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

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

August 1, 2021

Abstract

Through repeated interactions, firms today refine their understanding of individual users' preferences adaptively for personalization. In this paper, we use a continuous-time multi-agent bandit model to analyze firms that supply content to 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 firms that compete for users' attention focuses more on exploitation than exploration than a monopoly would. When users are impatient, competition decreases firms' incentives to develop forward-looking algorithms. On the other hand, development of the optimal forward-looking algorithm may hurt users under monopoly but always benefits users under competition. We are among the first to examine this multi-agent bandit problem under different competitive scenarios, and our results provide implications for AI adoption as well as for policy makers on the effect of market power on innovation and consumer welfare.

Keywords: AI, multi-agent bandit, recommendation algorithm, innovation, competition, reinforcement learning, experimentation.

*The authors thank J. Miguel Villas-Boas 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 Robert H. Smith School of Business, University of Maryland. Z. Eddie Ning is Assistant Professor of Marketing at Cheung Kong Graduate School of Business. 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, zhaoning@ckgsb.edu.cn, bhsun@ckgsb.edu.cn.

1 Introduction

Two inter-related 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 produce fine-grained digital consumer footprint which provide 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 in cloud computing, big data, machine learning, and AI has provided powerful tools to assist in large-scale automated decision making, which have greatly increased firms' ability to understand and fulfill customers' needs on a real-time basis (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 at e-commerce websites such as Amazon and Taobao. Digital advertisements are increasingly targeted at a personal level based on a user's past behaviors. Even more prevalent, popular social media and content platforms such as Facebook, Youtube, Spotify, Tiktok, and many news media sites, 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. This trend has also propelled a number of enterprise service sub-industries that provide technological and analytical services.

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 firms' recommendations. First, using historical data of various types, a firm can learn about consumer preferences at the group or individual level in a static fashion, and make customized recommendations based on the inferred segmentation. A rich body of literature developed over several decades, e.g. dynamic choice models, incorporates consumer heterogeneity in a 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), we call this paradigm *the non-adaptive recommendation algorithm*, which, given its prevalence, can be considered as the baseline.

In the second paradigm, going one step further from the baseline, 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 would continuously update its understanding of consumer preferences based on new information. At any point in time, the firm would make targeted offerings based on its best understanding of a consumer's preferences at that point. Many statistical techniques exist and machine learning algorithms are developed to 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, then use a predictive model to rank them by objectives such as click 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. Accordingly, 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 more step further from the adaptive but myopic algorithm, a third and more powerful recommendation paradigm is emerging. A firm not only learns adaptively from historical information, but also takes a forward-looking perspective in its recommendations to proactively gather new information in a guided manner. For example, while based on the current understanding, a consumer is most likely to enjoy a specific type of content, the firm may instead find it useful to recommend something different. This 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 consumer 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 each user's long-term satisfaction with the system (Chen et al. 2019). We call this third paradigm *the forward-looking recommendation algorithm*. Optimizing in a forward-looking framework, this recommendation paradigm is expected to outperform adaptive myopic recommendations.

The proliferation of consumer data has understandably attracted considerable attention from scholars in multiple fields. Research in computer science and machine learning has developed a vast and powerful tool set to administer large volumes of data and to extract

information from the data. Empirical research in marketing has consistently confirmed the value of consumers' digital footprint on understanding their preferences and decisions (Winer and Neslin 2014). Noticeably left out in both streams of research, however, is the theoretical implications of firms' continuous personalization, especially under competition. 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 source of value in different competitive situations.

In this paper, we address three questions: first, how does the presence of competition affects the optimal trade-off between exploration and exploitation? What characteristics of the users affect this trade-off? Second, how much value does a forward-looking algorithm provide to the firm over a myopic algorithm? How does competition affect the firms' incentives to invest and upgrade to a forward-looking algorithm? Third, how does upgrading from a myopic algorithm to a forward-looking algorithm affect consumer welfare with or without competition? These questions has theoretical and managerial implications on technology adoption as well as regulation. In light of rising concerns expressed by regulators over major tech firms' market power, our research contributes to the discussion by investigating the effect of market power on firms' incentives to develop advanced AI algorithms and their subsequent effects on consumer welfare.

We consider an online content consumption scenario, such as that of Youtube, which is a representative setting where firms offers adaptive personalization to its users. 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 updates its belief about the user's preferences and adjusts the recommendations adaptively. We formulate firms' decisions as a continuous-time multi-armed bandit problem. This framework incorporates key factors such as firms' continuous learning of users' preferences, adaptive responses to real-time information, and forward-looking optimization in a parsimonious manner.

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 situations: (1) when a firm acts as a monopoly, and (2) when two firms compete as a duopoly. In the competitive scenario, two firms compete for the attention of each user. The user chooses which firm to visit at each moment. This constitutes a 3-player bandit problem, where two firms and a user each faces a bandit problem, and the outcome of a decision depends on the decisions made by the other two players. Our study is among the first to study such multi-agent bandit problem arising from competition.

We derive closed-form solutions to the simultaneous dynamic optimization problem, and

the results reveal several important insights. Surprisingly, a monopoly that uses the myopic algorithm, which focus solely on exploitation, may not perform better than a monopoly that uses a non-adaptive algorithm. This shows that without competition, the value of adaptive personalization may be concentrated on the exploration aspect. The additional value from developing the forward-looking algorithm is also found to be non-monotonic in firms' prior knowledge about a user's preferences. To expedite exploration of consumer preferences, the forward-looking algorithm induces the firm to serve 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 would lead to a reduced profit in the near term, although the profit will increase later to more than compensate for the near-term loss.

The situation changes substantially, however, when a firm faces competition. The presence of competition pushes the forward-looking algorithm to shift towards exploitation by recommending less niche content compared to the monopoly case, due to less room for strategic experimentation. More importantly, in the case of competition, the user' discount rate becomes a crucial factor, depending on which the competitive forward-looking algorithm spans a continuous spectrum between the myopic algorithm and the monopoly's optimal forward-looking algorithm. 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 firm will be forced to adopt the myopic algorithm. Furthermore, in contrast to the monopoly case, the myopic algorithm perform better than the non-adaptive baseline under competition, as even myopic recommendations help a firm to retain its users from switching.

We also contribute by analyzing firms' incentives to adopt machine learning and AI technologies and the value of learning while taking into account firms' strategic behaviors. First, our analysis suggests that firms under competitive pressure may have lower incentives to invest in technologies such as reinforcement learning that enable forward-looking algorithms, and instead could be content with myopic algorithms. In contrast, a monopolist, under less competition, has more room to invest in forward-looking algorithms that bring long-term benefits at the cost of near-term profits.

On the flip side, there is also a trade-off between innovation incentives and consumer welfare. A monopoly may have a higher incentive to develop the forward-looking algorithm, but such technological advancement may hurt users due to overly aggressive customization. With competition, the development of forward-looking algorithms are always beneficial to users, but firms may have less incentives to do so when users are impatient or not forward-looking in their content consumption behaviors.

While we focus our analysis 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.

The rest of the paper is organized as follows. After reviewing the relevant literature in Section 2, we set up and analyze the monopoly model in Section 3. We then study the duopoly scenario in Section 4. In Section 5 we explore some extensions in which firms are asymmetric. Section 6 discusses managerial implications. Concluding remarks are in Section 7.

2 Literature Review

Dynamic Programming and Reinforcement Learning

The core idea of reinforcement learning (RL) is deriving solutions to stochastic dynamic programming problems under demand uncertainty in which the firm needs to learn about consumer preferences and trade off instantaneous cost with future payoff with the goal of maximizing long-term profit contribution. 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. There exist a stream of marketing research that derives and studies the properties of this problem in various applications of marketing decision support system. With application to catalogers, 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 maximizes profit over the long-term relationship with customers. Li, Sun, and Montgomery (2006) derive an optimal multi-step, multi-segment, and multi-channel cross-selling campaign process that tells 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 explicitly 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). 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 a index-based heuristic solution can perform nearly optimal and significantly better than myopic learning.

Recently, machine learning approaches are adopted to solve the DP problems and RL are applied to marketing problems by computer and data scientists with the same ideas

of continuously following consumers and predicting the next purchasing decisions of the target consumers and deliver the right message 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 historical activities of a user and the article recommendation policy adapts based on the user's 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 a RL problem to maximize lifetime value (LTV) and show the 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 optimal allocation of advertisement. 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. These studies show that by adaptively learning and adjusting participant assignment, 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).

Existing research based on dynamic programming and RL approaches are mostly empirical and provide specific applications to demonstrate that learning and acting on customer responses enable a firm to make more proactive and customized marketing decisions that reduce costs, increase consumer demand, and/or improve customer retention, and hence improve firms' long-term profit. The booming empirical research and/or algorithm development call for analytical studies to systematically investigate the properties of firms' real-time learning and acting, and the resulting inter-temporal trade-off between exploitation and exploration in algorithm-based, real-time decision making.

Analytical Modelling

From a modelling perspective, our research is related to the literature in economics that model learning and experimentation as multi-armed bandit (MAB) problems (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 under-

experiment as they try to free-ride on information from others' experiments. In the contexts of experience goods and labor market, respectively, Bergemann and Välimäki (1996) and Felli and Harris (1996) study the 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 the case where multiple agents face MAB problems while competing with each other for experiments, which is studied in this paper.

This problem of competition and MAB is also receiving attention in computer science. The papers closest to ours are Aridor et al. (2020) and Mansour, Slivkins, and Wu (2018). Both papers also 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) and Mansour, Slivkins, and Wu (2018) are short-lived and cannot observe other users' experience. In contrast, firms in our model face a population of long-lived users who are also solving MAB problems when choosing which firm to visit over time. Such differences allow us to investigate how firms' equilibrium strategies depend on users' long-term behaviors, which is absent in Aridor et al. (2020) and Mansour, Slivkins, and Wu (2018). Differences in our models also lead to very different conclusions. Aridor et al. (2020) and Mansour, Slivkins, and Wu (2018) 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, although it involves less exploration than the optimal algorithm of a monopoly. The impact of competition on consumer welfare also differs as a result.

A recent literature in economics and marketing builds 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 shows that human judgment can either complement or substitute machine prediction. Agrawal, Gans, and Goldfarb (2018b) consider subscription pricing of such prediction technology. Miklos-Thal and Tucker (2019) and Hansen, Misra, and Pai (2020) consider collusion between algorithms. 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 personal pricing when data analytics enable firms to have more information on consumer preferences

than consumers themselves. These papers focus on documenting the general impact of machine-aided decision-making without investigating the inter-temporal trade-off between exploitation and exploration.

Methodologically, we use a continuous-time model with sequential arrival of information. Such a model approximates the nature of real-time learning and acting on customer information. There is a related 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 the 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 total donation arrives over time. In contrast to the previous papers, the current paper features competition between two firms, each deciding its own experimentation strategy and receiving private information, while factoring in competitive responses by the other firm.

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 intensify price competition. Zhang (2011) expands the literature to allow for endogenous product designs which influence the information that firms collect. The current paper do not consider pricing. Instead, we focus our attention on personalized product offerings. We allow for rich dynamics where each firm makes personalized offerings over 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, with the crucial difference that personalization allows firms to actively compete for every consumer.

The paper also relates to the literature on the relationship between competition and innovation. Past theoretical literature is ambiguous on the relationship between the number of firms and innovation incentives. For example, 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 papers consider a very different type of innovation. We consider the technological upgrade from myopic algorithms to forward-looking algorithms, and show that competition decreases the return from such upgrade when consumers are impatient.

3 Monopoly Model

Consider a market in which users consume content over time. A monopolistic firm provides personalized content to each user. For example, platforms like Youtube, Tiktok, Spotify, and Google News recommend personalized content to users based on their historical behaviors on the site. The main objective of their content recommendation algorithms is to increase user engagement with the content on the site, which boosts monetization, such as advertising revenue, through increased views over a user's lifetime.[1] In this paper we abstract away from the advertising and pricing decisions, and focus solely on the recommendation decisions.

In this section, we first propose a dynamic model of content recommendations that captures the essential exploration vs. exploitation trade-off. For a given user, 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 $N1$ and $N2$.

Different types of content differ both in their intended audience as well as heterogeneity among users. All users enjoy mass-market content equally, but for niche-market content, they have different preferences. Some users enjoy $N1$ more often than $N2$, and vice versa. Let $T \in \{N1, N2\}$ denote the focal user's preferred type of niche-market content. This is drawn by Nature and is unknown to the firm.

Let $S_t \in \{M, N1, N2\}$ denote the type of content that the firm recommends to the focal user at time t. At a given time period t, each user has a unit demand for content. If a user is recommended niche-market content, she 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 is recommended mass-market content, she likes the content with probability

[1]There are various ways to display ads and generate revenue (pay per click / pay per impression), which 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, only customization of content.

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

So the probability that the user 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 WLOG.[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:

$$\pi_t = p * y(T, S_t) = y(T, S_t)$$

Let N_t be an indicator function that equals 1 if the firm recommends niche-market content to the user at time t. Let Y_t denote the firm's cumulative profit from the user after t periods. We then have:

$$E[Y_t] = \sum_{s=1}^{t} y(T, S_s) \quad \text{and} \quad Var(Y_t) = \alpha(1-\alpha) \sum_{0}^{t} N_t + c(1-c) \sum_{0}^{t} (1 - N_t) \tag{2}$$

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.

To capture the idea that these interactions are happening at a high frequency and the firm can monitor a user's behavior continuously, we use a continuous-time approximation of the cumulative profit, while maintaining the same expected value and variance from equation (2). The unique continuous-time process with independent noise in increments and that satisfies $Y_t \sim \mathcal{N}(E[Y_t],\ Var(Y_t))$, where

$$E[Y_t] = \int_0^t y(T, S_s) ds \quad \text{and} \quad Var(Y_t) = \alpha(1-\alpha) \int_0^t N_s ds + c(1-c) \int_0^t (1 - N_s) ds \tag{3}$$

[2]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. We find that the monopoly monetizes the content less under the forward-looking algorithm than under the myopic algorithm. The firm should also increase monetization as its knowledge about individual users' preferences improve.

is

$$dY_t = y(T, S_t)dt + \sigma_t dW_t, \tag{4}$$

where $y(T, S_t)$ as defined in equation (1) is the expected profit flow, $\sigma_t = \sqrt{\alpha(1-\alpha)N_t + c(1-c)(1-N_t)}$ represents the instantaneous standard deviation in profit, and the process W_t is a standard Wiener process.

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 $N1$ with probability λ_0, or has a prior belief that the user prefers $N2$ 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 $N1$ with probability λ_0.

Let λ_t denote the firm's *posterior* belief that the user prefers $N1$ over $N2$. The history of realized profit from the user serves as the information source. We have

$$\lambda_t = Pr(T = N1|F_t)$$

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 so that the user is more likely to like a mass-market content than any 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 choice between niche-market and mass-market content captures the trade-off between exploration and exploitation in a simple way.

Updating of Posterior Belief

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

$$y(\lambda_t, S_t) = E[y(T, S_t)|F_t] = \begin{cases} \lambda_t\alpha + (1-\lambda_t)(1-\alpha) & \text{if } S_t = N1 \\ (1-\lambda_t)\alpha + \lambda_t(1-\alpha) & \text{if } S_t = N2 \\ c & \text{if } S_t = M \end{cases} \tag{5}$$

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 serves niche-market content, follows the process

$$d\lambda_t = [\alpha - (1-\alpha)]\frac{\lambda_t(1-\lambda_t)}{\sigma_t^2}[y(T, S_t) - y(\lambda_t, S_t)]dt + [\alpha - (1-\alpha)]\frac{\lambda_t(1-\lambda_t)}{\sigma_t}dW_t \tag{6}$$

where the term $[y(T, S_t) - y(\lambda_t, S_t)]$ represents the new information, which is the difference between expected flow profit under the current belief and the true expected flow profit. The speed of updating is weighted by the difference in outcome between the right and the wrong action, which is captured by $\alpha - (1-\alpha) = 2\alpha - 1$. The term σ_t is the standard deviation of flow profit from equation (4).

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

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

We can then simplify equation (6) to

$$d\lambda_t = \frac{\sigma(\lambda_t)}{\sigma_t^2}[y(T, S_t) - y(\lambda_t, S_t)]dt + \sigma(\lambda_t)dW_t \tag{7}$$

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 have more varied tastes. As a result, they exhibit more varied responses to niche-market content. 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. The firm updates its belief about the user's preferred content type as it serves the user over time and observes the user's response to the

content. 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 time, there will be some amount of uncertainty regarding the user's preferences.

3.2 Firm's Decisions

The firm is risk neutral and maximize the present value of discounted expected profits with a discount rate of r by choosing recommendations S_t as a function of λ_t, which is the belief at time t that the user prefers type $N1$ over $N2$. 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

$$V_t(\{S_t\}, \lambda_0) = E \int_0^\infty e^{-rt} y(\lambda_t, S_t) dt$$

where $y(\lambda_t, S_t)$ is the expected flow profit defined in equation (5).

The firm's problem is to find an 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

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

for some coefficient b.

Consider the three types of algorithms discussed in the introduction. Here we describe them separately.

Non-adaptive Recommendation Algorithm

Assume that the firm does not have the capacity to track users' behaviors. The optimal strategy is to recommend based on the initial information about the user. The firm then never adjusts this recommendation. This is personalization based on static information used for segmentation, such as demographics. This is akin to empirical models in which decisions are made based on insights found in past data at a specific time, but with no real-time data tracking and updating of decisions, hence not adaptive.

The firm should recommend content type $N1$ over type M if lifetime expected profit from type $N1$, $\frac{\lambda_0\alpha+(1-\lambda_0)(1-\alpha)}{r}$, if greater than lifetime expected profit from type M, $\frac{c}{r}$. Let $\lambda^* = \frac{c-(1-\alpha)}{2\alpha-1}$ denote the break-even point.

The optimal non-adaptive algorithm is the following: the firm recommends content type $N1$ if $\lambda_0 > \lambda^*$, type $N2$ if $\lambda_0 < 1-\lambda^*$, and type M if $\lambda_0 \in [1-\lambda^*, \lambda^*]$.

Myopic Recommendation Algorithm

Now consider a firm that can track users' behaviors and update beliefs about their preferences in real time. The firm employs a myopic recommendation algorithm, which only aims to maximize the probability that a user likes the next recommendation. This resembles a firm with a supervised learning algorithm that continuously predicts the likelihood a user enjoys each piece 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 (5). The optimal myopic algorithm is the following: the firm recommends content type $N1$ if $\lambda_t > \lambda^* = \frac{c-(1-\alpha)}{2\alpha-1}$, type $N2$ if $\lambda_t < 1-\lambda^*$, and type M if $\lambda_t \in [1-\lambda^*, \lambda^*]$.

Note that the threshold λ^* is the same threshold from the non-adaptive algorithm. The myopic algorithm is different from the non-adaptive algorithm because the firm will adjust its recommendations over time based on a user' evolving history. As λ_t crosses the threshold λ^*, the firm's recommendation changes. However, because the myopic algorithm only seeks to maximize instantaneous profit, it focuses entirely on exploitation while ignoring exploration in the trade-off.

Forward-Looking Recommendation Algorithm

Now we solve the optimal forward-looking algorithm. It 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 serves niche-market content, it must serve type $N1$. Also as noticed earlier, once the firm serves mass-market content to the user, it stops learning, so it will always serve mass-market content in the future. Thus the firm's value function when it recommends mass-market content must be c/r. 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 $\lambda_t \leq \hat{\lambda}$. This is equivalent to an optimal stopping problem, where serving mass-market content is equivalent to a stopping option that gives a payoff of c/r. The optimal stopping threshold $\hat{\lambda}$ must satisfy the value-

matching and the smooth-pasting conditions (see, e.g., Dixit 1993) for the value function:

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

Plugging these boundary conditions back into the solution of the HJB equation (8) 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:

Proposition 1 *Define*

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

1. *If $\hat{\lambda} > 0.5$, then the optimal forward-looking algorithm recommends content type $N1$ if $\lambda_t > \hat{\lambda}$, type $N2$ 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 $\lambda_{it} > \hat{\lambda}$

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

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

2. *If $\hat{\lambda} \leq 0.5$, then the optimal forward-looking algorithm recommends $N1$ if $\lambda_t > 0.5$ and $N2$ if $\lambda_t < 0.5$.*

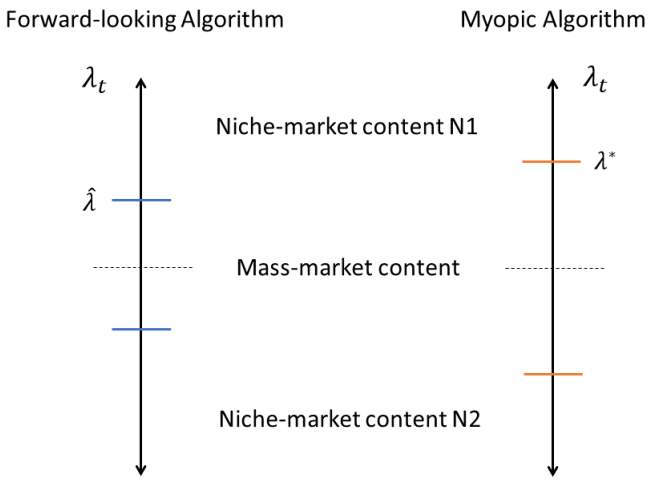
   The firm's value function is symmetric around 0.5 where for $\lambda_t > 0.5$,

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

Corollary 1.1 *The optimal threshold for the forward-looking algorithm, $\hat{\lambda}$, is strictly lower than the threshold in the myopic and non-adaptive algorithms, $\lambda^* = \frac{c-(1-\alpha)}{2\alpha-1}$.*

The implication of $\hat{\lambda} < \lambda^*$ is that the firm serves 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 $N1$, which is expected to generate less immediate 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 forward-looking vs. myopic algorithms



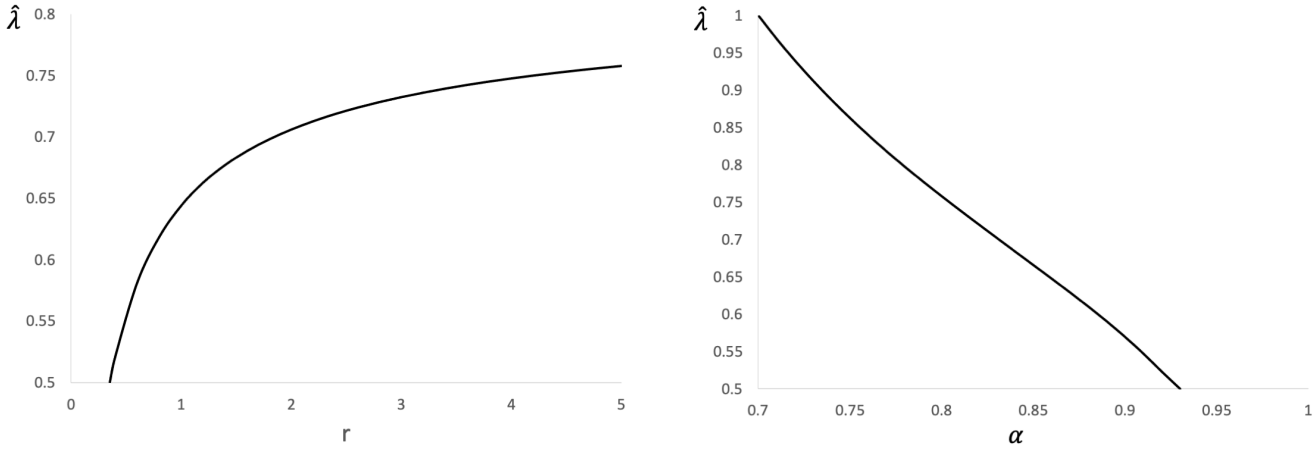
In Figures 2a and 2b, we plot $\hat{\lambda}$ as a function of r and α, respectively. Notice that $\hat{\lambda}$ is an increasing function of γ, while γ is an increasing function of r and a decreasing function of α. Thus, $\hat{\lambda}$ increases with r and decreases with α. Intuitively, when r is smaller, future profit becomes more important and thus it becomes more important for the firm to learn and adapt. Consequently, the firm is less likely to recommend mass-market content to a user. 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.

Corollary 1.2 *The optimal threshold $\hat{\lambda}$ increases with discount rate, r, and decreases with preference consistency, α.*

3.3 Value from Advanced Algorithms

Different recommendation algorithms require different levels of technology. An upgrade from the non-adaptive algorithm to the myopic algorithm requires learning and acting on users' behaviors over time. For example, a supervised-learning-based algorithm that predicts a user's likelihood to engage with different content and recommends the highest ranked content in real-time is a myopic algorithm, but it ignores the effect of the current decision on future information. 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 these technological upgrades.

Figure 2: The optimal threshold



(a) as a function of the discount rate for $\alpha = 0.8$, $c = 0.7$

(b) as a function of preference consistency for $c = 0.7$ and $r = 0.6$

This then can be interpreted as the monopoly's incentive to invest in such upgrades if they are costly. We denote the firm's ex-ante expected profits under the non-adaptive, the myopic, and the forward-looking algorithms as $V^{NA}(\lambda_t)$, $V^{MY}(\lambda_t)$, and $V^{FL}(\lambda_t)$, respectively.

The Additional Value from the Myopic Algorithm

First, consider the value of upgrading from the non-adaptive algorithm to the myopic algorithm. Under the non-adaptive algorithm, the firm always serves 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

$$V^{NA}(\lambda_0) = \begin{cases} \frac{\lambda_0\alpha+(1-\lambda_0)(1-\alpha)}{r} & \text{for } \lambda_0 > \lambda^* \\ \frac{c}{r} & \text{for } \lambda_0 \in (1-\lambda^*, \lambda^*) \\ \frac{\lambda_0(1-\alpha)+(1-\lambda_0)\alpha}{r} & \text{for } \lambda_0 < 1-\lambda^* \end{cases} \tag{9}$$

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 (8) with the boundary condition $rV(\lambda^*) = c$, from which we get $b_2 = 0$. Thus, the firm's expected profit at $t = 0$ is also

$$V^{MY}(\lambda_0) = \begin{cases} \frac{\lambda_0\alpha+(1-\lambda_0)(1-\alpha)}{r} & \text{for } \lambda_0 > \lambda^* \\ \frac{c}{r} & \text{for } \lambda_0 \in (1-\lambda^*, \lambda^*) \\ \frac{\lambda_0(1-\alpha)+(1-\lambda_0)\alpha}{r} & \text{for } \lambda_0 < 1-\lambda^* \end{cases} \tag{10}$$

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

Corollary 1.3 *The additional value from the myopic algorithm, $V^{MY}(\lambda_0) - V^{NA}(\lambda_0)$, is zero for all λ_0.*

The strong result of Corollary 1.3 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 5). 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 the 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 recommend 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.[3]

This analysis shows that, for a monopoly, simply recommending the "best" content at the moment could make the effort to learn the user's preferences fruitless. Exploitation of historical information alone may not provide value. Many recommender systems have supervised learning algorithms that can predict the likelihood that a user enjoys each content. This result highlights the inadequacy of recommending myopically based on this ranking. However, it is important to note that Corollary 1.3 is only for a monopoly, which does not need to worry about losing users to a competitor.

The Additional Value from the Forward-Looking Algorithm

Next, we consider the value of upgrading from the myopic algorithm to the forward-looking algorithm that balances the trade-off between exploration and exploitation. The value function $V^{FL}(\lambda_{it})$ is given in Proposition 1, from which we can get, for $\lambda_0 \geq \lambda^*$,

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

[3]If, instead, the expected profit flow from recommending N1 is concave in λ_t, then the net present value of always recommending N1 becomes a super-martingale 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. Section 5.1 provides an example where the existence of an outside mass-market content provider causes the additional value from the myopic algorithm to be positive. 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 N2, then the flow payoff can be made to be convex in λ_t.

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

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

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

Corollary 1.4 *The additional value from the forward-looking algorithm,* $V^{FL}(\lambda_0) - V^{MY}(\lambda_0)$*, is strictly positive for* $\lambda_0 \notin [1-\hat{\lambda}, \hat{\lambda}]$*, and is highest at* $\lambda_0 = \lambda^*$*. It increases in* λ_0 *for* $\lambda_0 < 1 - \lambda^*$ *and* $0.5 < \lambda_0 < \lambda^*$ *but decreases in* λ_0 *for* $1 - \lambda^* < \lambda_0 < 0.5$ *and* $\lambda_0 > \lambda^*$*. The value decreases with* r *and increases with* α*.*

In Figure 3, we plot the additional value from the forward-looking algorithm as a function of λ_0. The forward-looking algorithm creates positive value when $\lambda_t \in (\hat{\lambda}, \lambda^*)$, even though the algorithm recommends niche-market content which is less likely to match user preferences than mass-market content. 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 $\lambda < \lambda^*$.

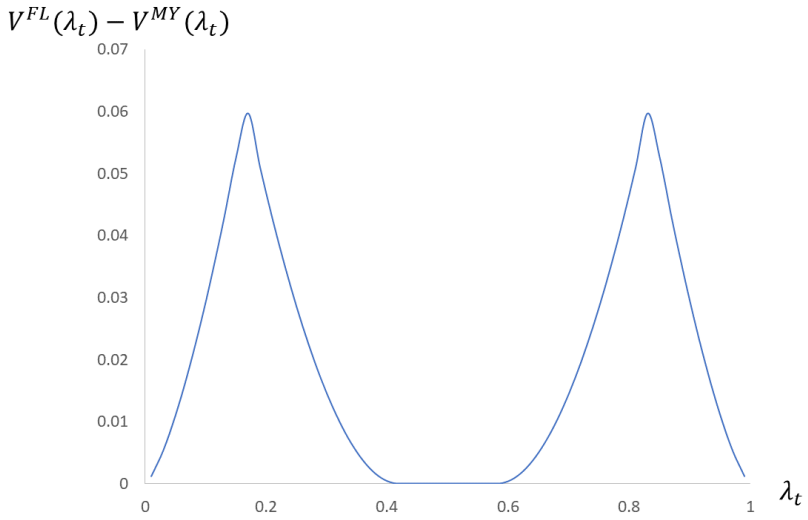
The additional value from the forward-looking algorithm peaks at $\lambda = \lambda^*$. Intuitively, the forward-looking algorithm and the myopic algorithm begin to diverge at λ^*. Under the myopic algorithm, users with $\lambda_t = \lambda^*$ are served mass-market content. Under the forward-looking algorithm, the firm can trade off short-term losses for more information on a user's preferences. This result has implications for when firms should prioritize investing in forward-looking algorithms. Conventional wisdom may suggest that firms should prioritize learning when there is higher uncertainty, but this is not always true. Corollary 1.4 shows that the benefit of the optimal algorithm is non-monotonic in the firm's knowledge about users.

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 in the population and describe the evolution of the firm's recommendations under the myopic and the forward-looking algorithms. We briefly describe our results here.

Proposition 2 *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*

Figure 3: Additional value from the forward-looking algorithm



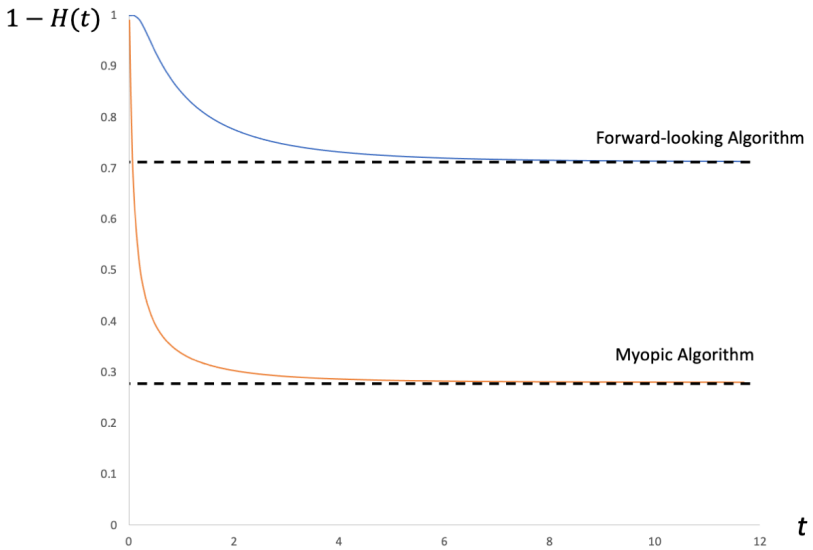
for $\alpha = 0.8$, $c = 0.7$ and $r = 0.6$

$\frac{\lambda_0 - \lambda^*}{1-\lambda^*}$ *fraction of users under the myopic algorithm. Both fractions decrease with* r *and increase with* α*. As* $t \to \infty$*, all users who are recommended niche-market content receive the content types that they prefer.*

We illustrate the evolution of the fraction of users who are recommended niche-market content in Figure 4. Notice that under both the forward-looking and the myopic algorithms, the fraction of users recommended niche-market content decreases and converges to a constant in the long-term steady state. This fraction decreases with discount rate r and increases with preference consistency α. Intuitively, with a bigger α, the firm cares more about learning users' preferences, and with a smaller r, the firm cares more about the long-term profit, so the steady-state amount of niche-market content increases. A more forward-looking algorithm serves more niche-market content both in the short-term and in the long-term, potentially with an increasing gap over time.

In the Appendix, we also track the evolution of expected flow profit and expected cumulative profit over time under different algorithms. We show that compared to the non-adaptive algorithm or 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 non-adaptive algorithm or the myopic algorithm in the long run. In such a case, it is important to know the evolution of profits for the firm for two reasons. First, the firm can know the approximate duration of financial losses before the implementation of such an algorithm is profitable. Second, the firm can identify

Figure 4: Fraction of Niche-Market Content Recommended



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

the maximum loss to make financial plans accordingly. In Figures 5a, we plot the expected flow profit under different algorithms. In Figure 5b, we plot the the difference in expected cumulative profit between the forward-looking and the myopic algorithms.

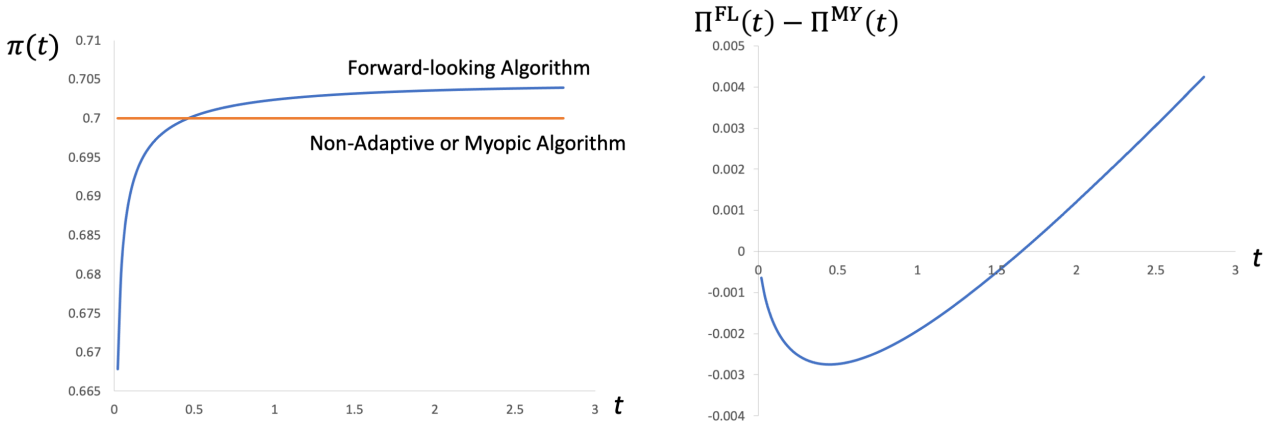
4 Duopoly Model

In this section, we study firm's optimal recommendation algorithms under competition. We explore how competition affects the exploration vs. exploitation trade-off, firms' incentives to develop forward-looking algorithms, and the impact of algorithms on welfare.

We expand the monopoly model to allow for two firms. There are two firms, firm 1 and firm 2, that provide content to users. At $t = 0$, both firms simultaneously choose their recommendation algorithms, which are functions mapping information sets to content types. After firms choose their recommendation algorithms, at each t, the user chooses to visit one of the two firms. Both firms observe the user's platform choice, but cannot observe a user's experience with 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 among three types of content: mass-market content and two types of niche-market content. Niche-market content from the two firms are different, so that there are a total of four types of niche-market content. Let

Figure 5: 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 $r = 0.6$

N_1^1 and N_2^1 denote two types of niche-market content on firm 1's platform, and let N_1^2 and N_2^2 denote the two types of niche-market content on firm 2's platform. For simplicity, and to capture the fact that different platforms carry different content, we assume the user's preferences for niche-market content types are independent between the two platforms.

The user's expected flow utility from seeing content $S_t^j \in \{M, N_1^j, N_2^j\}$, which represents the content type recommended by firm j at time t, is

$$u(T^j, S_t^j) = \begin{cases} \alpha & \text{if} \quad S_t^j = T^j \\ c & \text{if} \quad S_t^j = M \\ 1-\alpha & \text{otherwise} \end{cases} \tag{11}$$

where $T^j \in \{N_1^j, N_2^j\}$ represents the user's preferred niche-market content type 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:

$$y^j(T^j, S_t^j) = \begin{cases} \alpha & \text{if} \quad S_t^j = T^j \\ c & \text{if} \quad S_t^j = M \\ 1-\alpha & \text{otherwise} \end{cases} \tag{12}$$

The user has a discount rate of r_u, and both firms have a discount rate of r_f. We focus on the case of $r_u \geq r_f$, so that users are less patient than the firms.[4]

[4]This is motivated by the observation that online users often exhibit short attention spans and would

Information and 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 that firms can start with different amount 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.

With competition, we need to model how the user learns, which then determines her platform choices. Comparing to the monopoly model, we need to make one additional assumption. We assume that the user only observes whether she likes the recommended content, or her utility from the recommended content, but cannot directly observe content type. 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. Then firm 1 knows that the recommended content type is correct. If the user observes that firm 1 recommends the wrong niche-content type, then the user visits the competitor in the next period. Then firm 1 knows that the recommended content type is wrong. Either way, the user's preferred content type is immediately revealed to firm 1, so there is no more learning.[5]

Assumption 1 *The user does not observe content type.*

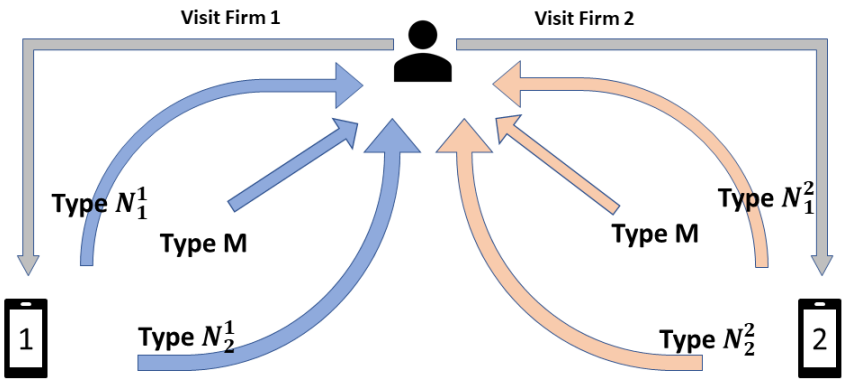
Under this assumption, the user has to update her belief on how well each firm's next recommendation will be based on past consumption experiences. Intuitively, if a users enjoyed recent videos recommended by YouTube, then her expectation for the next recommendation from YouTube increases. However, if YouTube recommended multiple videos that she did not enjoy, then she would have a lower expectation on the quality of the next recommendation, and may switch to 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. Given that

quickly abandon sites or content that do 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 examine how the two discount rates separately affect the equilibrium outcome.

[5]Such unravelling doesn't 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 need to learn about their own preferences over time as they consume various types of content.

Figure 6: A Multi-agent Bandit Problem



when the user consumes content from firm j, both the user and firm j receive the same information, the value of the user's posterior belief that firm j's inference is correct is also the value of firm j's posterior belief, λ_t^j.

Multi-Agent Bandit Problem

This duopoly model is a multi-agent bandit problem, in which three agents have to decide which "arm" to pull, while taking into consideration the other two agents' strategies. Each firm has to decide what content to recommend, and users decide 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 6 gives an intuitive illustration of the setup.

A solution to the problem has to simultaneously solve all three players' bandit problem. To solve the problem, we first characterize the user's optimal choice rule when presented a menu of content created by the two firms' recommendation algorithms. Then taking user behavior as given, we look for Nash equilibrium of the time 0 game in which the two firms simultaneously choose recommendation algorithms. Finally, we confirm that the user's choice rule is optimal under the equilibrium recommendation strategies. Then we can confirm that the equilibrium we characterize indeed solves all three agents' dynamic optimization problem simultaneously.

User's Behaviors

The user optimally chooses which firm to visit taking each firm's recommendation algorithm as given. When making choices, the user correctly anticipates whether she will be recommended niche-market or mass-market content upon visiting a firm.

First, consider the user's preferences when she can only choose between niche-market content from firm j and mass-market content from the other firm. Her continuation value from consuming niche-market content from firm j can be derived similar to the monopoly's value function from equation (21):

$$U(\lambda_t^j) = \frac{u(\lambda_t^j, S_t^j)}{r_u} + \frac{(\lambda_t^j)^2(1-\lambda_t^j)^2(2\alpha-1)^2}{2r_u\alpha(1-\alpha)}U''(\lambda_t^j) \tag{13}$$

This is a bandit problem with a stopping option (mass-market content). Solving for the user's optimal choice of between niche-market content and mass-market content is similar to solving for a monopoly's optimal recommendation. Using our results from 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

$$\widehat{\lambda_u} = \frac{[c-(1-\alpha)](\gamma_u-1)}{(2\alpha-1)(\gamma_u-1)+2(\alpha-c)} \qquad \text{where} \quad \gamma_u = \sqrt{1+\frac{8r_u\alpha(1-\alpha)}{(2\alpha-1)^2}} \tag{14}$$

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

Next, consider the user's preferences when she can only choose between 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. See Bank and Küchler (2007) for a derivation of Gittins index theorem in continuous time. The Gittins index for niche-market content from firm j at a given λ_t^j, $G(\lambda_t^j)$, is equivalent to a fixed flow payoff such that if the user can only choose between niche-market content from firm j and this fixed flow payoff, she switches to this fixed flow payoff exactly at λ_t^j. We can thus use equation (14) to solve for $G(\lambda_t^j)$, by replacing c with $G(\lambda_t^j)$. We then have the following implicit equation:

$$\lambda_t^j = \frac{[G(\lambda_t^j)-(1-\alpha)](\gamma_u-1)}{(2\alpha^j-1)(\gamma_u-1)+2[\alpha-G(\lambda_t^j)]} \qquad \text{where} \quad \gamma_u = \sqrt{1+\frac{8r_u\alpha(1-\alpha)}{(2\alpha-1)^2}} \tag{15}$$

Note that the right-hand side of equation (15) increases in $G(\lambda_t^j)$; thus $G(\lambda_t^j)$ is an increasing function of λ_t^j.

Finally, the user must be indifferent between mass-market content from the two firms. In this case, we assume that she randomly visits one of the two platforms.

Firms' Problem

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

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:

$$\begin{cases} D_t^1(\lambda_t^1, \lambda_t^2 | N, N) = \mathbb{I}\{G(\lambda_t^1) \geq G(\lambda_t^2)\} \\ D_t^1(\lambda_t^1, \lambda_t^2 | N, M) = \mathbb{I}\{\lambda_t^1 > \widehat{\lambda_u}\} \\ D_t^1(\lambda_t^1, \lambda_t^2 | M, N) = 1 - \mathbb{I}\{\lambda_t^2 > \widehat{\lambda_u}\} \\ D_t^1(\lambda_t^1, \lambda_t^2 | M, M) = \frac{1}{2} \end{cases} \tag{16}$$

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

Given the demand function, we can write firm j's expected flow profit from the user as:

$$\pi_t^j = y^j(\lambda_t^j, S_t^j) D_t^j dt$$

For a given pair of recommendation path, $(\{S_t^1\}, \{S_t^2\})$, the expected lifetime value of the user for firm 1 is

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

The firm's problem is to find an 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 set 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 $S^j(I_t^j) \in \{N_1^j, N_2^j, M\}$.

[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 the niche-market content.

In the monopoly model, 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 for such an equilibrium, if it exists, we do not need to calculate firm j's belief of the user's state on the competitor's platform, which can only updated from firm j's partial observation of the user's past behavior on its own platform. In such an equilibrium, no matter what firm j's belief of the user's state on the competitor's platform is, there is a weakly dominant action to take.

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. We then confirm that the equilibrium outcome maximizes the user's utility, hence a solution to the multi-agent bandit problem. The proof is presented in the Appendix.

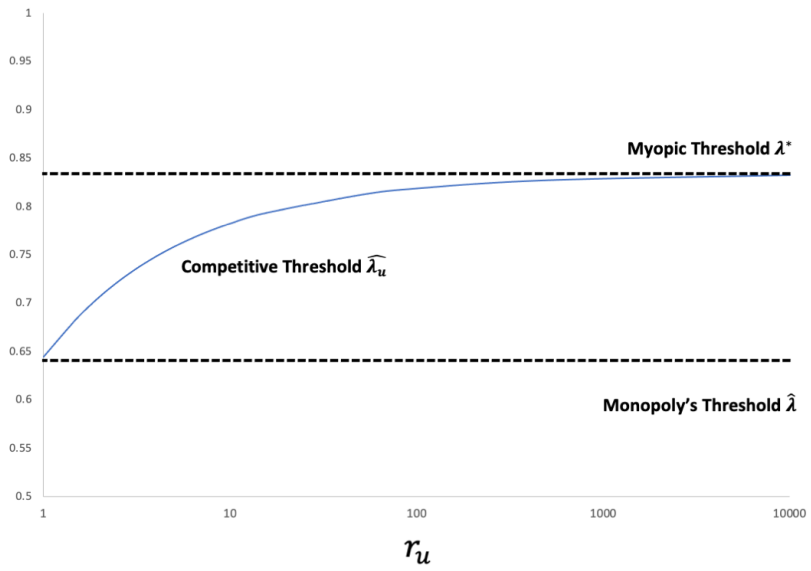
Proposition 3 *In a duopoly, firm j serves niche-market content if and only if $\lambda_t^j \notin [1-\widehat{\lambda_u}, \widehat{\lambda_u}]$. This is the unique stationary algorithm, characterized by $S_t^j = S^j(\lambda_t^j)$, that constitutes Nash equilibrium for all priors λ_0^1 and λ_0^2. Firms' equilibrium recommendations maximize the user's utility.*

Note that the equilibrium content choices maximize the user's welfare. 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 $G(\lambda_t^1) > \sup\{G(\lambda_t^2), c\}$, niche-market from firm 2 if $G(\lambda_t^2) > \sup\{G(\lambda_t^1), c\}$, and mass-market content if $c \geq \sup\{G(\lambda_t^1), G(\lambda_t^2)\}$. Note that this is exactly the user's content choice in equilibrium. In equilibrium, firm j offers niche-market content only when $\lambda_t^j > \widehat{\lambda_u}$ which is equivalent to $G(\lambda_t^j) > c$ by equations (14) and (15). Thus the user's most preferred content type is always available to her.

Given that, in equilibrium, both players (the two firms) also maximize their utility conditional on the user's and the competitor's actions, we can conclude that Proposition 3 is

[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.

Figure 7: The optimal threshold as a function of the user's discount rate



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

a solution to the multi-agent bandit problem, where each agent optimally solves its bandit problem conditional on the other two agents' strategies.

Figure 7 shows firms' equilibrium threshold as a function of the user's discount rate r_u, and compare them to the the myopic threshold and monopoly's forward-looking threshold derived from Section 3. Firms' optimal recommendation algorithms fall between the monopoly's forward-looking algorithm and the myopic algorithm. The forward-looking algorithm under competition recommends less niche-market content than under monopoly because $\widehat{\lambda_u}$ is higher than monopoly's threshold when $r_u > r_f$. Competition pushes firms away from exploration and towards exploitation. Also $\widehat{\lambda_u}$ is increasing in r_u but does not depend on r_f. Thus, firms are forced to be less forward-looking and recommend less niche-market content to prevent impatient users from switching.

From equation (14), we can see that, as $r_u \to \infty$, $\widehat{\lambda_u}$ approaches the myopic threshold, λ^*. As $r_u \to r_f^+$, $\widehat{\lambda_u}$ approaches the monopoly's forward-looking threshold, $\widehat{\lambda}$.

Corollary 3.1 *In a duopoly, firms recommend less niche-market content than a monopoly does if the user has a higher 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' own discount rate.*

Intuitively, when r_u is higher, users are more myopic in their content preferences. Con-

sequently, 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, who is more patient than the user, values the information collected. However, if a competitor recommends mass-market content to users visiting its platform, then the firm that recommends niche-market content will lose demand. If the competitor is offering niche-market content, then the firm can steal demand by offering mass-market content. 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. However, when competing for users' attention, the value from exploration is muted if users are less patient and can take their demand elsewhere.

As Figure (7) shows, when users become more myopic, firms' optimal forward-looking algorithms also approach their myopic algorithms. In the limit as $r_u \to \infty$, the optimal forward-looking threshold $\widehat{\lambda_u}$ converges to the myopic threshold, $\frac{c-(1-\alpha)}{2\alpha-1}$. Thus, when users are myopic, the myopic algorithm itself is optimal. This implies that, when facing myopic users, the forward-looking algorithm provides no extra value than the myopic algorithm, regardless of whether the competitor has the forward-looking algorithm or the myopic algorithm.

Corollary 3.2 *As $r_u \to \infty$, the equilibrium forward-looking algorithm under competition converges to the the myopic algorithm, and the additional value from the forward-looking algorithm converges to zero.*

Comparing Corollary 3.2 to Corollary 1.4, we see that the presence of a competitor decreases the value from the forward-looking algorithm if the users' discount rate is sufficiently high. This also implies that competition lowers firms' incentives to invest in the technological upgrade from the myopic algorithm to the forward-looking algorithm when users are sufficiently impatient.

Note that from Proposition 3, we know that the equilibrium outcome under the competitive threshold, $\widehat{\lambda_u}$, maximizes the user's utility. This implies that, the impact on user welfare of both firms upgrading technology from the myopic algorithm to the forward-looking algorithm must be positive. However, note that the monopoly threshold, $\widehat{\lambda}$, does not depend on the user's time preferences. The monopoly's forward-looking algorithm is not optimal for the user unless $r_u = r_f$. A monopoly with the forward-looking algorithm recommends too much niche-market content with respect to user welfare. If the user is sufficiently myopic, as $r_u \to \infty$, technological upgrade from the myopic algorithm to the forward-looking algorithm actually decreases user welfare. Thus the development of the forward-looking algorithm may

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

Our findings illustrate that competition lowers the optimal level of exploration in the exploration vs. exploitation trade-off. In the case of sufficiently impatient or myopic users, competition also presents a new trade-off between incentives in technological upgrade and the effect of such upgrade on consumer welfare. We summarize our findings conceptually in Figure 8.

Figure 8: Impact of Competition When Users are Impatient

	Optimal level of exploration	Firms' incentives in algorithmic upgrade	Effect of upgrade on user welfare
Monopoly	High	High	Negative
Duopoly	Low	Low	Positive

5 Asymmetric Extensions

In Section 4, we studied competition between two firms with symmetric technology. In this section, we consider extensions where firms have asymmetric capabilities to learn and recommend.

5.1 Against a Mass-Market Content Provider

A focal firm that recommends content competes with a traditional content provider that only serves mass-market content. This is also equivalent to adding an outside option to the monopoly model. At each moment, the user chooses to visit either the recommendation firm or the mass-market content provider.

Consider the user's preferences between niche-market content and mass-market content. The user's optimal content choice is marked by the threshold $\widehat{\lambda_u}$ from equation 14. She prefers niche-market content from the focal firm for $\lambda_t > \widehat{\lambda_u}$ or $\lambda_t < 1 - \widehat{\lambda_u}$, and prefers mass-market content from either firm for $\lambda_t \in [1 - \widehat{\lambda_u}, \widehat{\lambda_u}]$. If both firms offer mass-market content, we assume she visit randomly.

Similar to our analysis from Section 4, one can show that the optimal forward-looking

recommendation algorithm must follow the threshold $\widehat{\lambda_u}$. The firm switches from niche-market content to mass-market content as λ_t drops to this threshold. To see this, note that the firm's profit from mass-market content is $\frac{1}{2}c$. Given our assumption that $c < 1$, the firm has no incentive to offer mass-market content as long as the demand for niche-market content is positive. However, if $\lambda_t \in [1 - \widehat{\lambda_u}, \widehat{\lambda_u}]$, then the user prefers to consume mass-market content. Given the presence of a competitor who always offers mass-market content, the firm has to recommend mass-market content in order to receive demand.

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

Note that the optimal forward-looking algorithm in this case is exactly the same as the equilibrium algorithm in the symmetric duopoly model. Thus our results from Section 4 replicate. When competing against a mass-market content provider, the firm offers less niche-market content than a monopoly does if users have a higher discount rate than the firm. The firm offers less niche-market content as users' discount rate increases. As $r_u \to \infty$, the user becomes myopic. This forces the firm to focus on exploitation. In the limit, the additional value from the forward-looking algorithm converges to zero. Under monopoly, the development of the optimal algorithm can lower user welfare. However, with the presence of an outside option, developing of forward-looking algorithms is strictly beneficial to users.

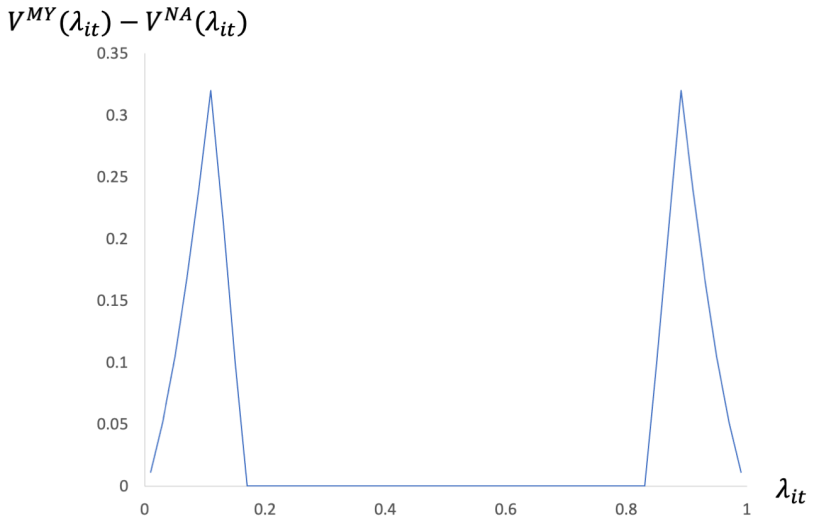
In the monopoly model, the value of upgrading from the non-adaptive algorithm to the myopic algorithm is zero. However, this is no longer true when users have an outside option. Intuitively, with the threat from an outside option, the myopic algorithm creates value by preventing the user from switching to the other platform. If the firm employs the non-adaptive algorithm, then it would lose the user forever when λ_t drops to $\widehat{\lambda_u}$. By using the myopic algorithm, the firm is able to 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 myopic algorithms even though it lowers their incentives to develop forward-looking algorithms. We calculate the additional value from the myopic algorithm in the Appendix and prove the following Corollary.

Corollary 4.1 *When competing against a mass-market content provider, the additional value from the myopic algorithm is strictly positive if* $\lambda_0 > \frac{c-(1-\alpha)}{2\alpha-1}$.

Figure 9 shows the additional value of the myopic algorithm.

Similar to Proposition 2, one can show that as $t \to \infty$, $\frac{\lambda_0 - \widehat{\lambda_u}}{1-\widehat{\lambda_u}}$ fraction of the users consume niche-market content from the recommendation firm. The rest of the users split

Figure 9: Additional value from the myopic algorithm



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

their time evenly between the recommendation firm and the mass-market content provider. Thus, the steady-state demand for the recommendation firm is $\frac{(1+\lambda_0)/2-\widehat{\lambda_u}}{1-\widehat{\lambda_u}}$, which decreases in r_u and increases in α. The steady-state demand for the traditional content provider is $\frac{(1-\lambda_0)/2)}{1-\widehat{\lambda_u}}$, which increases in r_u and decreases in α. The traditional content provider faces stronger competition from the content recommendation firm if users are more patient and their preferences are more consistent over time.

5.2 Different Speed of Learning

We modify the duopoly model by allowing the the user to have different values of α on the two platforms, so the user is more consistent in her preferences for niche-market content on one platform versus the other platform. In such a case, the platform with a higher α has a greater ability to learn the user's preferences, because the noise in the user's response, σ_t (from equation 4), decreases in α. Let α^j denote the parameter for firm j. Without loss of generality, we assume that $\alpha^1 \geq \alpha^2$, so that firm 1 has an advantage in learning over firm 2.

The user's expected utility from firm j's recommendation at time t, S_t^j, becomes

$$u_t^j = u(T^j, S_t^j) = \begin{cases} \alpha^j & \text{if} \quad S_t^j = T^j \\ c & \text{if} \quad S_t^j = M \\ 1-\alpha^j & \text{otherwise} \end{cases} \tag{17}$$

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

Consider the user's preferences between niche-market content from firm j and mass-market content from the other firm. With different α^j's, equation (14) becomes

$$\widehat{\lambda_u^j} = \frac{[c-(1-\alpha^j)](\gamma_u^j-1)}{(2\alpha^j-1)(\gamma_u^j-1)+2(\alpha^j-c)} \qquad \text{where} \quad \gamma_u^j = \sqrt{1+\frac{8r_u\alpha^j(1-\alpha^j)}{(2\alpha^j-1)^2}} \tag{18}$$

which is a different threshold for each firm.

Next, consider the user's preferences when she can only choose between niche-market content from two different firms. Equation (15) now becomes:

$$\lambda_t^j = \frac{[G^j(\lambda_t^j)-(1-\alpha^j)](\gamma_u^j-1)}{(2\alpha^j-1)(\gamma_u^j-1)+2[\alpha^j-G^j(\lambda_t^j)]} \qquad \text{where} \quad \gamma_u^j = \sqrt{1+\frac{8r_u\alpha^j(1-\alpha^j)}{(2\alpha^j-1)^2}} \tag{19}$$

which gives a different Gittins index function for each firm, $G^j(\lambda_t^j)$. Note that the right-hand side of equation (19) decreases in α^j; thus we have $G^1(\lambda) \geq G^2(\lambda)$, because by assumption, we have $\alpha^1 \geq \alpha^2$. The user prefers niche-market content from firm 1 over firm 2 when $\lambda_t^1 = \lambda_t^2$.

One can then prove that, similar to the symmetric duopoly model, there is a unique equilibrium in which firm j's recommendation is governed by a threshold. The different now is that, due to different α^j, the two firms follow different thresholds. This leads to the following modification of Proposition 3.

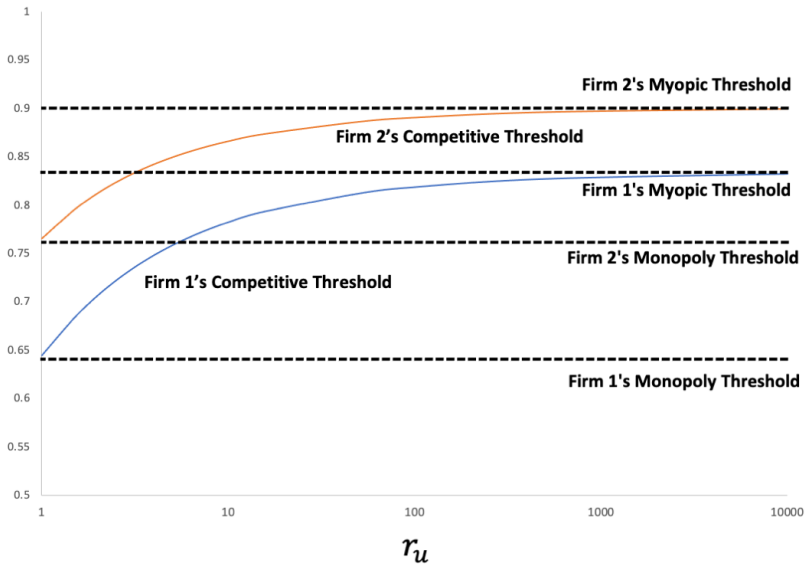
Proposition 5 *In a duopoly, firm j serves niche-market content if and only if $\lambda_t^j \notin [1-\widehat{\lambda_u^j}, \widehat{\lambda_u^j}]$.*

The fact that $\alpha^1 \geq \alpha^2$ implies $\widehat{\lambda_u^1} \leq \widehat{\lambda_u^2}$, so firm 1 recommends more niche-market content than firm 2 does. Intuitively, having a higher α means that the user's behavior is less noisy, which facilitates faster learning. A higher α also means higher profit from serving niche-market content. Both factors encourage firm 1 to serve more niche-market content than firm 2 even when the user dislikes previous recommendations.

Figure 10 shows both firms' equilibrium thresholds as functions of the user's discount rate r_u.

It then follows that the qualitative results from the symmetric duopoly model remain robust. Competition lowers the optimal level of exploration, so the optimal forward-looking algorithm under competition is between the myopic algorithm and the monopoly's optimal

Figure 10: Different optimal thresholds as functions of the users' discount rate



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

forward-looking algorithm. Competition lowers firms' incentives to invest in the technological upgrade from the myopic algorithm to the forward-looking algorithm when users are sufficiently impatient. The development of the forward-looking algorithm may lower user welfare under monopoly, but always benefits the user under competition.

6 Discussion and Managerial Implications

With companies across industries investing heavily in both hardware and software to expand their capacity to acquire, administer, and analyze large volumes of diverse data, there is much excitement about the potential to learn and act upon customer information. Complementing the extant research that focuses on the technical aspects of this phenomenon, our study analyzes the theoretical and strategic implications of such practices. Our findings have a number of important implications for managers and policy makers.

First, when making personalized and adaptive interventions, it is important for firms to take a forward-looking approach. Firms should recognize the value of information collected from observing customer responses, and customize offers to expedite this learning. Our analysis shows that in the absence of competition, the entire value of adaptive learning can come from the forward-looking aspect. For companies that are leaders in certain markets, it is important to understand the inadequacy of supervised learning-based myopic algorithms

that focus solely on exploitation of existing data and ignore forward-looking exploration, and to invest in technologies such as reinforcement learning to guide information acquisition.

However, proactive learning may imply near-term sacrifice. As our analysis shows, the optimal forward-looking algorithm recommends more personalized and niche-market offerings than a myopic algorithm does. When a company does not have much information about a customer's preferences, 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 infrastructure, it is important to recognize this implication and be prepared to tolerate worse performance in the near term. A long-term perspective is essential for the success of such initiatives.

Second, the optimal trade-off between exploration and exploitation is significantly different for companies facing competition. Managers should realize that a customer may switch back and forth between their platforms and their competitors'. Consequently, the customer's future switching behavior needs to be incorporated into the forward-looking analysis. The need to compete for customers' attention should shift the focus away from exploration and toward exploitation, especially when customers have short attention spans and seek instant gratification. Ultimately, if a customer does not like the company's offerings and switches to a competitor's platform, then the company can neither gather information nor profit from that customer. Thus, companies that do not enjoy the status of a local monopoly may have to curtail learning and make less niche-market offerings. Companies with superior learning capacities are more likely to derive competitive advantages in markets where consumers are more patient.

Third, competition may not foster investment and innovation. Instead, the incentive to invest in developing learning capacity is a complicated point. For a monopoly that does not need to compete for users' attention, the value of learning is concentrated on exploration. This incentivizes the company to develop forward-looking learning capacities. However, the value of learning is non-monotonic in the monopoly's prior knowledge about customer preferences. 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 is. Successful exploration requires tolerating sub-par offerings in the near term which results in lower engagement and lower profit.

In contrast to the monopoly case, competition makes myopic adaptive learning algorithms valuable by retaining customers from switching. However, the incentive to invest in forward-looking capabilities can be lower, and this incentive will disappear if customers have very

short attention spans and only seek immediate gratification. If users do not factor in how their current platform and consumption choices affect the future offerings, and instead only seek to maximize their immediate satisfaction, then it would be pointless for the competing companies to develop forward-looking learning infrastructures, as myopic adaptive learning is the optimal strategic choice. Competition boosts companies' incentives to develop myopic adaptive-learning capacities, but dampens their incentives to develop forward-looking ones. Understanding the complexity of the incentive to develop learning capabilities is important for both managers and policy makers.

Even though our model does not directly deal with privacy, it has implications on how algorithms and competition affect the way companies learn and act on the information collected on customers' preferences. Competing firms' ability to learn individual preferences through strategic customization depends on how forward-looking customers are. The more myopic customers act, the less information firms will be able to infer from their actions. In terms of consumer welfare, a monopoly with a myopic algorithm learns too little, while a monopoly implementing a forward-looking algorithm learns too aggressively. When customers are very impatient compared to the monopoly, the monopoly's adoption of forward-looking algorithms could hurt consumer welfare by recommending too much sub-par content in the short term in pursuit of collecting more information on user preferences. In comparison, competition forces firms to behave somewhere between the two, which benefits customers. Since the competitive pressure forces companies to align their customization strategies with the time preferences of the customers, the adoption of forward-looking learning by competing companies is always beneficial to users.

Our discussion suggest that competition can have reverse effects on innovation and consumer welfare. A firm with monopolistic power has more incentives to develop forward-looking algorithms, but such technology may hurt users by lowering near-term service quality. In contrast, competitive pressure lowers firms' incentives to develop forward-looking algorithms, even though such technology is beneficial to users. Our findings add new perspectives to the discussion on how to regulate major tech firms. In 2020, policy makers around the world have expressed increasing concern over monopoly power held by firms such as Google, Facebook, Amazon, and Apple, and the effects of such market power on innovation and consumer welfare. In July 2020, the CEOs of the four companies testified in the U.S. Congress regarding market power and alleged anti-competitive behaviors (Romm 2020). Around the same time, lawmakers and regulators in Europe proposed new laws and inquiries aimed at limiting the market power of large tech companies (Satariano 2020). In China, regulators released a new antitrust guideline in November 2020 specifically targeting internet giants such as Alibaba and Tencent (Liu and Ren 2020).

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 lead 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 the year, Britain's antitrust agency published a similar report on online platforms, expressing concern that 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 warn 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 shows how competition could potentially discourage innovation when it comes to development and adoption of more advanced learning algorithms.[8] 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.

7 Conclusion

The rise of AI and machine learning has dramatically changed marketing practices. Interaction between firms and consumers is increasingly frequent, personalized, automated, and more importantly guided by deliberate strategies. As marketing decisions become more evidence-based and algorithm-aided, companies are investing heavily in the technology and analytical expertise that enable real-time collection of customer data and performing dynamic interventions based on the data. However, while extensive research has focused on data modeling and analysis techniques, noticeably less attention has been paid to the theoretical and strategic implications, especially under competition.

In this paper, we develop an analytical model of adaptive content recommendations under competition. We formulate content recommendation algorithms as solutions to a stochastic dynamic programming problem under demand uncertainty. We compare recommendation algorithms under different learning regimes, and investigate how the value from more advanced algorithms vary across different competitive conditions. The model allows us to

[8]This occurs in our model with no switching cost and when consumers are myopic or sufficiently impatient. Whether this result remains in settings other than content consumption is an important topic for future research.

address three questions: how competition affects the optimal exploration vs. exploitation trade-off, how competition affects firms' incentives to invest in more advanced algorithms, and how such technological upgrades affect consumer welfare.

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 campaign, service assignment, and product recommendation 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 instantaneous 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 stylized model only features two user types and three content types. It would be interesting to see what happens under a more general distribution of preferences and choices. Third, we only consider competition between two firms, with independent user preferences for content from the two firms. Future research may consider a more competitive scenario or when preferences are correlated across platforms. Finally, learning in our model is symmetric between a user and the firm she visits (although asymmetric between firms). Adding private information to users may significantly complicates the problem, but is nonetheless an important direction for future research.

References

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

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

[3] 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.

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

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

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

[7] 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

[8] Aridor G, Mansour Y, Slivkins A, Wu ZS (2020). Competing bandits: The perils of exploration under competition. *Working paper*, arXiv preprint arXiv:2007.10144.

[9] Athey S, Bryan K, Gans J (2020). The Allocation of Decision Authority to Human and Artificial Intelligence. *Stanford University Graduate School of Business Research Paper No. 3517287*

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

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

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

[13] 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.

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

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

[16] 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.

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

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

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

[20] Dogan M, Jacquillat A, Yildirim P (2018). Strategic Automation and Decision-Making Authority. MIT Sloan School of Management.

[21] 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

[22] 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.

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

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

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

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

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

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

[29] Hansen K, Misra K, Pai M (2020). Algorithmic Collusion: Supra-Competitive Prices via Independent Algorithms. *CEPR Discussion Paper No. DP14372*. University of California San Diego.

[30] 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.

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

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

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

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

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

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

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

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

[39] 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).*

[40] 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.

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

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

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

[44] 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.

[45] Liu Y, Yildirim TP, Zhang ZJ (2019). Consumer Attitudes Toward Articial Intelligence and Price Discrimination. *Working paper.*

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

[47] Mansour Y, Slivkins A, Wu ZS (2018). Competing bandits: Learning under competition. *In 9th Innovations in Theoretical Computer Science Conference*, ITCS 2018, January 11-14, 2018, Cambridge, MA, USA, pages 48:1-48:27, 2018.

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

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

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

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

[52] 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.

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

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

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

[56] 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.

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

[58] 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).*

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

[60] 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.

[61] 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.

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

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

[64] 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).*

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

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

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

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

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

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

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

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

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

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

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

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

Appendix

Proof of Equation 8

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

$$\begin{aligned} V(\lambda_t) &= y(\lambda_t, S_t)dt + (1 - rdt)E[V(\lambda_{t+dt})] \\ &= y(\lambda_t, S_t)dt + V(\lambda_t) - rV(\lambda_t)dt + \frac{\sigma(\lambda_t)^2}{2}V''(\lambda_t)dt \end{aligned} \tag{20}$$

which simplifies to the following ordinary differential equation:

$$V(\lambda_t) = \frac{y(\lambda_t, S_t)}{r} + \frac{\sigma(\lambda_t)^2}{2r}V''(\lambda_t) \tag{21}$$

The value function has two terms which can be understood in the following ways. The first term, $y(\lambda_t, S_t)/r$, 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. Notice 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 (21) is

$$V(\lambda_t) = \frac{y(\lambda_t, S_t)}{r} + b_1\lambda_t^{(\gamma+1)/2}(1 - \lambda_t)^{-(\gamma-1)/2} + b_2\lambda_t^{-(\gamma-1)/2}(1 - \lambda_t)^{(\gamma+1)/2}, \tag{22}$$

with

$$\gamma = \sqrt{1 + \frac{8r\alpha(1 - \alpha)}{(2\alpha - 1)^2}}. \tag{23}$$

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

$$b_1 = 0. \tag{24}$$

Thus the solution simplifies to

$$V(\lambda_t) = \frac{y(\lambda_t, S_t)}{r} + b_2\lambda_t^{-(\gamma-1)/2}(1 - \lambda_t)^{(\gamma+1)/2}, \tag{25}$$

Proof of Proposition 3

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 $\lambda_t^1 > 0.5$ and $\lambda_t^2 > 0.5$. Suppose $\lambda_t^1 \leq \widehat{\lambda}_u$, and consider firm 1's deviation from mass-market content to niche-market content. By equation (16), firm 1 will receive no demand for any λ_t^2. Given that firm 1 also does not learn any information if the user visits the competitor, this deviation cannot be profitable. Suppose $\lambda_t^1 > \widehat{\lambda}_u$, and consider firm 1's deviation from niche-market content to mass-market content. If $\lambda_t^2 > \widehat{\lambda}_u$, then firm 1 has no demand after deviating, which cannot be profitable. If $\lambda_t^2 \leq \widehat{\lambda}_u$, then firm 1 gets a flow profit of $\frac{1}{2}c$ with no new information. Given that $\frac{1}{2}c < 0.5 < \lambda_t^1$, the deviation cannot be profitable. One can show that firm 2 does not have a profitable deviation in the same way.

Next, we prove this is the unique equilibrium 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 serves 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 $\tilde{\lambda} > \widehat{\lambda}_u$. 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 3, offering niche-market content if and only if $\lambda_t^j \notin [1 - \widehat{\lambda_u}, \widehat{\lambda_u}]$. After deviating, firm 1 gets flow payoff of $\lambda_t^1 > \frac{1}{2}c$ until λ_t^1 hits $\widehat{\lambda_u}$, and gets flow payoff of $\frac{1}{2}c$ after λ_t^1 hits $\widehat{\lambda_u}$. 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 $\tilde{\lambda} \leq \widehat{\lambda}_u$. 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 $\frac{c}{2r}$. 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 3 is the unique stationary strategy profile, where $S_t^j = S(\lambda_t^j)$, that constitutes equilibrium from possible priors.

Proof of Corollary 4.1

Note that the firm's value function for $\lambda_t > \lambda^*$ follows the same general solution as equation (8). However, at $\lambda_t = \lambda^*$, the user splits her time between the two providers. We have the following boundary condition

$$rV^{MY}(\lambda^*) = \frac{c}{2r}$$

Solving this gives $b_2 = -\frac{c}{2r}\lambda^{*\frac{\gamma-1}{2}}(1-\lambda^*)^{-\frac{\gamma+1}{2}}$. Thus for $\lambda_t > \lambda^*$, the value function is:

$$V^{MY}(\lambda_t) = \frac{\lambda_t\alpha + (1-\lambda_t)(1-\alpha)}{r} - \frac{c}{2r}(\frac{\lambda_t}{\lambda^*})^{-\frac{\gamma-1}{2}}(\frac{1-\lambda_t}{1-\lambda^*})^{\frac{\gamma+1}{2}}$$

If the firm uses the non-adaptive algorithm, then if it recommends niche-market content, the firm would lose users at λ^*. If the firm recommends mass-market content, it splits demand and earns $\frac{c}{2}$ as flow profit. The value function becomes:

$$V^{NA}(\lambda_t) = \max\{\frac{c}{2r}, \frac{\lambda_t\alpha + (1-\lambda_t)(1-\alpha)}{r} - \frac{c}{r}(\frac{\lambda_t}{\lambda^*})^{-\frac{\gamma-1}{2}}(\frac{1-\lambda_t}{1-\lambda^*})^{\frac{\gamma+1}{2}}\}$$

For $\lambda_0 > \lambda^*$, we have $V^{MY}(\lambda_t) > \frac{c}{2r}$, and

$$\frac{c}{2r}(\frac{\lambda_t}{\lambda^*})^{-\frac{\gamma-1}{2}}(\frac{1-\lambda_t}{1-\lambda^*})^{\frac{\gamma+1}{2}} > 0$$

This implies that the additional value from the myopic algorithm, $V^{MY} - V^{NA}$, is strictly positive for $\lambda_0 > \lambda^*$.

Evolution of Recommendations for Monopoly

Users who are 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 N1 at time t.

Let

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

Then we have:

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

here, $h(\lambda, t)$ and $p(z, t)$ are probability density function of λ and z at time t.

For users who prefer content type N1, 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 are $1 - \lambda_0$ proportion of users who prefer content type $N2$. 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 we have for users who prefer N2:

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

The first hitting time probability density for users who prefer N1 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 N2:

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

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

$$\begin{aligned} h(z_0,t) &= \lambda_0 h_1(z_0,t) + (1-\lambda_0)h_2(z_0,t) \\ &= (\lambda_0 + (1-\lambda_0)\exp(z_0-\hat{z}))(z_0-\hat{z})\frac{1}{\sqrt{\sigma_z^2 t^3}} n\left(\frac{z_0-\hat{z}+\sigma_z^2 t/2}{\sigma_z\sqrt{t}}\right) \end{aligned} \tag{26}$$

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

$$\begin{aligned} H(t) &= \lambda_0 H_1(z_0,t) + (1-\lambda_0)H_2(z_0,t) \\ &= \frac{\lambda_0}{\hat{\lambda}}\left[\Phi\left(\frac{(\hat{z}-z_0)-\sigma_z^2 t/2}{\sigma_z\sqrt{t}}\right) + \exp(\hat{z}-z_0)\Phi\left(\frac{(\hat{z}-z_0)+\sigma_z^2 t/2}{\sigma_z\sqrt{t}}\right)\right] \end{aligned} \tag{27}$$

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 N1 conditional on that the firm recommends content type N1 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

$$\bar{\lambda}_t = \frac{\lambda_0 - \hat{H}(t)\hat{\lambda}}{1-\hat{H}(t)} \tag{28}$$

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 N1 may receive N1 content. Users who prefer type N2, but were incorrectly recommended type N1 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 N1 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

$$H^*(t) = \frac{\lambda_0}{\lambda^*}\left[N\left(\frac{(z^*-z_0)-\sigma_z^2 t/2}{\sigma_z\sqrt{t}}\right) + \exp(z^*-z_0)\Phi\left(\frac{(z^*-z_0)+\sigma_z^2 t/2}{\sigma_z\sqrt{t}}\right)\right] \tag{29}$$

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 above results are summarized in Proposition 2.

Evolution of Profit for 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

$$\begin{aligned}
\pi_t &\equiv E\int y(\lambda_t, S(\lambda_t)) \\
&= (1-\hat{H}(t))y(\bar{\lambda}_t, S(\bar{\lambda}_t)) + \hat{H}(t)c \\
&= y(\lambda_0, S(\lambda_0)) - \hat{H}(t)y(\hat{\lambda}, S(\hat{\lambda})) + \hat{H}(t)c \\
&= \lambda_0\alpha + (1-\lambda_0)(1-\alpha) - \hat{H}(t)[\hat{\lambda}\alpha + (1-\hat{\lambda})(1-\alpha) - c]
\end{aligned} \tag{30}$$

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 firms expects to suffer losses from deploying the forward-looking algorithm until the proportion of users that receive mass-market content reach $\frac{y(\lambda_0,S(\lambda_0))-c}{y(\hat{\lambda},S(\hat{\lambda}))-c}$.

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 5a and 5b, and compare to flow and cumulative profit under myopic or non-adaptive algorithms. 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. There must exist such t_1, 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 non-adaptive or the myopic algorithms for $t < t_0 = \hat{H}^{-1}\left(\frac{y(\lambda_0, S(\lambda_0))-c}{y(\hat{\lambda}, S(\hat{\lambda}))-c}\right)$.*

Online Appendix

Endogenous Monetization Level for Monopoly

In the model 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 viewed, 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 viewed 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:

$$dY(Q) = y(T, S_Q)dQ + \sqrt{\alpha(1-\alpha)p_Q^2}dW(Q) \tag{31}$$

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

$$dY(t) = y(T, S_t)q_t dt + \sqrt{q_t}\sqrt{\alpha(1-\alpha)p_t^2}d\tilde{W}_t \tag{32}$$

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:

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

$$= \frac{\lambda_t(1-\lambda_t)(2\alpha-1)(\frac{\beta_1}{\beta_2} - \frac{1}{\beta_2}p_t)}{\alpha(1-\alpha)p_t}[y(T) - y(\lambda_t)]dt + \frac{\lambda_t(1-\lambda_t)(2\alpha-1)\sqrt{(\frac{\beta_1}{\beta_2} - \frac{1}{\beta_2}p_t)}}{\sqrt{\alpha(1-\alpha)}}dW_t$$

Notice 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:

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

The HJB equation gives us:

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

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 choose 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 1.3, 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 recommendation 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 (21):

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

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, such as 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 A.2 *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