investment on both firms' expected payoffs. Figure 9a plots $V^1(FL, MY) - V^1(MY, MY)$ and $V^2(FL, MY) - V^2(MY, MY)$, i.e., the changes in firm 1's and firm 2's payoffs when only firm 1 upgrades to the forward-looking algorithm, for different r_u. There are two benefits to firm 1 when firm 1 upgrades its algorithm. First, by providing better content than its competitor, it steals demand from firm 2, causing firm 1's payoff to increase and firm 2's payoff to decrease. Second, the forward-looking algorithm allows for better learning of user preferences, creating a higher social surplus. This is reflected in the fact that $|V^1(FL, MY) - V^1(MY, MY)| > |V^2(FL, MY) - V^2(MY, MY)|$. Because of the demand-stealing effect, firm 1 places a negative externality on firm 2 when firm 1 upgrades its algorithm. If $K > |V^1(FL, MY) - V^1(MY, MY)| - |V^2(FL, MY) - V^2(MY, MY)|$, then firm 1's investment leads to a decrease in the industry profit.

Figure 9b plots $V^1(FL, FL) - V^1(FL, MY)$ and $V^2(FL, FL) - V^2(FL, MY)$, i.e., the changes to firm 1's and firm 2's payoffs when firm 2 also upgrades to the forward-looking algorithm. Similarly, firm 2 takes demand away from firm 1 when firm 2 upgrades its algorithm. For K not too high, both firms upgrade to the forward-looking algorithm in equilibrium. Because firms do not factor in demand externality in their investment decisions, the total industry profit is lower when both firms invest compared to when neither firm

Figure 9: Effect of Investment on Profit



(a) Only firm 1 invests

(b) Firm 2 also invests

for $\lambda_0^1 = \lambda_0^2 = 0.7$, $\alpha^1 = \alpha^2 = 0.75$, $c = 0.6$ and $r_f = 0.01$

invests, unless the investment cost K is sufficiently small.

The effect of consumer impatience on the equilibrium profit is more nuanced. Locally, a higher r_u decreases the return from having the forward-looking algorithm, as shown in Corollary 5.3. Thus, the impact of the upgrade on the firms' payoffs becomes smaller. However, a higher r_u also lowers the investment thresholds. Thus, a higher degree of consumer myopia can benefit both firms by helping them to avoid the prisoner's dilemma.

Note that as $r_u \to \infty$, there must be no incentive for either firm to invest, as the myopic algorithm is itself optimal. Thus for any $K > 0$, the equilibrium must be no investment if the user's discount rate is sufficiently high. This confirms our previous finding that competition lowers the firms' incentives to adopt the forward-looking algorithm if users are sufficiently impatient in their content consumption behaviors.

6 Conclusion

The increased use of AI and machine learning has dramatically changed marketing practices. Interactions between firms and consumers are increasingly frequent, personalized, and automated. However, while extensive research has focused on techniques that enable real-time collection of customer data and dynamic, personalized interventions based on such data, less attention has been paid to the strategic implications of such algorithmic capability, especially under competition.

In this paper, we study the competition between content recommendation algorithms as a representative setting of adaptive, algorithm-based personalization. We formulate con-

tent recommendation algorithms as solutions to a stochastic dynamic programming problem under demand uncertainty. Firms compete for the attention of a multi-homing user who switches between firms over time. We compare the simpler, myopic recommendation algorithm to the more advanced, forward-looking algorithm, and investigate how the additional value from the more advanced algorithm varies across different competitive scenarios. The model allows us to address three questions. How does competition affect the optimal exploration vs. exploitation trade-off? How does competition affect firms' incentives to invest in more advanced algorithms? And how do such technological upgrades affect consumer welfare?

We find that the optimal recommendation algorithm under competition should cater to users' time preferences in order to prevent users from switching. If users are less patient than firms, competition shifts the optimal algorithm away from exploration and towards exploitation, as firms under competition recommend less niche-market content and learn less information about user preferences. For impatient users, competition reduces firms' incentive to invest in the optimal forward-looking algorithm. On the other hand, upgrading from the myopic algorithm to the optimal algorithm always improves user welfare under competition, while hurting user welfare under monopoly when users are sufficiently impatient. There exists a prisoner's dilemma when firms endogenously decide whether to invest in algorithmic upgrades. In such a case, a higher cost of investment or a higher degree of user myopia can raise industry profit by avoiding the prisoner's dilemma.

Our findings add new perspectives to the discussion on the regulation of algorithm-driven tech firms. In October 2020, the U.S. Congress released a report on competition in digital markets including search, e-commerce, social media, and digital advertising. The report suggests that the concentrated market power held by these firms has led to less innovation as well as lower service quality by deterring entrepreneurs from entering the market (U.S. House, 2020, pp. 46-56). Earlier in 2020, the UK's antitrust agency published a similar report on online platforms, expressing concern that the market power held by Google and Facebook in their consumer-facing markets hampers innovation and lowers service quality (Competition and Markets Authority, 2020, pp. 310-313). Both reports also warned against over-collection of consumer data as a consequence of market power.

Our study presents a complementary view on the effects of market power on innovation and consumer welfare. While our study echoes the concern that market power leads to over-learning and lower service quality, we show how competition could potentially discourage the development and adoption of more advanced learning algorithms. According to OECD, by the end of 2019, 50 countries (including the European Union) "have launched, or have plans to launch, national AI strategies" (Berryhill et al. 2020). For policy makers in these

countries, it is important to understand how the market structure affects the development of AI capacities in a global race, while balancing AI development with other factors such as consumer protection, privacy, and general entrepreneurship in digital markets. Additionally, our results imply that having more forward-looking users is beneficial to all parties: firms under competition have a higher incentive to develop better algorithms and users enjoy better recommendations. Thus educating consumers about how their current actions affect their future experiences on online platforms can be a mutually beneficial intervention for policy makers and competing firms.

Several limitations of this paper open avenues for future research. First, we focus on the specific context of content recommendation. Many other marketing decisions such as pricing, coupon distribution, advertising campaigns, service assignments, and product recommendations are also solutions to a stochastic dynamic programming problem under demand uncertainty, in which the firm needs to learn about consumer preferences and trade off short-term cost with future payoff in order to maximize long-term profit. While the key insights revealed from our analysis have general implications, these contexts also have specific attributes that warrant more focused examination. Second, our model abstracts away important elements of competition, including pricing and differentiation. These strategic decisions can have interactions that significantly alter the optimal algorithm and its value. Our model also focuses on the hypothetical optimal algorithm instead of algorithms that are used in practice. Third, in this model, learning is only useful for improving recommendation for the focal user. We do not consider other benefits of data, such as selling information to third parties, or learning of preferences across similar consumers.

Our stylized model only features two user types and three content types. It would be interesting to study a more general distribution of preferences and choices. Future research may also consider more than two firms or when a user's preferences are correlated across firms. Finally, learning in our model is symmetric between a user and the firm she visits (although asymmetric between firms). Adding private information to users significantly complicates the problem, but is nonetheless an important direction for future research.

Table 1: Notation for the Monopoly Model

Variable	Description
t	time
M	mass-market content
N_1, N_2	niche-market content
T	the user's preferred niche-market content type
α	the consistency of the user's preference for niche-market content
c	the probability the user likes a mass-market content
S_t	the firm's choice of content type at time t
$y(T, S_t)$	the probability that a user of type T likes content type S_t
σ_t	standard deviation of the flow profit
$Y_t(x)$	cumulative profit
λ_t	belief that the user prefers N_1 over N_2
$y(\lambda_t, S_t)$	the firm's expected flow profit from recommending S_t at time t
$\sigma(\lambda_t)$	instantaneous standard deviation of λ_t
r_f	the firm's discount rate
$V(\lambda_t)$	the firm's value function at state λ_t
λ^*	the optimal threshold for the myopic algorithm
$\hat{\lambda}$	the optimal threshold for the forward-looking algorithm
γ	$\sqrt{1+\frac{8r_f\alpha(1-\alpha)}{(2\alpha-1)^2}}$
FL	the forward-looking algorithm
MY	the myopic algorithm

References

[1] Abdollahpouri H, Burke R, Mobasher B (2017). Controlling popularity bias in learning-to-rank recommendation. *Proceedings of the eleventh ACM conference on recommender systems*, 42-46.

[2] Abdollahpouri H, Burke R, Mobasher B (2019). Managing popularity bias in recommender systems with personalized re-ranking. *The thirty-second international flairs conference.*

[3] Acquisti A, Taylor C, Wagman L (2016). The economics of privacy. *Journal of Economic Literature*, 54(2), 442-492.

[4] Acquisti A, Varian HR (2005). Conditioning prices on purchase history. *Marketing Science*, 24(3), 367-381.

[5] Agarwal D, Chen BC, Elango P (2008). Explore/exploit schemes for web content optimization. *Yahoo Research paper series.*

[6] Aghion P, Bloom N, Blundell R, Griffith R, Howitt P (2005). Competition and Innovation: An Inverted U Relationship. *The Quarterly Journal of Economics*, 120, 701-728.

Table 2: Notation for the Competition Model

Variable	Description
t	time
M^j	mass-market content from firm j
N_1^j, N_2^j	niche-market content from firm j
T^j	the user's preferred niche-market content type on firm j's platform
α^j	the consistency of the user's preference for niche-market content on firm j' platform
c	the probability the user likes a mass-market content
S_t^j	firm j's choice of content type at time t
$u(T^j, S_t^j)$	the type T_j user's expected flow utility from content S_t^j
$y^j(T^j, S_t^j)$	the probability that a user of type T^j likes content type S_t^j
λ_t^j	firm j's belief that the user prefers N_1^j over N_2^j
$G^j(\lambda_t^j)$	the Gittins index for firm j's niche-market content at state λ_t^j
$y^j(\lambda_t^j, S_t^j)$	firm j's expected flow profit from recommending S_t^j at time t if the user visits firm j
r_f	firms' discount rate
r_u	the user's discount rate
$V^j(\lambda_t^j)$	firm j's value function at state λ_t^j
I_t^j	firm j's information set at time j
λ^*	the optimal threshold for the myopic algorithm
$\widehat{\lambda_u^j}$	the optimal threshold for the user on firm j's platform
D_t^j	the user's demand for firm j at time t
γ_u^j	$\sqrt{1+\frac{8r_u\alpha^j(1-\alpha^j)}{(2\alpha^j-1)^2}}$
TN_j	firm j's technology
FL	the forward-looking algorithm
MY	the myopic algorithm
$V^j(TN_1, TN_2)$	firm j's expected profit when firm 1 uses TN_1 and firm 2 uses TN_2
K	the cost of upgrading from the myopic algorithm to the forward-looking algorithm

[7] Agrawal A, Gans J, Goldfarb A (2018a). *Prediction machines: the simple economics of artificial intelligence*. Harvard Business Press.

[8] Agrawal A, Gans J, Goldfarb A (2018b). Human Judgment and AI Pricing. *AEA Papers and Proceedings*, 108, 58-63.

[9] Agrawal A, Gans J, Goldfarb A (2019). Exploring the impact of artificial intelligence: Prediction versus judgment. *Information Economics and Policy*, 47, 1-6.

[10] Akamai (2017). State of Online Retail Performance - 2017 Holiday Retrospective. Accessed online on October 11, 2020 at https://www.akamai.com/us/en/multimedia/documents/report/akamai-state-of-online-retail-performance-2017-holiday.pdf

[11] Aridor G, Mansour Y, Slivkins A, Wu ZS (2021). Competing bandits: The perils of exploration under competition. *Working paper*, arXiv preprint arXiv:2007.10144.

[12] Asker J, Fershtman C, Pakes A (2022). The Impact of AI Design on Pricing. *Working Paper*.

[13] Athey S, Bryan K, Gans J (2020). The Allocation of Decision Authority to Human and Artificial Intelligence. *AEA Papers and Proceedings*, 110, 80-84.

[14] Athey S, Imbens GW (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, 11, 685-725.

[15] Banchio M, Mantegazza (2022). Adaptive Algorithms and Collusion Via Coupling. *Working Paper*.

[16] Banchio M, Skrzypacz A (2022). Artificial Intelligence and Auction Design. *Working Paper*.

[17] Bank P, Küchler C (2007). On Gittins index theorem in continuous time. *Stochastic Processes and Their Applications*, 117(9), 1357-1371.

[18] Bergemann D, Välimäki J (1996). Learning and Strategic Pricing. *Econometrica*, 64, 1125-49.

[19] Berman R, Katona Z (2020). Curation Algorithms and Filter Bubbles in Social Networks. *Marketing Science*, 39(2), 296-316.

[20] Berryhill J, Heang KK, Clogher R, McBride K (2020). Hello, World: Artificial intelligence and its use in the public sector. *OECD Observatory of Public Sector Innovation*. Available online at oecd-opsi.org.

[21] Blake E (2020) Data Shows 90 Percent of Streams Go to the Top 1 Percent of Artists. *Rolling Stone*. Accessed on 11/24/2022 at https://www.rollingstone.com/pro/news/top-1-percent-streaming-1055005/.

[22] Bolton P, Harris C (1999). Strategic Experimentation. *Econometrica*, 67, 349-374.

[23] Branco F, Sun M, Villas-Boas JM (2012). Optimal Search for Product Information *Management Science*, 58(11), 2037-2056.

[24] Calvano E, Calzolari G, Denicolò V, Pastorello S (2020). Artificial intelligence, algorithmic pricing and collusion. *American Economic Review*, 110(10), 3267-3297.

[25] Chen M, Beutel A, Covington P, Jain S, Belletti F, Chi EH (2019). Top-k off-policy correction for a REINFORCE recommender system. *In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, 456-464.

[26] Chen Y, Narasimhan C, Zhang ZJ (2001). Individual marketing with imperfect targetability. *Marketing Science*, 20(1), 23-41.

[27] Ching A (2010). A Dynamic Oligopoly Structural Model for the Prescription Drug Market After Patent Expiration. *International Economic Reviews*, 51, 1175-1207.

[28] Chintagunta P, Hanssens DM, Hauser JR (2016). Editorial-Marketing Science and Big Data. *Marketing Science*, 35(3), 341-342.

[29] Dasgupta P, Stiglitz, J (1980). Industrial Structure and the Nature of Innovative Activity. *The Economic Journal*, 90, 266-293.

[30] Deb J, Öry A, Williams K (2018). Aiming for the Goal: Contribution Dynamics of Crowdfunding. *Working paper*.

[31] Dogan M, Jacquillat A, Yildirim P (2021). Strategic Automation and Decision-Making Authority. *Working Paper*.

[32] Facebook IQ (2016). Capturing Attention in Feed: The Science Behind Effective Video Creative. Accessed online on October 11, 2020 at https://www.facebook.com/business/news/insights/capturing-attention-feed-video-creative

[33] Fader PS, Winer RS (2012). Introduction to the special issue on the emergence and impact of user-generated content. *Marketing Science*, 31(3), 369-371.

[34] Felli L, Harris C (1996). Job Matching, Learning and Firm-Specific Human Capital. *Journal of Political Economy*, 104, 838-868.

[35] Fudenberg D, Strack P, Strzalecki T (2018). Speed, Accuracy, and the Optimal Timing of Choices. *American Economic Review*, 108(12), 3651-84.

[36] Fudenberg D, Tirole J (2000). Customer poaching and brand switching. *RAND Journal of Economics*, 31(4), 634-657.

[37] Godes D, Mayzlin D (2004). Using online conversations to study word-of-mouth communication. *Marketing science*, 23(4), 545-560.

[38] Gonul F, Shi M (1998). Optimal Mailing of Catalogs: A New Methodology Using Estimable Structural Dynamic Programming Models. *Management Science*, 44(9).

[39] Google Developers (2020). Recommendation Systems. Accessed online on November 11, 2020 at https://developers.google.com/machine-learning/recommendation.

[40] Hansen K, Misra K, Pai M (2020). Algorithmic Collusion: Supra-Competitive Prices via Independent Algorithms. *Marketing Science*, 40(1), 1-12.

[41] Hauser JR, Liberali G, Urban GL (2014). Website morphing 2.0: Technical and implementation advances and a field experiment. *Management Science*, 60(6), 1594-1616.

[42] Huang MH, Rust RT (2018). Artificial Intelligence in Service. *Journal of Service Research*, 21(2), 155-172.

[43] Johnson J, Rhodes A, Wildenbeest MR (2020). Platform Design When Sellers Use Pricing Algorithms. *Working Paper*.

[44] Kamakura WA, Russell GJ (1989). A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research*, 26(4), 379-390.

[45] Kannan PK, Li H (2017). Digital marketing: A framework, review and research agenda. *International Journal of Research in Marketing*, 34(1), 22-45.

[46] Ke TT, Li C, Safronov M (2021). Learning by Choosing: Career Concerns with Observable Actions. *Working Paper*.

[47] Ke TT, Shen ZJM, Villas-Boas JM (2016). Search for Information on Multiple Products. *Management Science*, 62(12), 3576-3603.

[48] Ke TT, Sudhir K (2022). Privacy rights and data security: GDPR and personal data driven markets. *Working paper*.

[49] Ke TT, Villas-Boas JM (2019). Optimal learning before choice. *Journal of Economic Theory*, 180, 383-437.

[50] Keller G, Rady S (1999). Optimal Experimentation in a Changing Environment. *Review of Economic Studies*, 66, 475-507.

[51] Keller G, Rady S, Cripps M (2005). Strategic Experimentation with Exponential Bandits. *Econometrica*, 73, 39-68.

[52] Lewis M (2005). A Dynamic Programming Approach to Customer Relationship Pricing. *Management Science*, 51(6), 986-994.

[53] Li L, Chu W, Langford J, Schapire RE (2010). A contextual-bandit approach to personalized news article recommendation. *In the International World Wide Web Conference (WWW)*.

[54] Li S, Montgomery A, Sun B (2011). Cross-Selling the Right Product to the Right Customer at the Right Time. *Journal of Marketing Research*, 48(4), 683-700.

[55] Li X, Li L, Gao J, He X, Chen J, Deng L, He J (2015). Recurrent Reinforcement Learning: A Hybrid Approach. *ArXiv e-prints*.

[56] Lin S, Zhang J, Hauser JR (2015). Learning from Experience, Simply. *Marketing Science*, 34(1), 1-19.

[57] Liptser RS, Shiriaev AN (1977). Statistics of random processes: General theory (Vol. 394). New York: Springer-verlag.

[58] Liu Y, Ren D (2020). China drafts new antitrust guideline to rein in tech giants, wiping US$102 billion from Alibaba, Tencent and Meituan stocks. South China Morning Post, November 10, 2020. Accessed online at https://www.scmp.com/business/china-business/article/3109188/china-drafts-new-antitrust-guideline-rein-tech-giants on November 20, 2020.

[59] Ma L, Sun B (2020). Machine learning and AI in marketing-Connecting computing power to human insights. *International Journal of Research in Marketing*, forthcoming.

[60] Martinez S (2021). Streaming Service Algorithms Are Biased, Directly Affecting Content Development. *AMT Lab Blog*. Accessed on 11/24/2022 at https://amt-lab.org/blog/2021/11/streaming-service-algorithms-are-biased-and-directly-affect-content-development.

[61] Miklós-Thal J, Tucker C (2019). Collusion by algorithm: Does better demand prediction facilitate coordination between sellers? *Management Science*, 65(4), 1552-1561.

[62] Misra K, Schwartz EM, Abernethy J (2019). Dynamic online pricing with incomplete information using multi-armed bandit experiments. *Marketing Science*, 38(2), 226-252.

[63] Ning ZE (2021). List Price and Discount in A Stochastic Selling Process. *Marketing Science*, 40(2), 366-387.

[64] Pazgal A, Soberman D (2008). Behavior-based discrimination: Is it a winning play, and if so, when? *Marketing Science*, 27(6), 977-994.

[65] Romm T (2020). Amazon, Apple, Facebook and Google grilled on Capitol Hill over their market power. The leaders behind the tech giants testified before Congress virtually. *The Washington Post*, July 29, 2020. Available online at www.washingtonpost.com.

[66] Rossi PE, McCulloch RE, Allenby GM (1996). The value of purchase history data in target marketing. *Marketing Science*, 15(4), 321-340.

[67] Rothschild M (1974). A Two-Armed Bandit Theory of Market Pricing, *Journal of Economic Theory*, 9, 185-202.

[68] Rubinstein A (1982). Perfect Equilibrium in a Bargaining Model. *Econometrica*, 50, 97-109.

[69] Satariano A (2020). 'This Is a New Phase': Europe Shifts Tactics to Limit Tech's Power. *The New York Times*, July 30, 2020. Availabel online at www.nytimes.com.

[70] Schafer JB, Frankowski D, Herlocker J, Sen S (2007). Collaborative filtering recommender systems. *The Adaptive Web*. Springer, Berlin, Heidelberg. 291-324.

[71] Schwartz EM, Bradlow ET, Fader PS (2017). Customer Acquisition via Display Advertising Using Multi-Armed Bandit Experiments. *Marketing Science*, 36(4), 500-522.

[72] Silver D, Newnham L, Barker D, Weller S, McFall J (2013). Concurrent reinforcement learning from customer interactions. *In the International Conference on Machine Learning (ICML)*.

[73] Spence M (1984). Cost Reduction, Competition, and Industry Performance. *Econometrica*, 52, 101-121.

[74] Steckel JH, Winer RS, Bucklin RE, Dellaert BG, Drèze X, Häubl G, Jap SD, Little JDC, Meyvis T, Montgomery AL, Rangaswamy A (2005). Choice in interactive environments. *Marketing Letters*, 16(3-4), 309-320.

[75] Sun B, Li S (2011). Learning and Acting Upon Customer Information: A Simulation-Based Demonstration on Service Allocations with Offshore Centers. *Journal of Marketing Research*, 48(1), 72-86.

[76] Sun B, Li S, Zhou C (2006). "Adaptive" Learning and "Proactive" Customer Relationship Management. *Journal of Interactive Marketing*, 20(3/4), 82-96.

[77] Sutton RS, Barto AG (2018). *Reinforcement learning: An introduction*. MIT press.

[78] Theocharous G, Thomas PS, Ghavamzadeh M (2015). Personalized ad recommendation systems for life-time value optimization with guarantees. *In the International Joint Conference on Artificial Intelligence (IJCAI).*

[79] Competition and Markets Authority (2020). *Online platforms and digital advertising.*

[80] Urban GL, Liberali G, Bordley R, MacDonald E, Hauser JR (2014). Morphing banner advertising. *Marketing Science*, 33(1), 27-46.

[81] U.S. House, Committee on the Judiciary, Subcommittee on Antitrust, Commercial and Administrative Law (2020). *Investigation of Competition in Digital Markets*

[82] Villas-Boas JM (1999). Dynamic competition with customer recognition. *RAND Journal of Economics*, 30(4), 604-631.

[83] Villas-Boas JM (2004). Price cycles in markets with customer recognition. *RAND Journal of Economics*, 35(3), 486-501.

[84] Villas-Boas JM, Yao Y (2020). A Dynamic Model of Optimal Retargeting. *Marketing Science*, forthcoming.

[85] Vives X (2008). Innovation and Competitive Pressure. *The Journal of Industrial Economics*, 56, 419-469.

[86] Wedel M, Kannan PK (2016). Marketing analytics for data-rich environments. *Journal of Marketing*, 80(6), 97-121.

[87] Weitzman M (1979). Optimal Search for the Best Alternative. *Econometrica*, 47, 641-654.

[88] Winer RS, Neslin SA (Eds.) (2014). *The history of marketing science*. New York, NY: World Scientific.

[89] Xu Z, Dukes A (2020). Personalization, Customer Data Aggregation, and The Role of List Price. *Management Science*, forthcoming.

[90] Zhang J (2011). The Perils of Behavior-Based Personalization. *Marketing Science*, 30(1), 170-186.

[91] Zhang J, Krishnamurthi L (2004). Customizing promotions in online stores. *Marketing Science*, 23(4), 561-578.

Appendix

Proof of Equation 7

To derive the Hamilton-Jacobi-Bellman equation, notice that when the firm recommends niche-market content, the firm's value function satisfies

$$\begin{aligned} V(\lambda_t) &= y(\lambda_t, S_t)dt + (1 - r_f dt)E[V(\lambda_{t+dt})] \\ &= y(\lambda_t, S_t)dt + V(\lambda_t) - r_f V(\lambda_t)dt + \frac{\sigma(\lambda_t)^2}{2}V''(\lambda_t)dt \end{aligned} \tag{A1}$$

which simplifies to the following ordinary differential equation:

$$V(\lambda_t) = \frac{y(\lambda_t, S_t)}{r_f} + \frac{\sigma(\lambda_t)^2}{2r_f}V''(\lambda_t) \tag{A2}$$

The value function has two terms, which can be understood in the following ways. The first term, $y(\lambda_t, S_t)/r_f$, can be viewed as the present value of the profit if the firm stops learning information about the user. The second term corresponds to the value from learning and adapting to user behaviors in the future. Note that it is proportional to the instantaneous volatility of λ_t and V''. Consequently, V must be convex in λ_t for the value of learning and adapting to be positive. The value from adapting to new information is higher when λ_t is more volatile.

The general solution for equation (A2) is

$$V(\lambda_t) = \frac{y(\lambda_t, S_t)}{r} + b_1\lambda_t^{(\gamma+1)/2}(1-\lambda_t)^{-(\gamma-1)/2} + b_2\lambda_t^{-(\gamma-1)/2}(1-\lambda_t)^{(\gamma+1)/2}, \tag{A3}$$

with

$$\gamma = \sqrt{1 + \frac{8r_f\alpha(1-\alpha)}{(2\alpha-1)^2}}. \tag{A4}$$

However, because $\gamma > 1$, as $\lambda_t \to 1$ there will be no more uncertainty and thus the value function should satisfy $V(1) = y(1, S(1))/r_f$. Consequently, we must have:

$$b_1 = 0. \tag{A5}$$

Thus the solution simplifies to

$$V(\lambda_t) = \frac{y(\lambda_t, S_t)}{r_f} + b_2\lambda_t^{-(\gamma-1)/2}(1-\lambda_t)^{(\gamma+1)/2}, \tag{A6}$$

Proof of Proposition 3

Users who are the same ex-ante become heterogeneous from the firm's view as they exhibit different behaviors towards past recommendations. The population density starts as uni-modular and becomes bi-modular. As time goes to infinity, the mass moves toward 1 or $\hat{\lambda}$. In the limit, all users who are recommended niche-market content must receive the correct type of content.

Let $\hat{H}(t)$ denote the probability that the user hits threshold $\hat{\lambda}$ before time t. By the law of large numbers, $\hat{H}(t)$ is also the proportion of users in the population that hits $\hat{\lambda}$. So $1-\hat{H}(t)$ is the number of users being recommended content type N_1 at time t.

Let

$$z = \ln\left(\frac{\lambda}{1-\lambda}\right)$$

Then we have:

$$\begin{aligned}\lambda = g(z) &\equiv \frac{e^z}{1+e^z}\\ h(\lambda,t) &= p(z,t)/g'(z)\\ g'(z) &= \frac{e^z}{(1+e^z)^2}\end{aligned}$$

where $h(\lambda,t)$ and $p(z,t)$ are the probability density functions of λ and z at time t.

For users who prefer content type N_1, we have

$$dz = \frac{1}{2}\sigma_z^2 dt - \sigma_z dW$$

with $\sigma_z \equiv (2\alpha-1)/\sqrt{\alpha(1-\alpha)}$.

The probability density of z is

$$p_1(z,t) = \frac{1}{\sqrt{2\pi\sigma_z^2 t}}\exp\left(-\frac{(z-z_0-\sigma_z^2 t/2)^2}{2\sigma_z^2 t}\right)$$

The probability density for λ at time t is

$$h_1(\lambda,t) = p_1(z,t)dz/d\lambda = p_1(z(\lambda),t)/[\lambda(1-\lambda)]$$

And we have

$$\lambda_t = \frac{(\lambda_0/(1-\lambda_0))\exp\left(\sigma_z^2 t/2 - \sigma_z W(t)\right)}{1+(\lambda_0/(1-\lambda_0))\exp\left(\sigma_z^2 t/2 - \sigma_z W(t)\right)}$$

Moreover, there is $1-\lambda_0$ proportion of users who prefer content type N_2. For this group of users, their posterior belief λ follows the following stochastic differential equation:

$$dz = -\frac{1}{2}\sigma_z^2 dt - \sigma_z dW$$

The probability density of y is

$$p_2(z,t) = \frac{1}{\sqrt{2\pi\sigma_z^2 t}} \exp\left(-\frac{(z - z_0 + \sigma_z^2 t/2)^2}{2\sigma_z^2 t}\right)$$

The probability density for λ at time t is

$$h_2(\lambda, t) = p_2(z,t)dz/d\lambda = p_2(z(\lambda), t)/[\lambda(1-\lambda)]$$

and for users who prefer N_2:

$$\lambda_t = \frac{(\lambda_0/(1-\lambda_0))\exp(-\sigma_z^2 t/2 - \sigma_z W(t))}{1 + (\lambda_0/(1-\lambda_0))\exp(-\sigma_z^2 t/2 - \sigma_z W(t))}$$

The first hitting time probability density for users who prefer N_1 is

$$h_1(z_0, t) = (z_0 - \hat{z})\frac{1}{\sqrt{\sigma_z^2 t^3}} n\left(\frac{z_0 - \hat{z} + \sigma_z^2 t/2}{\sigma_z\sqrt{t}}\right)$$

The cumulative probability distribution of hitting times for this case is

$$H_1 = \Phi\left(\frac{(\hat{z} - z_0) - \sigma_z^2 t/2}{\sigma_z\sqrt{t}}\right) + \exp(\hat{z} - z_0)\Phi\left(\frac{(\hat{z} - z_0) + \sigma_z^2 t/2}{\sigma_z\sqrt{t}}\right)$$

For users who prefer N_2:

$$h_2(z_0, t) = (z_0 - \hat{z})\frac{1}{\sqrt{\sigma_z^2 t^3}} n\left(\frac{z_0 - \hat{z} - \sigma_z^2 t/2}{\sigma_z\sqrt{t}}\right)$$

The cumulative probability distribution of hitting times for this case is

$$H_2 = \Phi\left(\frac{(\hat{z} - z_0) + \sigma_z^2 t/2}{\sigma_z\sqrt{t}}\right) + \exp(z_0 - \hat{z})\Phi\left(\frac{(\hat{z} - z_0) - \sigma_z^2 t/2}{\sigma_z\sqrt{t}}\right)$$

The probability density is

$$\begin{aligned} h(z_0, t) &= \lambda_0 h_1(z_0, t) + (1-\lambda_0)h_2(z_0, t) \\ &= (\lambda_0 + (1-\lambda_0)\exp(z_0 - \hat{z}))(z_0 - \hat{z})\frac{1}{\sqrt{\sigma_z^2 t^3}} n\left(\frac{z_0 - \hat{z} + \sigma_z^2 t/2}{\sigma_z\sqrt{t}}\right) \end{aligned} \tag{A7}$$

and the total cumulative probability distribution of hitting times for the model is

$$\begin{aligned} H(t) &= \lambda_0 H_1(z_0, t) + (1 - \lambda_0) H_2(z_0, t) \\ &= \frac{\lambda_0}{\hat{\lambda}} \left[\Phi \left(\frac{(\hat{z} - z_0) - \sigma_z^2 t/2}{\sigma_z \sqrt{t}} \right) + \exp(\hat{z} - z_0) \Phi \left(\frac{(\hat{z} - z_0) + \sigma_z^2 t/2}{\sigma_z \sqrt{t}} \right) \right] \end{aligned} \tag{A8}$$

where $z = ln\left(\frac{\lambda}{1-\lambda}\right)$ and $\sigma_z = (2\alpha - 1)/\sqrt{\alpha(1-\alpha)}$. As t approaches infinity, $\hat{H}(t)$ converges to a constant:

$$\lim_{t \to \infty} \hat{H}(t) = \frac{1 - \lambda_0}{1 - \hat{\lambda}}$$

Let $\bar{\lambda}_t$ denote the probability that a user who prefers content type N_1 conditional on that the firm recommends content type N_1 to her at time t. Since λ_t is a martingale for all i, we must have

$$(1 - \hat{H}(t))\bar{\lambda}_t + \hat{H}(t)\hat{\lambda} = \lambda_0$$

and thus

$$\bar{\lambda}_t = \frac{\lambda_0 - \hat{H}(t)\hat{\lambda}}{1 - \hat{H}(t)} \tag{A9}$$

and

$$\lim_{t \to \infty} \bar{\lambda} = 1$$

Since $\hat{H}(t)$ is increasing in t, $\bar{\lambda}_t$ must also increase in t. Note that $\bar{\lambda}$ approaching 1 implies that in the limit, only users who prefer type N_1 may receive N_1 content. Users who prefer type N_2, but were incorrectly recommended type N_1 content initially under the prior, receive mass-market content in the limit. This also implies that there will be $\frac{1-\lambda_0}{1-\hat{\lambda}} - (1 - \lambda_0)$ fraction of users who prefer type N_1 but are incorrectly recommended mass-market content in the long run.

Similarly, for the myopic algorithm, we can show that the probability that the user hits the myopic threshold λ^* before time t is

$$H^*(t) = \frac{\lambda_0}{\lambda^*} \left[N \left(\frac{(z^* - z_0) - \sigma_z^2 t/2}{\sigma_z \sqrt{t}} \right) + \exp(z^* - z_0) \Phi \left(\frac{(z^* - z_0) + \sigma_z^2 t/2}{\sigma_z \sqrt{t}} \right) \right] \tag{A10}$$

and

$$\lim_{t \to \infty} H^*(t) = \frac{1 - \lambda_0}{1 - \lambda^*} \qquad \text{and} \qquad \lim_{t \to \infty} \bar{\lambda} = 1$$

The results above are summarized in Proposition 3.

Evolution of Profit for a Monopoly

Recall that $\hat{H}(t)$ is the fraction of users who have hit the absorbing barrier $\hat{\lambda}$ at time t and $\bar{\lambda}_t$ denotes the population average of λ_t among the remaining users. The profit flow at time t is

$$\begin{aligned}\pi_t &\equiv E\int y(\lambda_t, S(\lambda_t)) \\ &= (1-\hat{H}(t))y(\bar{\lambda}_t, S(\bar{\lambda}_t)) + \hat{H}(t)c \\ &= y(\lambda_0, S(\lambda_0)) - \hat{H}(t)y(\hat{\lambda}, S(\hat{\lambda})) + \hat{H}(t)c \\ &= \lambda_0\alpha + (1-\lambda_0)(1-\alpha) - \hat{H}(t)[\hat{\lambda}\alpha + (1-\hat{\lambda})(1-\alpha) - c]\end{aligned} \tag{A11}$$

Note that $c > y(\hat{\lambda}, S(\hat{\lambda}))$ and $\hat{H}(t)$, which is a cumulative density function, is an increasing function of t. Thus, π_t must be increasing in t. If $\lambda_0 \in (\hat{\lambda}, \lambda^*)$, then $\pi(t)$ is smaller than c for small t. For $\pi_t = c$, we must have

$$\hat{H}(t) = \frac{y(\lambda_0, S(\lambda_0)) - c}{y(\hat{\lambda}, S(\hat{\lambda})) - c} = \frac{\lambda_0(2\alpha-1) + (1-\alpha) - c}{\hat{\lambda}(2\alpha-1) + (1-\alpha) - c}$$

The turning point for when expected profit flow is above c is given by $t_0 = \pi^{-1}(c) = \hat{H}^{-1}\left(\frac{y(\lambda_0, S(\lambda_0))-c}{y(\hat{\lambda}, S(\hat{\lambda}))-c}\right)$. This means that the firm expects to suffer losses from deploying the forward-looking algorithm until the proportion of users that receive mass-market content reaches $\frac{y(\lambda_0, S(\lambda_0))-c}{y(\hat{\lambda}, S(\hat{\lambda}))-c}$.

We define the discounted cumulative profit function up to time t as Π_t. For $\lambda_0 \in (\hat{\lambda}, \lambda^*)$, we plot function $\pi(t)$ and $\Pi(t)$ in Figures 4a and 4b, and compare them to flow and cumulative profit under the myopic algorithm. The expected profit flow increases over time but remains below myopic flow profit until t_0. The gap between $\Pi(t)$ and the cumulative profit under the myopic algorithm first widens over time, then begins to narrow after $t > t_0$, and eventually becomes positive after some later time t_1. Such t_1 must exists, otherwise the algorithm must not be optimal.

The results are summarized as follows:

Corollary A.1 *The expected flow profit under the forward-looking algorithm increases over time. For $\lambda_0 \in (\hat{\lambda}, \lambda^*)$, the expected flow profit under the forward-looking algorithm is lower than the expected flow profit under the myopic algorithm for $t < t_0 = \hat{H}^{-1}\left(\frac{y(\lambda_0, S(\lambda_0))-c}{y(\hat{\lambda}, S(\hat{\lambda}))-c}\right)$.*

Proof of Proposition 5

First, we prove that this is an equilibrium. We show that neither firm has an incentive to deviate. Due to symmetry around $\lambda = 0.5$, we assume WLOG that $\lambda_t^1 \geq 0.5$ and $\lambda_t^2 \geq 0.5$.

Suppose $\lambda_t^1 \leq \widehat{\lambda}_u^1$, and consider firm 1's deviation from mass-market content to niche-market content. By equation (15), firm 1 will receive no demand for any λ_t^2. This deviation cannot be profitable. Suppose $\lambda_t^1 > \widehat{\lambda}_u^1$, and consider firm 1's deviation from niche-market content to mass-market content. If $\lambda_t^2 > \widehat{\lambda}_u^2$, then firm 1 has no demand after deviating, which cannot be profitable. If $\lambda_t^2 \leq \widehat{\lambda}_u^2$, then firm 1 gets a flow profit of $\frac{1}{2}c$ with no new information. Given our assumption that $c < 1$, we have $\lambda_t \alpha^1 + (1-\lambda_t)(1-\alpha^1) > \frac{1}{2}c$. That is, the expected flow profit from offering niche-market content is always higher than the flow profit from splitting demand with mass-market content. This deviation cannot be profitable. One can show that firm 2 does not have a profitable deviation in the same way.

Next, we prove that this is the unique equilibrium with stationary algorithm that is robust to all priors λ_0^1 and λ_0^2. First, consider the case of $\lambda_0^1 \neq 0.5$ and $\lambda_0^2 = 0.5$. We must have $S^2(0.5) = M$ in equilibrium, because firm 2 always gets no demand if it recommends niche-market content when $\lambda_t^2 = 0.5$. Consider an alternative strategy for firm 1. Suppose there exists an equilibrium strategy profile such that $S^1(\tilde{\lambda}) = M$ for some $\tilde{\lambda} > \widehat{\lambda}_u^1$. Then select $\lambda_0^1 = \tilde{\lambda}$ and $\lambda_0^2 = 0.5$. At time 0, firm 1 can profitably deviate by switching to the strategy in Proposition 5. After deviating, firm 1 gets a flow payoff of $\lambda_t^1 > \frac{1}{2}c$ until λ_t^1 hits $\widehat{\lambda_u^1}$, and gets a flow payoff of $\frac{1}{2}c$ after λ_t^1 hits $\widehat{\lambda_u^1}$. This is strictly more profitable than getting $\frac{1}{2}c$ at all t. Now suppose there exists an equilibrium strategy profile with $S^1(\tilde{\lambda}) = N$ for some $\tilde{\lambda} \leq \widehat{\lambda}_u^1$. Then let $\lambda_0^1 = \tilde{\lambda}$ and $\lambda_0^2 = 0.5$. Then firm 1 can profitably deviate to offering mass-market content forever, which increases the total payoff from 0 to $\frac{c}{2r_f}$. Thus no other strategy for firm 1 can be equilibrium for all priors.

Now by symmetry we have established that $S^1(0.5) = M$, which then can be used to prove that there is no alternative strategy for firm 2 that can be equilibrium for all priors. The strategy profile in Proposition 5 is the unique stationary strategy profile, $S_t^j = S(\lambda_t^j)$, that constitutes equilibrium for all possible priors.

Proof of Proposition 6

The proof is separated into three steps. We first consider user preferences. Then we solve for firm 1's optimal strategy in the simpler case where firm 1 can observe λ_t^2. We prove that the algorithm proposed in 6 is optimal if firm 1 can observe λ_t^2. Because the algorithm does not depend on λ_t^2, it must also be optimal when firm 1 cannot observe λ_t^2. Then we show that any other stationary algorithm must be sub-optimal for some priors.

Step 1: User Behavior

Firm 2's myopic algorithm recommends niche-market content if and only if λ_t^2 is greater than the myopic threshold, λ^*. Note that the user always weakly prefers firm 2's content to

mass-market content. Thus the user never strictly prefers the mass-market content from firm 1. The user is indifferent between the mass-market content from firm 1 and the recommended content from firm 2 when $\lambda_t^2 \leq \lambda^*$.

Consider the user's preferences between niche-market content from firm 1 and content recommended by firm 2. The Gittins index for niche-market content from firm 1 is provided by equation 14.

Let $G^{MY}(\lambda_t^2)$ denote the Gittins index for the myopic firm's content. If $\lambda_t^2 \leq \lambda^*$, then $G^{MY}(\lambda_t^2) = c$ because firm 2 always offers mass-market content. Consider now the case of $\lambda_t^2 > \lambda^*$. By the definition of the Gittins index, with a static arm paying a flow payoff of $G^{MY}(\lambda_t^2)$, the user would switch from firm 2 to the static arm at or below λ_t^2. However, because firm 2 only offers niche-market content above λ_t^2, we must have $G^{MY}(\lambda_t^2) = G(\lambda_t^2)$ where $G(\lambda_t^2)$ is the Gittins index for the niche-market content from firm 2 from equation 14. Thus the Gittins index for the myopic firm is the same as the Gittins index for the niche-market content for $\lambda_t^2 > \lambda^*$, then jump discretely down to c at λ^*

We can then summarize firm 1's demand, $D(\lambda_t^1, \lambda_t^2 | S_t^1)$, as:

$$D(\lambda_t^1, \lambda_t^2 | N) = \mathbb{1}\{G(\lambda_t^1) \geq G^{MY}(\lambda_t^2)\}$$

$$D(\lambda_t^1, \lambda_t^2 | M) = \frac{1}{2}(1 - \mathbb{1}\{\lambda_t^2 \leq \lambda^*\})$$

Step 2: Optimal Algorithm

We only need to prove that the algorithm proposed in 6 is optimal if firm 1 can observe λ_t^2. Because the stationary algorithm does not depend on the actual value of λ_t^2, it must also be optimal when firm 1 cannot observe λ_t^2.

Consider deviation from 6. The state, λ_t^1 and λ_t^2, has three cases.

Case 1: $\lambda_t^2 \leq \lambda^*$. Firm 2 always recommends mass-market content starting from time t. Given our analysis in Section 4.1, one can conclude that firm 1 offering niche-market content if and only if $\lambda_t^1 \notin [1 - \widehat{\lambda_u^1}, \widehat{\lambda_u^1}]$ is the unique optimal solution starting from time t. There is no profitable deviation.

Case 2: $\lambda_t^2 > \lambda^*$ and $\lambda_t^1 \notin [1 - \widehat{\lambda_u^1}, \widehat{\lambda_u^1}]$. In this case, the user strictly prefers firm 2's content to mass-market content. Thus firm 1 receives no demand at time t if it deviates to recommending mass-market content, which cannot be a profitable deviation.

Case 3: $\lambda_t^2 > \lambda^*$ and $\lambda_t^1 \in [1 - \widehat{\lambda_u^1}, \widehat{\lambda_u^1}]$. In this case, the user does not visit firm 1 at time t regardless of firm 1's content choice. Thus there is no profitable deviation at time t.

Combining the three cases, we can conclude that the stationary algorithm where firm 1 offers niche-market content if and only if $\lambda_t^1 \notin [1 - \widehat{\lambda_u^1}, \widehat{\lambda_u^1}]$ is optimal.

Step 3: Uniqueness

If $\lambda_0^2 \leq \lambda^*$, any other stationary algorithm must be sub-optimal for some λ_0^1, as shown in the case 1 above. This completes the proof.

Online Appendix

1 The Additional Value from the Myopic Algorithm

In the main text, we study the value of upgrading from the myopic algorithm to the forward-looking algorithm. Here we consider the value of upgrading from making non-adaptive recommendations to the myopic algorithm.

We denote the firm's ex-ante expected profits under the non-adaptive and the myopic algorithm as $V^{NA}(\lambda_t)$ and $V^{MY}(\lambda_t)$, respectively.

Under the non-adaptive algorithm, the firm always recommends niche-market content if the initial belief, λ_0, is not in $(1-\lambda^*, \lambda^*)$, where $\lambda^* = \frac{c-(1-\alpha)}{2\alpha-1}$. The expected lifetime value for a given λ_0 is

$$V^{NA}(\lambda_0) = \begin{cases} \frac{\lambda_0\alpha+(1-\lambda_0)(1-\alpha)}{r} & \text{for } \lambda_0 > \lambda^* \\ \frac{c}{r} & \text{for } \lambda_0 \in (1-\lambda^*, \lambda^*) \\ \frac{\lambda_0(1-\alpha)+(1-\lambda_0)\alpha}{r} & \text{for } \lambda_0 < 1-\lambda^* \end{cases} \tag{OA1}$$

Under the myopic algorithm, the firm's expected profit at $t = 0$ is given in equation 8:

$$V^{MY}(\lambda_0) = \begin{cases} \frac{\lambda_0\alpha+(1-\lambda_0)(1-\alpha)}{r} & \text{for } \lambda_0 > \lambda^* \\ \frac{c}{r} & \text{for } \lambda_0 \in (1-\lambda^*, \lambda^*) \\ \frac{\lambda_0(1-\alpha)+(1-\lambda_0)\alpha}{r} & \text{for } \lambda_0 < 1-\lambda^* \end{cases} \tag{OA2}$$

We call $V^{MY}(\lambda_0) - V^{NA}(\lambda_0)$ the additional value from the myopic algorithm.

Corollary OA.1 *The additional value from the myopic algorithm, $V^{MY}(\lambda_0) - V^{NA}(\lambda_0)$, is zero for all λ_0.*

Corollary OA.1 suggests that for a monopoly, the value from continuously learning the user's preferences is zero if recommendations are made myopically. Many recommender systems have supervised learning algorithms that can predict the likelihood that a user enjoys each type of content. This result highlights the inadequacy of recommending myopically based on this ranking.

The strong result of Corollary OA.1 is due to the fact that the expected profit flow from recommending niche-market content, $y(\lambda_t, S_t)$, is linear in posterior belief λ_t (see equation 4). This linearity means that the net present value of recommending N1 forever is a martingale in λ_t. When $\lambda_t > \lambda^*$, the myopic algorithm and the non-adaptive algorithm make the same recommendation. They begin to diverge exactly at the moment when λ_t drops to the

myopic threshold, λ^*. However, at the myopic threshold, the net present value of always recommending N1 equates to the net present value of always recommending M. Thus, this implies that the value functions under the myopic algorithm and the non-adaptive algorithm are the same under all priors.

If, instead, the expected profit flow from recommending N1 is concave around λ^*, then the net present value of always recommending N1 becomes a supermartingale in λ_t, which implies that at λ^*, recommending M forever is better than recommending N1 forever. So the additional value from the myopic algorithm becomes positive. This can happen, for example, when the user has an outside option, which is explored below.

On the other hand, if the expected profit flow from recommending N1 is convex in λ_t, then the net present value of always recommending N1 becomes a submartingale in λ_t, which implies that the additional value from the myopic algorithm is negative. For example, if besides customizing content, the firm also has to choose between two types of advertisements, one targeted to users who prefer N1 and the other targeted to users who prefer N_2, then the flow payoff can be made to be convex in λ_t.

It is important to note that Corollary OA.1 depends on a few model assumptions besides linearity. The result is only for a monopoly with fixed demand that does not need to worry about losing users to a competitor. Additionally, we assume that the firm does not learn any information when recommending mass-market content. The myopic algorithm will have more value if we allow the firm to continue learning, at a slower pace, when recommending mass-market content.

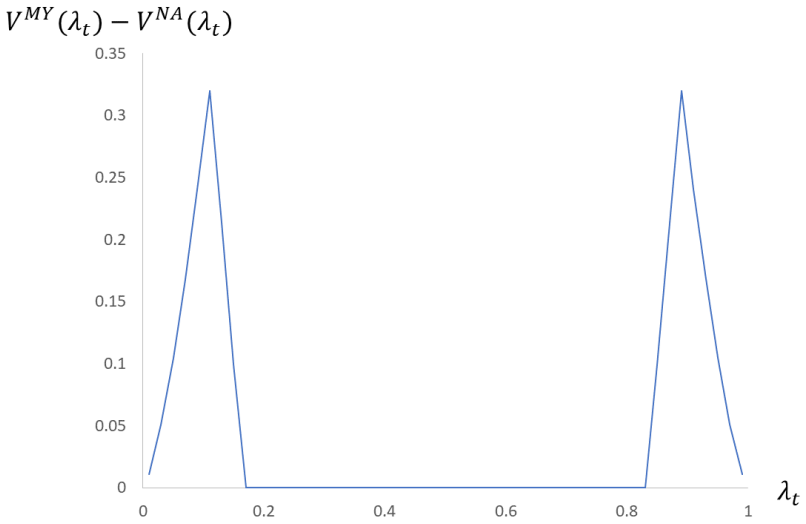
In the monopoly model, the value of upgrading from the non-adaptive algorithm to the myopic algorithm is zero. However, this is no longer true when users have an outside option. Intuitively, with the threat from an outside option, the myopic algorithm creates value by preventing the user from switching to the other platform. If the firm employs the non-adaptive algorithm, then it will lose the user forever when λ_t drops to $\widehat{\lambda_u}$. By using the myopic algorithm, the firm can keep half of the user's demand on its platform even when λ_t drops to $\widehat{\lambda_u}$. This also implies that the presence of an outside option may increase firms' incentives to develop the myopic algorithm even though it lowers their incentives to develop the forward-looking algorithm.

Corollary OA.2 *When competing against a mass-market content provider, the additional value from the myopic algorithm is strictly positive if* $\lambda_0 > \lambda^* = \frac{c-(1-\alpha)}{2\alpha-1}$.

Figure 10 shows how the additional value from the myopic algorithm change with λ_t.

In contrast to the monopoly case, competition makes the myopic algorithm valuable by

Figure 10: Additional value from the myopic algorithm



for $\alpha = 0.8$, $c = 0.7$, and $r = 0.6$

deterring customers from switching. Competition boosts companies' incentives to develop myopic adaptive-learning capacities, but dampens their incentives to develop forward-looking capacities. Understanding the complexity of the incentive to develop learning capabilities is important for both managers and policy makers.

2 Endogenous Monetization Level for Monopoly

In the monopoly model, the firm earns a fixed margin when a user engages with the content, and each user only consumes one unit of content per "period." The speed of learning is constant. In this section, we consider an extension in which the firm's margin, the quantity of content that a user consumes, and the speed of learning are endogenous. For simplicity, we assume that users are myopic in that they only maximize their instantaneous utility.

The firm chooses the level of monetization, which affects the quantity of content that users consume and the firm's speed of learning. For example, the firm may generate profit from advertising embedded in the content. The amount of advertising can be seen as a price levied on users. A higher level of monetization, such as by increasing the amount of advertising, increases the margin that the firm gets per content consumed, but decreases the amount of content that a user views on the platform. We also assume that in each period, users have diminishing marginal utility on the quantity of content consumed in this period.

At time t, the user's marginal utility from content is $\frac{du}{dq_t} = \beta_1 - \beta_2 q_t - p_t$, where q_t is the

amount of content consumed by the user at time t, and p_t is the level of monetization.

The process for the cumulative profit from the user is the same as in the base model, but with respect to the cumulative amount of content viewed, Q, instead of time t:

$$dY(Q) = y(T, S_Q)dQ + \sqrt{\alpha(1-\alpha)p_Q^2}dW(Q) \tag{OA3}$$

where the cumulative amount of content viewed follows $dQ_t = q_t dt$. This implies:

$$dY(t) = y(T, S_t)q_t dt + \sqrt{q_t}\sqrt{\alpha(1-\alpha)p_t^2}d\tilde{W}_t \tag{OA4}$$

for some Wiener process $\tilde{W}_t$. Because the user maximizes her instantaneous consumption utility, we have

$$q_t = \frac{\beta_1}{\beta_2} - \frac{1}{\beta_2}p_t$$

The expected profit flow at time t becomes:

$$\pi_t = q_t\big[p_t[\lambda_t\alpha + (1-\lambda_t)(1-\alpha)]\big]$$

The learning process becomes:

$$\begin{aligned} d\lambda_t &= \frac{\lambda_t(1-\lambda_t)(2\alpha-1)p_t q_t}{\alpha(1-\alpha)p_t^2}[y(T) - y(\lambda_t)]dt + \frac{\lambda_t(1-\lambda_t)(2\alpha-1)p_t q_t}{\sqrt{\alpha(1-\alpha)p_t^2 q_t}}dW_t \\ &= \frac{\lambda_t(1-\lambda_t)(2\alpha-1)(\frac{\beta_1}{\beta_2} - \frac{1}{\beta_2}p_t)}{\alpha(1-\alpha)p_t}[y(T) - y(\lambda_t)]dt + \frac{\lambda_t(1-\lambda_t)(2\alpha-1)\sqrt{(\frac{\beta_1}{\beta_2} - \frac{1}{\beta_2}p_t)}}{\sqrt{\alpha(1-\alpha)}}dW_t \end{aligned}$$

Note that the standard deviation of λ_t, $\frac{\lambda_t(1-\lambda_t)(2\alpha-1)\sqrt{(\frac{\beta_1}{\beta_2} - \frac{1}{\beta_2}p_t)}}{\sqrt{\alpha(1-\alpha)}}$, decreases in p_t. Thus, lowering the level of monetization can increase the speed of learning by increasing the amount of content users view, which increases the speed of data collection.

The firm's value function, or maximized lifetime value is given by:

$$V(\lambda_t) = \max_{p_t} V(p_t, \lambda_0) = V(p_t, \lambda_0) = E\int_0^\infty e^{-r_f t}\pi_t dt$$

The HJB equation gives us:

$$0 = 0 + \Big[\max_{p_t} q_t\big(p_t[\lambda_t\alpha + (1-\lambda_t)(1-\alpha)]\big)\Big] - r_f V(\lambda_t) + \frac{\lambda_t^2(1-\lambda_t)^2(2\alpha-1)^2 p_t^2 q_t}{2(\alpha(1-\alpha)p_t^2)}V''(\lambda_t)$$

For an interior solution, we take the first-order condition of the right hand side with respect to p_t. If there is no learning, the myopic strategy is to set the monetization level at

$$p_t^* = M^{-1}(0) = \frac{\beta_1}{2}$$

The difference between the myopic level and the forward-looking level is:

$$p_t^* - \hat{p}_t = \frac{\lambda_t^2(1-\lambda_t)^2(2\alpha-1)^2}{4\alpha(1-\alpha)}V''(\lambda_t)$$

Note that in Corollary OA.1, we show that if the firm makes myopic recommendations, then $V''(\lambda_t) = 0$, and the additional value from the myopic algorithm is zero. This means that if the firm makes recommendations myopically, it should monetize the content at a constant level of $p^* = \frac{\beta_1}{2}$.

We can obtain the ODE for the value function (for niche-market content) by rearranging equation (A2):

$$V(\lambda) = \hat{q}_t\frac{\hat{p}_t[\lambda_t\alpha + (1-\lambda_t)(1-\alpha)]}{r_f} + \hat{q}_t\frac{\lambda^2(1-\lambda)^2(2\alpha-1)^2}{2r_f\alpha(1-\alpha)}V''(\lambda)$$

Because information adds value, we have $V''(\lambda_t) > 0$, which implies that the forward-looking monetization level is always strictly lower than the myopic monetization level. Thus when there is opportunity to learn and adapt to each user's preference, a forward-looking firm should reduce monetization. Less monetization, for example, by limiting the amount of advertising, encourages users to view more content, which increases the firm's speed of learning. There is no closed form solution to the ODEc but we can solve it numerically. Figure 11 shows this graphically.

For niche-market content, when λ_t increases, the need for experimentation also falls. As a result, the firm increases advertising. As λ_t approaches 0 or 1, the need for information vanishes, and the forward-looking monetization level must approach the myopic level. However, if the firm is serving mass-market content, the myopic level is optimal because there is no more information to learn. As a result, the optimal forward-looking monetization level is non-monotonic.

Proposition OA.1 *With the myopic algorithm, the firm monetizes with a constant rate of* $\frac{\beta_1}{2}$*. With the optimal forward-looking algorithm, the firm monetizes less when recommending niche-market content. The reduction in monetization goes to zero as* $\lambda_t \to 0$ *or* 1*, or as* $t \to \infty$*.*

Figure 11: The monetization level as a function of λ