INFORMATION DESIGN OF ONLINE PLATFORMS*

T. TONY KE
Chinese University of Hong Kong
tonyke@cuhk.edu.hk

SONG LIN
Hong Kong University of Science and Technology
mksonglin@ust.hk

MICHELLE Y. LU
China Europe International Business School
michellelu@ceibs.edu

November 2022

*We thank Jiwoong Shin, Duncan Simester, Miguel Villas-Boas, Mark Whitmeyer, Robert Zeithammer, Jidong Zhou for helpful comments and suggestions. We also thank seminar participants at Monash University, University of Science and Technology of China, 2020 UTD Bass FORMS Conference, 2021 POMS Conference, Mannheim Virtual IO Conference, Consumer Search Digital Seminar. An earlier version of the paper was circulated under the title of "Personalization Trap".

INFORMATION DESIGN OF ONLINE PLATFORMS

ABSTRACT

We consider the strategic use of information by an online platform to both guide consumers' search through product recommendations and influence sellers' targeted advertising decision. Our model unifies the analysis of personalized product recommendation and targeted advertising under the information design framework. We illustrate a fundamental tradeoff facing the platform between increasing trades through higher match efficiency and extracting seller surplus by inducing their competition for prominence. The optimal information design may be socially inefficient, as it balances the tradeoff by limiting consumer search and mixing the matched product with a long tail of unmatched ones for recommendation.

Keywords: *information design, platform design, product recommendation, personalization, targeted advertising, consumer search*

1 Introduction

Growing availability of consumer data and rapid advances in data analytics have fueled the growth of online marketplaces in recent years. Harnessing the power of granular data on customer characteristics and purchasing history, online platforms can predict the match between buyers and sellers and thus make personalized product recommendations to buyers. Enhanced match efficiency can increase platforms' sales revenue.[1] Meanwhile, these platforms are gaining ground on the digital advertising market. Alibaba and Amazon have grown into the third- and forth-largest online ad platforms in the world respectively, only trailing behind Google and Facebook (eMarketer, 2019). Customer data analytics have improved the accuracy of targeted advertising, increasing sellers' bidding incentives and thereby raising the ad revenue for online platforms.[2]

However, these two concurrent trends are seemingly at odds with each other. With more precise data, platforms can recommend sellers to consumers at a higher accuracy and thus generate more sales revenue. Yet, this may create thinner advertising markets where the recommended sellers can establish local monopoly power. Consequently, sellers' incentives to bid for advertising are reduced, eventually hurting platforms' advertising revenue.

This research seeks to explain the phenomenon from the perspective of information design and shed light on how the above trade-off is influenced by consumer search. Drawing on the Bayesian persuasion framework (Kamenica and Gentzkow 2011), we posit that a platform can design a public signal that influences the beliefs of both consumers and sellers. In practice, such a signal takes the forms of personalized product recommendation on the consumer side and targeted advertising on the seller side. Upon observing the signal, a consumer can conduct a sequential search with perfect recall among the sellers. After visiting a seller, the consumer observes the product price and whether the product is a match or not (Wolinsky 1986). Besides setting the price, sellers also decide how much to bid in a second-price ad auction for each consumer, where the winner is granted a prominent position. The consumer can obtain the price and match information of the seller in the prominent position at no cost. Beyond the prominent position, the consumer incurs search cost to visit additional sellers. In essence, we aim to build a unified model that examines three interrelated information problems stemming from the matching uncertainty between consumers and sellers: (1) consumers can search

[1]Schrage (2021) reports that online shoppers are 4.5 times more likely to complete a purchase after clicking on any product recommendation, and personalized product recommendation accounted for almost 31% of the revenues in the global e-commerce industry. For streaming services like YouTube and Netflix, video recommendation accounts for 60-75% of revenue generated. Agrawal et al. (2018) envisioned that as the recommendation accuracy further improves, retail platforms' current business model of "shop-then-ship" may be replaced by a new model of "ship-then-shop", where companies can anticipate their customers' needs and ship relevant products to them before they place an order, and then, customers only need to return the ones they do not want.

[2]For example, Amazon's interest-based ads allow advertisers to target Amazon's customers based on their interactions with Amazon sites, content, and services, as well as cookies that track their viewing and clicking activities on third-party sites.

for the best matched product; (2) sellers can direct consumer search by advertising; and (3) a platform can design an information environment to influence both the consumers' search and sellers' advertising.

The crux of the information design problem is the platform's dilemma in increasing trades through higher match efficiency and extracting seller surplus by inducing their competition for prominence. By revealing more information, the platform improves match efficiency, but it also gives the matched seller a competitive edge over the unmatched ones. Thus, the unmatched sellers are discouraged from competitive bidding, impairing the platform's ability to extract sellers' surplus from their competitive ad-bidding.

We illustrate the basic trade-off with two extreme designs—full information and no information. In the full information environment, the platform perfectly reveals the match information—the matched seller will win the prominent position but pay zero price. Given that, the consumer does not need to search beyond the prominent seller. As a result, the platform profits only from sales commissions with no advertising revenue. In other words, even though the match is enabled by an auction, it is as if the auction were shut down and instead, the platform uses personalized product recommendation solely to monetize consumer attention. Conversely, in a no-information environment, the platform randomly matches the consumer with one seller. The inefficiency in matching may motivate the consumer to search for the matched seller, which in turn provides an incentive for sellers to bid aggressively to become prominent so as to influence consumer search. Their competitive motive to corner competitors out of the prominent position ultimately leads to the prisoner's dilemma: all sellers' bidding will not change the random matching outcome but the platform can fully exploit seller surplus. In other words, we end up with the other extreme where no product recommendation is used by the platform; instead, sponsored ads are the sole means to monetizing consumer attention.

The optimal information design lies in-between these two polar cases. Equipped with data capability, the platform can design a posterior belief that induces any search order of a consumer. The induced search order further affects sellers' bidding incentive for the prominent position. To solve for the optimal information design, we show that the platform's profit-maximization problem over posterior beliefs is equivalent to the problem of finding the optimal consumer search length induced by an optimal belief.

If a consumer's search cost is sufficiently low, then it is optimal for the consumer to keep searching until she finds the matched product. Anticipating the perfect match, sellers have no incentive to advertise. Any information design of the platform leads to the same expected profit, which only consists of commissions from sales. On the other hand, if consumer's search cost is not that low, then the optimal information design hinges on the platform's ability to collect commissions from sales. In practice, commissions are often shaped by the bargaining power between the platform and sellers, platform competition, and/or regulatory policies. When a platform is able to demand a high commission fee, it will find it optimal to adopt the full information design, which maximizes match efficiency, despite of zero advertising revenue. In contrast, when the commission rate is relatively lower, the platform finds it optimal to adopt a coarse information design. That is, it recommends two sellers ("contenders" hereafter) that have an equal and relatively high probability to be the match, while recommending the remaining sellers ("a long tail" hereafter) with an equal and relatively low probability to be the match.

Under this optimal information design, one of the two contenders will win the prominent position, and consumers do not search beyond the prominent seller, even if it turns out to be a mismatch. By recommending two contenders with a relatively high matching probability, the platform provides "winning rewards" to the them, who upon winning the prominent position can expect a relatively high probability of being matched. In the meantime, by keeping a long tail of sellers with a relatively low but positive matching probability, the platform ensures that, upon finding out the prominent seller is not a match, a consumer is still unsure which remaining seller could be a match and thus stops searching. The limited search length essentially serves as a "losing punishment" to the two contenders, who upon losing the prominent position will never be visited. The winning rewards and the losing punishment together incentivize the two contenders to bid for advertising while ensuring the chance of matching to be as high as possible.

Our findings yield several implications for online marketplaces. First, when a platform's ability to appropriate from sales revenue is relatively low, the optimal "noisy design" introduces social inefficiency to the market. This result implies that a policy that only regulates a platform's market power without limiting its data power could hurt social welfare. Second, the optimal noisy design implies that platforms may not need to invest in collecting all pieces of information. They do not necessarily get hurt from regulatory protection of consumer privacy. Last, platform designs may have non-monotonic effects on the optimal information design and social welfare. Notably, the match efficiency and social welfare under the optimal information design is non-monotonic in the number of sellers in one product category. Moreover, a platform would have an incentive to obfuscate consumer search because it is easier for the platform to limit consumer search and thus provide sellers with higher incentives to bid for advertising and compete for prominence.

This paper enriches a growing literature of information design (a là Kamenica and Gentzkow 2011) and applies the method to a two-sided market with a public signal. The information design literature has examined the impact of information structure on market outcomes such as market segmentation (Bergemann et al. 2015), consumer search and price competition (Board and Lu 2018, Bergemann et al. 2021, Dogan and Hu 2022, Whitmeyer 2021), and product differentiation (Armstrong and Zhou 2022). Compared with the existing studies, our application is novel as the design of the signal factors in the strategic interactions between consumers' search and purchase decisions on one side of the market and sellers' targeting and advertising decisions on the other side.

When designing the information, the platform faces a trade-off between match efficiency and surplus extraction. A similar tradeoff between targetability and market thickness in online advertising markets has been discussed by Levin and Milgrom (2010) and studied by Bergemann and Bonatti (2011). In their framework, targetability improves match efficiency between sellers and buyers but inevitably creates many thin markets that limit the sellers' competitive incentives. Different from them, we do not impose exogenous market segmentation but allow the platform to use information design to directly moderate the sellers' incentives to become prominent in consumer search.

Our paper builds on the seminal search model of Wolinsky (1986) and Anderson and Renault (1999) and contributes to the literature, especially those with applications in retail industry (e.g., Zhou 2014, Rhodes 2015, Rhodes and Zhou 2019, Jiang and Zou 2020, Rhodes et al. 2021). Since we focus on a third-party retail platform instead of multiple competing retail outlets, it seems natural to abstract away from the consumers' multi-product search problem that has been much of the focus of the literature. On the other hand, we allow the platform to use sponsored ads to influence consumer search order; in this sense, our paper also contributes to the literature of position auctions (e.g., Athey and Ellison 2011, Chen and He 2011).

The rest of the paper is organized as follows. We present the model in Section 2. Section 3 demonstrates the main tradeoff by analzying the two polar information designs (full information in 3.1 and no information in 3.2). Then we solve for the optimal information design problem in 4. Section 5 concludes the paper and discusses some limitations and future research directions.

2 The Model

Consider an online platform that matches $N \geq 3$ sellers with $I \geq 1$ consumers. Sellers are indexed by $n \in \{1, \cdots, N\} \equiv \mathcal{N}$, each offering one product. Their marginal production costs are assumed to be the same and normalized as zero. For each consumer, there is one and only one product that matches with her preference. This assumption greatly simplifies the state space for information design; we discuss what happens under alternative setups in Section 5. Given the selling price of p, each consumer's demand function is $D(p)$ for the matched product and zero for the unmatched ones. It is assumed that $D(\cdot)$ is strictly decreasing, differentiable and log-concave. This guarantees that the monopolistic price, p^*, given by $D(p^*) + p^* D'(p^*) = 0$ exists and is positive. Moreover, it is further assumed that $D(p^*) > 0$ so that consumers earn positive surplus under the monopolistic price. All players share the common prior belief that for each consumer, all sellers are equally likely to be a match. That is, for each consumer, the match probability is $1/N$ a priori for every product. Consumers also have an outside option normalized as zero. We impose the tie-breaking rule that consumers prefer the outside option than an unmatched product.

The true state of the world, i.e., the identity of the matched seller for each consumer, is represented by $\omega = (n_1, n_2, \cdots, n_I) \in \mathcal{N}^I$, where $n_i \in \mathcal{N}$ is the index of the matched seller for consumer $i \in \{1, \cdots, I\} \equiv \mathcal{I}$. The platform is the *Sender* who designs an information structure that consists of a signal realization space $\mathcal{S}$ and a family of probability distribution $\{\pi(s|\omega)\}_{\omega \in \mathcal{N}^I}$ for $s \in \mathcal{S}$, where without loss of generality we can restrict $\mathcal{S} = \mathcal{N}^I$. Both the consumers and sellers are *Receivers*, who observe the platform's choice of the information structure and a signal realization $s = (s_1, \cdots, s_I) \in \mathcal{N}^I$. In practice, the platform's information design can be operationalized via a two-sided recommendation strategy. Particularly, a signal realization of $s = (s_1, \cdots, s_I)$ means that the platform will recommend seller s_i to consumer i for purchase and recommend consumer i to seller s_i for targeted advertising for $i \in \mathcal{I}$. All consumers and sellers understand that they get this recommendation s with probability $\pi(s|\omega)$ that depends on ω, the true state of the world. For example, a fully revealing design with $\pi(\omega|\omega) = 1$ for $\forall \omega \in \mathcal{N}^I$ induces recommendation of perfect matches. An uninformative design with $\pi(s|\omega) = 1/N^I$ for $\forall s, \omega \in \mathcal{N}^I$ induces a purely random recommendation of products.

Once the platform determines the information structure, it holds a second-price auction for each consumer $i \in \mathcal{I}$ to sell a prominent position that displays an advertisement from sellers to that consumer. In Section 5, we discuss other selling mechanisms such as take-it-or-leave-it offers. We impose the tie-breaking rule that given multiple sellers' bids being the same, the platform will choose the winner that maximizes its revenue under the information environment it commits to. Each seller $n \in \mathcal{N}$ sets his selling price $p_n \geq 0$ as well as bid $b_{ni} \geq 0$ in the auction for consumer i. The winner of the auction will be advertised in a prominent position for the consumer, who learns the match value and price of the seller in the prominent position freely. Based on the information structure and the realized signal, the consumer updates her belief of match values of the remaining sellers and decides whether to continue to search for information on them sequentially. Upon continuation, she decides which seller to visit next. By paying a search cost $c > 0$, the consumer visits a seller and learns her match value and price. At any time point, she chooses between continuing to search and stopping searching to make a purchase decision from one of the sellers that she has visited or to take the outside option. Our utility specification implies that a consumer will make a purchase only if the product is a match. When a transaction takes place, the platform pockets a fraction $\alpha \in [0, 1]$ of the revenue as a commission while the seller keeps the remaining fraction $1 - \alpha$. The commission α may be determined by various factors out of the scope of the model, such as competition between platforms, and thus is treated as exogenous in our main model; we discuss endogenous α as well as two-part tariffs of commission in Section 5.

We conclude the model setup by summarizing the game timeline. First, the platform commits to an information structure $\{\pi(\cdot|\omega)\}_{\omega \in \mathcal{N}^I}$. Second, a signal s realizes according to the information structure. Both sellers and consumers observe the signal realization. Third, sellers decide their price p_n and bid b_{ni} (for $n \in \mathcal{N}$ and $I \in \mathcal{I}$). Lastly, each consumer observes the prominent seller determined by the ad auction and then makes her search and purchase decisions.

3 Preliminary Analysis

We first analyze sellers' pricing decisions and consumers' optimal search strategy under any given information structure and realized signal. In our model, consumers do not observe sellers' prices beforehand. They form rational expectation about the unobserved prices, which coincide with the actual ones in equilibrium. We follow the literature (e.g., Wolinsky 1986, Anderson and Renault 1999) and assume passive belief updating in that if a consumer observes a seller's deviation from the equilibrium price, she does not update her belief about the prices of other sellers that she has not visited yet. The following lemma characterizes the equilibrium prices and consumer search strategy in a similar flavor of Diamond (1971). Proofs to all statements in this paper are provided in Appendix.

Lemma 1 (Equilibrium Price)**.** *In any perfect Bayesian equilibrium, every seller with a positive demand sets price at the monopolistic level of p^*.*

We have set up the model in such a way that sellers' strategic decisions in pricing is shut down. This greatly simplifies the equilibrium analysis and allows us to focus on the impact of the platform's information design on sellers' targeted advertising decisions on one side of the market and consumers' search decisions on the other. The key driving force behind Lemma 1 is the binary nature of match values—it does not depend on the assumption of one and only one matched seller for each consumer.

Lemma 1 implies that the platform's information design does not affect sellers' pricing decision. Moreover, notice that sellers' advertising decisions are targeted at individual consumer level. These two observations together imply that sellers' pricing and advertising decisions for different consumers are independent so that we can without loss of generality restrict our attention to just one representative consumer i. With slight abuse of notation, the platform's information design problem can be decomposed at consumer level with $\pi(s|\omega) = \Pi_{i\in\mathcal{I}}\pi(s_i|n_i)$. Under this decomposition, consumer i only needs to observe $\{\pi(s_i|n_i)\}_{n_i\in\mathcal{N}}$, while sellers observe $\{\pi(s|\omega)\}_{\omega\in\mathcal{N}^I}$ and s. Therefore, a consumer only needs to observe the product recommendation targeted at her. For notational simplicity, we drop subscript i for the remaining analysis when we study a representative consumer.

Let us define a seller's monopolistic profit before commission as V and a consumer's surplus from a matched seller as U, where

$$V = p^* D(p^*),$$
$$U = \int_{p^*}^{\infty} D(p)dp.$$

To avoid the trivial case where consumers never search, we make the following assumption

Assumption 1. *$0 < c < U$.*

Denote a representative consumer's posterior belief given signal $s \in \mathcal{N}$ by $\mu_s(n)$ for $n \in \mathcal{N}$, where by Bayes' rule,

$$\mu_s(n) = \frac{\pi(s|n)}{\sum_{n \in \mathcal{N}} \pi(s|n)}. \tag{1}$$

The following lemma characterizes the consumer's optimal search strategy given her posterior belief.

Lemma 2 (Optimal Search Strategy)**.** *If a consumer searches beyond the prominent seller, then it is optimal for her to search by the descending order of her posterior belief.*

The lemma illustrates a key point that the platform can use information design to influence consumers' posterior belief and consequently, guide their search process. Notice that we cannot leverage Weitzman (1979) directly to prove the lemma, because for a consumer, there is only one matched seller in the market so that sellers are not independent (Ke and Lin, 2020, Janssen and Ke, 2020).

Next, to illustrate the main trade-off, we first analyze two polar cases of the information environment: (1) the platform chooses a fully informative information structure; (2) it chooses a completely uninformative one. The former case can be understood as an environment where the platform exploits rich consumer data to precisely tell which seller a consumer likes. Contrarily, the latter depicts an environment in which the platform commits not to utilize (or do not have access to) any consumer data.

3.1 Full Information

Let us consider the bidding strategy of seller n, who is the match for the representative consumer. Note that if he wins the prominent position, he earns a profit of $R_{win} = (1 - \alpha)V$; on the other hand, if he loses, the consumer will still visit him as $c < U$, so his profit upon losing, $R_{lose} = R_{win}$. This implies that the seller will bid $b_n^* = R_{win} - R_{lose} = 0$. Obviously, other sellers will bid zero. Given our tie-breaking rule, the platform will choose the matched seller as the winner, as it maximizes the platform's revenue. As a result, seller n wins the prominent position at no cost and the platform gets zero advertising revenue. Ad auctions are essentially inactive, and the platform profits solely from sales commission. The following proposition summarizes these equilibrium outcomes.

Proposition 1 (Full Information)**.** *Under the fully informative environment, in equilibrium:*

1. *all sellers bid zero, and the matched seller wins and earns a profit of* $(1 - \alpha)V$*;*
2. *consumers do not search and buy from the prominent seller, each expecting a surplus of* U*;*
3. *the platform obtains a profit of* αIV*.*

3.2 No Information

With no information from the platform, consumers' posterior belief is the same as the prior, where all sellers are equally likely to be a match, so if they decide to search, they will search among the unadvertised sellers randomly.

Lemma 3 ("None or All"). *Under the uninformative environment, if $0 < c \leq 2U/N$, the consumer keeps searching until she finds a match; if $2U/N < c < U$, the consumer never searches beyond the advertised seller.*

If the consumer has the incentive to start searching, then not seeing a match after a product search will strengthen her belief about the possibility of a match in the remaining pool of unsearched products, making her more willing to search than before. Thus, she will continue to search until a match is found.

Next, we examine a seller's bidding strategy. If he wins the auction, he has an $1/N$ chance to be a match, yielding an expected payoff of $R_{win} = (1-\alpha)V/N$; if he loses, there are two cases to consider according to Lemma 3. First, if $0 < c \leq 2U/N$, the consumer keeps searching until she finds a match, so we have $R_{lose} = R_{win}$. The seller will bid $b_n^* = R_{win} - R_{lose} = 0$, and the platform gets commission of αIV and zero advertising revenue. Second, if $2U/N < c < U$, the consumer never searches beyond the advertised seller, so we have $R_{lose} = 0$. The seller will bid $b_n^* = R_{win} - R_{lose} = (1-\alpha)V/N$. Given the platform commits to disclose no information, it can only choose the winner randomly. By doing so, it gets commission of $\alpha IV/N$ and advertising revenue of $(1-\alpha)IV/N$. The following proposition summarizes these equilibrium outcome.

Proposition 2 (No Information). *Under the uninformative environment, in equilibrium:*

1. *If $0 < c \leq 2U/N$,*
   - (a) *all sellers bid zero, and a random seller wins the prominent position, and the matched seller earns a profit of $(1-\alpha)V$;*
   - (b) *consumers purchase the advertised product if it is a match; otherwise, they keep searching until finding a match; each consumer expects a surplus of $U - Nc/2$;*
   - (c) *the platform obtains a profit of αIV.*
2. *Otherwise, if $2U/N < c < U$,*
   - (a) *all sellers bid $(1-\alpha)V/N$, and a random seller wins the prominent position and earns zero profit;*
   - (b) *consumers purchase the advertised product if it is a match; otherwise, they exit the market; each consumer expects a surplus of U/N;*
   - (c) *the platform obtains a profit of IV/N.*

3.3 Comparison

By comparing the equilibrium outcomes between the two polar cases above, we can establish the following proposition.

Proposition 3 (Full vs. No Information)**.** *When $0 < c \leq 2U/N$, full and no information environments generate the same revenue for the platform, but consumer surplus is lower under the no information environment. When $2U/N < c < U$, the platform earns a higher revenue under the full information environment if and only if $\alpha N > 1$.*

The first case ($0 < c \leq 2U/N$) is more likely to happen when the number of sellers N is small, or consumer search cost c is low. In this case, even without any information provided by the platform, matching uncertainty can be resolved through consumer search. As a result, the platform's information design becomes inconsequential.

On the contrary, the second case ($2U/N < c < U$) illustrates the platform's tradeoff between match efficiency and surplus extraction in designing information environment. By revealing all information, the platform perfectly matches sellers and consumers. However, as sellers have been matched with right consumers, they have no incentive to pay for advertising. In contrast, by suppressing all information, the platform randomly matches the two sides and suffers from mismatch inefficiency. Nonetheless, such a mismatch outcome provides an opportunity for sellers to influence consumer purchase by advertising at the prominent position. Consequently, the platform sacrifices match volume but fully extracts sellers' surplus by inducing them to compete in advertising.

The discussion on the two extreme cases naturally invites a question of the optimal information design, which we address next.

4 Optimal Information Design

Denote the representative consumer's posterior belief given signal $s \in \mathcal{N}$ by $\boldsymbol{\mu}_s = (\mu_s(1), \cdots, \mu_s(N))$. Signal s can be intuitively understood as the recommendation of seller s. Notice that all sellers are ex-ante symmetric, the realization of different signals, s will only induce a permutation of the posterior belief among sellers. Therefore, we can without loss of generality omit the dependence of $\boldsymbol{\mu}_s$ on s and assume $\mu(1) \geq \mu(2) \geq \cdots \geq \mu(N)$. That is, when the platform is designing the information structure and correspondingly, the posterior belief, the identity of the sellers does not matter; only the distribution of $\mu(n)$ across seller n matters. We first solve for the posterior belief, $\boldsymbol{\mu}^*$ that maximizes the platform's revenue. Then, we show that it is Bayes plausible by constructing the signal distribution $\{\pi(s|\omega)\}_{\omega \in \mathcal{N}^I}$.

Define the platform's revenue as $\Pi(\boldsymbol{\mu})$ given the consumer's posterior belief $\boldsymbol{\mu}$. We can write down the platform's revenue maximization problem as the following.

$$\Pi^* = \max_{\boldsymbol{\mu}} \Pi(\boldsymbol{\mu}) \text{ s.t. } \mu(1) \geq \cdots \geq \mu(N) \geq 0 \text{ and } \sum_{n=1}^{N} \mu(n) = 1. \tag{2}$$

Notice that because of the permutation symmetry among sellers, the Bayes plausibility is naturally satisfied, so it does not appear in the constrained optimization problem above. This will become clearer later when we construct the the signal distribution that realizes the optimal posterior belief.

4.1 Search Length and Platform Profit

To derive $\Pi(\boldsymbol{\mu})$, we need to determine the consumers' search behavior and sellers' bidding behaviors induced by the posterior belief $\boldsymbol{\mu}$. Given a consumer's posterior belief upon receiving the signal and the seller at the prominent position, Lemma 2 implies that she will search among sellers in the descending order of her posterior belief. Therefore, to characterize a consumer's search behavior, we only need to know how long she will continue to search before seeing a match. This observation motivates the notion of *search length*.

Definition 1 (Search Length)**.** *Given the posterior belief $\boldsymbol{\mu}$ and seller n taking the prominent position, a consumer's search length, $L(\boldsymbol{\mu}, n)$, is defined as the number of sellers, including the prominent seller, that she will visit given no match has been found.*

Given this definition, we can completely pin down the consumer's search behavior as the following: she will first visit the prominent position and then keep searching among the sellers in the descending order of her posterior belief until she either finds a match or has visited $L(\boldsymbol{\mu}, n)$ sellers, whichever comes first.

Let's define $\Pi(\boldsymbol{\mu}, L)$ as the platform's profit given the consumer's posterior belief $\boldsymbol{\mu}$ and any exogenously fixed search length L. One can intuitively understand $\Pi(\boldsymbol{\mu}, L)$ by imaging that the platform could assign any search length, L exogenously to the consumer, who still follows the search order determined by Lemma 2 given her posterior belief $\boldsymbol{\mu}$. $\Pi(\boldsymbol{\mu}, L)$ is well defined because the consumer's search strategy is completely pinned down given $\boldsymbol{\mu}$ and the exogenously given search length, L.

We derive $\Pi(\boldsymbol{\mu}, L)$ next, for which, we need to consider two cases. First, if $L = N$, the consumer will always keep searching until she finds the match, implying that all sellers will bid zero and the platform expects zero advertising revenue, similar to case 1 in Proposition 2. Therefore, we have $\Pi(\boldsymbol{\mu}, N) = \alpha IV$. Second, consider $L = 1, \cdots, N-1$. For seller $n = 1, \cdots, L-1$ (if $L = 1$, this is an empty set), his profit upon winning the auction is $R_{win} = (1-\alpha)\mu(n)V$. If he loses the auction, the consumer will still visit him if he is the match, so $R_{lose} = (1-\alpha)\mu(n)V = R_{win}$. This implies that seller n will bid $b_n^* = R_{win} - R_{lose} = 0$ for $n = 1, \cdots, L-1$. This further implies that the winner of the auction will be among sellers $L, \cdots, N$. For seller $n = L, \cdots, N$, his profit upon winning the auction is $R_{win} = (1-\alpha)\mu(n)V$. If he loses the auction, the consumer will not visit him, so $R_{lose} = 0$. This implies that seller n will bid $b_n^* = R_{win} - R_{lose} = (1-\alpha)\mu(n)V$ for $n = L, \cdots, N$. Therefore, seller L will win the auction and pay seller $L+1$'s bid, and consequently, the consumer follows the sequence of $L, 1, \cdots, L-1$ to search. The platform's revenue is

$$\Pi(\boldsymbol{\mu}, L) = \underbrace{\alpha IV \sum_{n=1}^{L} \mu(n)}_{\text{commission fee}} + \underbrace{(1-\alpha)IV\mu(L+1)}_{\text{advertising revenue}}, \text{ for } L = 1, \cdots, N-1. \tag{3}$$

Next, we turn to endogenous search length, and establish the relationship between $\Pi(\boldsymbol{\mu})$ and $\Pi(\boldsymbol{\mu}, L)$. Denote $n^*(\boldsymbol{\mu})$ as the seller who wins the prominent position in equilibrium induced by any belief $\boldsymbol{\mu}$. Then the endogenous equilibrium search length is $L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu}))$. We have:

Lemma 4. *In any sub-game equilibrium induced by $\boldsymbol{\mu}$, the profit of the platform satisfies*

$$\Pi(\boldsymbol{\mu}) = \Pi(\boldsymbol{\mu}, L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu}))). \tag{4}$$

This lemma is not as trivial as it may appear at first glance. In fact, in order to calculate $\Pi(\boldsymbol{\mu})$, we need to know the platform's advertising revenue, for which, we need to know each seller's bid, including those who lose the auction in equilibrium. This would further entail an analysis of the off-equilibrium path, on which, a seller who loses in equilibrium but deviates to win the auction. On this off-equilibrium path, there is no guarantee that the consumer's search length is still $L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu}))$, because the prominent seller is no longer $n^*(\boldsymbol{\mu})$. This is the disconnection between $\Pi(\boldsymbol{\mu})$ and $\Pi(\boldsymbol{\mu}, L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu})))$, as the latter always entails a fixed search length of $L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu}))$. Nevertheless, we prove in Appendix that equation (4) holds.

Based on Lemma 4, we can reformulate the problem in (2) as follows:

$$\Pi^* = \max_{L=1,\cdots,N} \Pi^*(L), \text{ where,} \tag{5}$$

$$\Pi^*(L) = \max_{\boldsymbol{\mu}} \Pi(\boldsymbol{\mu}, L) \text{ s.t. } L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu})) = L,\ \mu(1) \geq \cdots \geq \mu(N) \geq 0 \text{ and } \sum_{n=1}^{N} \mu(n) = 1.$$

In this reformulation, we first solve for the optimal belief for a given search length. This step requires that the belief is chosen such that the consumer follows the given search length. We then solve for the optimal search length.

4.2 A Relaxed Problem with Exogenous Search Length

In the following, we will first work on a relaxed problem of (5) by relaxing the constraint of $L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu})) = L$:

$$\Pi^*_{relexed}(L) = \max_{\boldsymbol{\mu}} \Pi(\boldsymbol{\mu}, L) \text{ s.t. }, \mu(1) \geq \cdots \geq \mu(N) \geq 0 \text{ and } \sum_{n=1}^{N} \mu(n) = 1. \tag{6}$$

That is, we will first investigate the platform's design of $\boldsymbol{\mu}$ given a consumer's arbitrary and fixed search length, $L \in \mathcal{N}$. This gives us some intuition about the platform's tradeoff in designing the optimal posterior belief. To solve problem (6), notice that by equation (3), $\Pi(\boldsymbol{\mu}, L)$ increases with $\mu(1), \cdots, \mu(L+1)$ and does not depend on $\mu(L+2), \cdots, \mu(N)$. Thus, we shall minimize $\mu(L+2), \cdots, \mu(N)$ by setting $\mu(L+2) = \cdots = \mu(N) = 0$. Therefore, $\mu(L+1) = 1 - \sum_{n=1}^{L} \mu(n)$, and correspondingly, $\Pi(\boldsymbol{\mu}, L) = (1-\alpha)IV + (2\alpha - 1)IV \sum_{n=1}^{L} \mu(n)$. To maximize $\Pi(\boldsymbol{\mu}, L)$ subject to $\mu(1) \geq \cdots \geq \mu(L+1) \geq 0$, we have the optimal solution as the following,

$$\begin{cases} \sum_{n=1}^{L} \mu(n) = 1 \text{ and } \mu(1) \geq \cdots \geq \mu(L) \geq \mu(L+1) = \cdots = \mu(N) = 0, & \text{if } \alpha \geq \frac{1}{2}, \\ \mu(1) = \cdots = \mu(L+1) = \dfrac{1}{L+1} \text{ and } \mu(L+2) = \cdots = \mu(N) = 0, & \text{otherwise.} \end{cases} \tag{7}$$

Intuitively, if the commission rate α is above $1/2$, the platform allocates all the match probability to the first L products to maximize commission; otherwise, the platform will essentially maximize $\mu(L+1)$ so as to maximize the advertising revenue subject to the constraint $\mu(1) \geq \cdots \geq \mu(L+1)$. Hence, the optimal solution is $\mu(1) = \cdots = \mu(L+1) = 1/(L+1)$.

4.3 Solving the Original Problem with Endogenous Search Length

Next, we work on the original problem (5) with the constraint $L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu})) = L$. Equation (7) implies that there are two cases to consider.

First, consider $\alpha \geq 1/2$. Notice that given the solution in equation (7), $\Pi(\boldsymbol{\mu}, L) = \alpha IV$, which does not depend on L. $\Pi^* = \max_{L=1,\cdots,N-1} \Pi^*(L) \leq \max_{L=1,\cdots,N-1} \Pi^*_{relaxed}(L) = \alpha IV$. Meanwhile, by Proposition 1, we know under full information, the consumer' search length is one and the platform's profit is αIV. This implies that $\Pi^* \geq \Pi^*(1) \geq \alpha IV$. Therefore, we must have $\Pi^* = \alpha IV$ and the optimal solution to problem (2) is given by the full information design in Proposition 1.

Second, consider $\alpha < 1/2$. Given the solution in equation (7), Lemma 3 implies that $L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu})) = 1$ or $L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu})) = L + 1$. Therefore, we need to "modify" the solution in equation (7) to ensure $L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu})) = L$. We solve problem (5) in Appendix and present the solution by the following two lemmas. Let us first introduce the following notation.

$$\kappa_L \equiv \frac{U - \max\left\{0, U - \frac{N-L}{2}c\right\}}{c - \max\left\{0, U - \frac{N-L}{2}c\right\}} > 1, \text{ for } L = 1, \cdots, N-1.$$

Moreover, for any $x \in \mathbb{R}$, let's define $x \pm 0$ as $x \pm \varepsilon$ with $\varepsilon > 0$ and $\varepsilon \to 0$.

Lemma 5 (Optimal Posterior Belief with Search Length of One)**.** *The solution to* $\Pi^*(1)$ *in problem (5) is:*

$$\begin{cases} \mu(1) = 1 \textit{ and } \mu(2) = \cdots = \mu(N) = 0, & \textit{if } 0 < c \leq \dfrac{2U}{N} \textit{ or } \alpha \geq \dfrac{1}{1+\kappa_1}, \\ \mu(1) = \mu(2) = \dfrac{1}{1+\kappa_1} - 0 \textit{ and} & \\ \quad \mu(3) = \cdots = \mu(N) = \dfrac{\kappa_1 - 1}{(N-2)(1+\kappa_1)} + 0, & \textit{otherwise.} \end{cases} \tag{8}$$

To understand Lemma 5 above, first notice that by equation (3), the platform's objective function under $L = 1$ is given by $\Pi(\boldsymbol{\mu}, 1) = \alpha\mu(1)IV + (1-\alpha)\mu(2)IV$, where $\alpha\mu(1)IV$ is the commission and $(1-\alpha)\mu(2)IV$ is the advertising revenue. We can understand Lemma 5 by comparing it with the solution in equation (7) under $L = 1$. There are essentially two differences between equations (7) and (8).

First, when $0 < c \leq 2U/N$, Lemma 5 implies that $\mu(1) = 1$ and $\mu(2) = \cdots = \mu(N) = 0$, regardless of the value of α. In this case, the platform gives up the advertising revenue and relies on commission solely. This is optimal because under low search cost $c \leq 2U/N$, consumers will search all sellers anyway and thus sellers have no incentive to bid for the prominent position. We have obtained the same result in Proposition 2 under no information. In contrast, equation (7) does not involve this condition on search cost, because the consumer's search length L is exogenously fixed.

Second, under a high search cost $c > 2U/N$ and low commission fee $\alpha < (1+\kappa_1)^{-1}$, we have $\mu(3) = \cdots = \mu(N) > 0$ in Lemma 5, rather than being zero in equation (7). This is because if $\mu(3) = \cdots = \mu(N) = 0$, then after visiting seller 1 in the prominent position and learning that it is not a match, a consumer will infer that seller 2 must be the match and will surely continue to visit seller 2, resulting in a search length of $L = 2$. To ensure that the search length is one, the platform needs to deter the consumer from searching seller 2 by allocating some positive match probability to $\mu(3), \cdots, \mu(N)$. At the same time, the platform wants to keep the total of these match probabilities, $\mu(3), \cdots, \mu(N)$, as small as possible so that it can maximize the posterior beliefs for sellers 1 and 2 so as to maximize commission and ad revenue.

We show in Appendix that the optimal way to allocate these probabilities is uniform allocation, $\mu(3) = \cdots = \mu(N) \equiv \mu_{tail}$. Intuitively, the consumer finds it most reluctant to continue to search if the expected search cost to identify the match is the highest. The uniform allocation can achieve this under the constraint $\mu(3) \geq \cdots \geq \mu(N)$. We can then apply Lemma 3 to derive the expected value of continuing search. In particular, upon knowing that seller 1 is not the match, the consumer is just unwilling to search seller 2 if

$$-c + \frac{\mu(2)}{\mu(2)+(N-2)\mu_{tail}}U + \frac{(N-2)\mu_{tail}}{\mu(2)+(N-2)\mu_{tail}}\max\{0, U - \frac{N-1}{2}c\} = 0 - 0.$$

We then have $\mu_{tail} = (\kappa_1 - 1)(N-2)^{-1}(1+\kappa_1)^{-1} + 0$. Given this result, maximizing the advertising revenue implies $\mu(1) = \mu(2) = (1+\kappa_1)^{-1} - 0$ which is less than $1/2$. That is, the advertising revenue is lower compared with the case of exogenous search length. Thus, the threshold for α in equation (8), $(1+\kappa_1)^{-1}$ is lower than $1/2$, which is the threshold in equation (7).

Next, we solves problem (5) for the other cases with $2 \leq L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu})) \leq N$. Notice that given $L = N$ or any $2 \leq L \leq N-1$ such that $\mu(L+1) = 0$, by equation (3) we have that the platform's profit come solely from the commission. In this case, the full information design characterized by Proposition 1 enables every consumer to be matched and thus maximizes the commission. Under the full information design, the search length is equal to 1. This implies that the cases with $L = N$ or $2 \leq L \leq N-1$ and $\mu(L+1) = 0$ are weakly dominated by the full information design with $L = 1$. Therefore, we can without loss of generality restrict ourselves to $\mu(L+1) > 0$ and $2 \leq L \leq N-1$ when solving problem (5).

Lemma 6 (Optimal Posterior Belief with Search Length of $L \geq 2$)**.** *If $\alpha \geq (1+\kappa_L)^{-1}$ or $L \geq N+1-2U/c$, there does not exist a solution to problem (5) that satisfies $\mu(L+1) > 0$; otherwise if $\alpha < (1+\kappa_L)^{-1}$ and $2 \leq L < N+1-2U/c$, the solution that satisfies $\mu(L+1) > 0$ is:*

$$\mu(n) = \frac{c}{U-c}\left(\frac{U}{U-c}\right)^{L-n-1}\frac{\kappa_L}{1+\left(\frac{U}{U-c}\right)^{L-1}\kappa_L} - 0, \textit{ for } n = 1, \cdots, L-1,$$

$$\mu(L) = \mu(L+1) = \frac{1}{1+\left(\frac{U}{U-c}\right)^{L-1}\kappa_L} - 0,$$

$$\mu(L+2) = \cdots = \mu(N) = \frac{1}{N-L-1}\cdot\frac{\kappa_L - 1}{1+\left(\frac{U}{U-c}\right)^{L-1}\kappa_L} + 0.$$

To understand Lemma 6, notice that similar to Lemma 5, there are two cases. When α is above the threshold of $(1+\kappa_L)^{-1}$, in Appendix, we show that the optimal design entails $\mu(L+1) = 0$ and the platform's profit comes from commission almost solely. In this case, the full information design is optimal as argued above. Moreover, if $L \geq N+1-2U/c$, there are only a few sellers left unexplored, after a consumer has visited seller $L-1$ and found it a mismatched. As such, the consumer will be tempted to continue to search and it is impossible for the search length to reach L. On the other hand, when α is below the threshold of $(1+\kappa_L)^{-1}$ and $2 \leq L < N+1-2U/c$, we still have $\mu(L) = \mu(L+1) > \mu(L+2) = \cdots = \mu(N)$, similar to the case of $L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu})) = 1$ in Lemma 5.

What is new here is $\mu(n)$ for $n = 1, \cdots, L-1$, which decreases exponentially over

n. They are set in a way such that the consumer is just willing to search seller $n+1$ after learning that seller n is not a match. This can minimize the probability allocated to $\mu(1), \cdots, \mu(L-1)$ and thus reserves most probability to $\mu(L+1)$ so as to maximize the advertising revenue.

Given the consumer's search length of L, she will stop searching after visiting L sellers, so her continuation value after visiting L sellers is zero. Given the previous $L-1$ sellers unmatched (including seller L at the prominent position), the consumer will search seller $L-1$ if and only if

$$-c + \frac{\mu(L-1)}{\mu(L-1) + \mu(L+1) + \sum_{n=L+2}^{N} \mu(n)} U \geq 0$$
$$\Leftrightarrow \mu(L-1) \geq \frac{c}{U-c} \left[\mu(L+1) + \sum_{n=L+2}^{N} \mu(n) \right].$$

In Appendix, we show that this constraint is binding so that $\mu(L-1) = \frac{c}{U-c}\big[\mu(L+1) + \sum_{n=L+2}^{N} \mu(n)\big]$. By applying the same argument above recursively, we have for $n = 1, \cdots, L-1$,

$$\begin{aligned} \mu(n) &= \frac{c}{U-c} \left[\sum_{k=1}^{L-n-1} \mu(L-k) + \mu(L+1) + \sum_{n=L+2}^{N} \mu(n) \right] \\ &= \frac{c}{U-c} \left(\frac{U}{U-c} \right)^{L-n-1} \left[\mu(L+1) + \sum_{n=L+2}^{N} \mu(n) \right]. \end{aligned}$$

This explains the exponential decline of $\mu(n)$ with respect to n for $n = 1, \cdots, L-1$. To further determine $\mu(n)$ for all $n \in \mathcal{N}$, we need to further utilize four other conditions: (1) $\mu(L) = \mu(L+1)$, (2) the consumer is indifferent between continuing to search and stopping after visiting L sellers, (3) the normalization condition such that $\sum_{n=1}^{N} \mu(n) = 1$, and (4) the uniform belief allocation for the long tail, $\mu(L+2) = \cdots = \mu(N)$. The details are relegated to Appendix.

4.4 Optimal Posterior Belief

After solving problem (5) by Lemmas 5 and 6, we can compare $\Pi^*(L)$ for all $L = 1, \cdots, N$ to determine the optimal search length for the platform. We prove the following proposition in Appendix. To simplify notation, we define:

$$\alpha^* \equiv \frac{1}{1+\kappa_1} = \frac{c - \max\left\{0, U - \frac{N-1}{2}c\right\}}{U + c - 2\max\left\{0, U - \frac{N-1}{2}c\right\}}.$$

Proposition 4 (Optimal Posterior Belief)**.**

1. *If $0 < c \leq 2U/N$, any posterior belief $\boldsymbol{\mu}$ yields the same profit for the platform.*

2. *Otherwise, if $2U/N < c < U$, the optimal posterior belief $\boldsymbol{\mu}^*$ is given by equation (8). That is,*

$$\begin{cases} \mu^*(1) = 1 \text{ and } \mu^*(2) = \cdots = \mu^*(N) = 0, & \text{if } \alpha \geq \alpha^*, \\ \mu^*(1) = \mu^*(2) = \alpha^* - 0 \text{ and } \mu^*(3) = \cdots = \mu^*(N) = \frac{1-2\alpha^*}{N-2} + 0, & \text{otherwise,} \end{cases} \tag{9}$$

under which, consumers' search length is equal to one so that they only visit the prominent seller.

When $0 < c \leq 2U/N$, Proposition 3 has already shown that the full and no information yield the same platform profit. Proposition 4 extends this result to any information design. This result is quite intuitive—when consumers' search cost is sufficiently low, they will always visit all sellers to find the match, leaving little room for the platform to design an information environment so as to influence their search behaviors. In contrast, when $2U/N < c < U$, Proposition 4 implies that it is optimal for the platform to design the information environment that induces consumers' search length to be one.[3] Essentially, we find that $L = 1$ maximizes $\Pi^*(L)$.

To understand why it is optimal for the platform to limit consumer search, notice that the platform's revenue $\Pi^*(L)$ consists of two parts: commission and advertising revenue. By Lemmas 5 and 6, we can calculate the platform's advertising revenue under consumer search length of L:

$$(1-\alpha)IV\mu(L+1) = \frac{(1-\alpha)IV}{1 + \left(\frac{U}{U-c}\right)^{L-1} \kappa_L} - 0,$$

which decreases with L as κ_L increases with L. This implies that $L = 1$ maximizes the platform's advertising revenue. The intuition is that by limiting consumer search, the platform incentivizes the sellers to bid high because losing the auction means no demand. When the platform's commission rate $\alpha < \alpha^*(< 1/2)$, advertising revenue is more important. Therefore, it is optimal for the platform to limit consumer search so as to maximize advertising revenue. However, when $\alpha \geq \alpha^*$, as shown by Proposition 1, full information design maximizes the platform's commission, under which the consumer's search length is also equal to one.

4.5 Optimal Information Design

We now construct an information structure $\{\pi(s|n)\}_{n \in \mathcal{N}}$ that can implement the optimal posterior belief $\boldsymbol{\mu}^*$ in Proposition 4. The following proposition presents one optimal information design as well as the welfare of consumers, sellers and the platform under the optimal design.

Proposition 5 (Optimal Information Design)**.**

[3] It is possible that there exist other optimal designs that yield the same maximum platform profit.

1. *If $0 < c \leq 2U/N$, any information design $\pi(\cdot|\cdot)$ yields the same profit of αIV for the platform.*

2. *If $2U/N < c < U$ and $\alpha \geq \alpha^*$, the optimal information design entails full information disclosure with $\pi^*(n|n) = 1$ and $\pi^*(s|n) = 0$ for $s \neq n$. In equilibrium:*

   (a) all sellers bid zero, and the matched seller, n wins and earns a profit of $(1-\alpha)V$;

   (b) consumers buy the prominent products and do not continue to search, each expecting a surplus of U;

   (c) the platform obtains a profit of αIV.

3. *If $2U/N < c < U$ and $\alpha < \alpha^*$, one optimal information design is:*

$$\pi^*(n|n) = \pi^*(n+1|n) = \alpha^* - 0 \text{ and } \pi^*(s|n) = \frac{1-2\alpha^*}{N-2} + 0 \text{ for } s \neq n, n+1,$$

   where we have used the cyclic indexing so that $\pi^*(N+1|N) \equiv \pi^*(1|N)$. In equilibrium:

   (a) sellers s and $s-1$ bid $(1-\alpha)V\alpha^* - 0$, and the other sellers bid $(1-\alpha)V\frac{1-2\alpha^*}{N-2} + 0$; sellers s and $s-1$ win the auction with equal probability and make zero profit;

   (b) consumers buy the prominent products and do not continue to search, each expecting a surplus of $U\alpha^* - 0$;

   (c) the platform obtains a profit of $\alpha^* IV - 0$.

We make a few remarks about the implications of the optimal information design next. Let us first focus on the case with $2U/N < c < U$ so that the platform's information design is relevant. There are two different scenarios depending on whether α is above or below the threshold of α^*. Casual observations suggest that, in practice, the commission rate α typically ranges from 5% to 20% on e-commerce platforms. The model suggests that the threshold α^* is less than $1/2$. This implies that either of these two cases could be relevant in practice. Furthermore, notice that the full information design implemented under $\alpha \geq \alpha^*$ is the first-best design that maximizes the social welfare. Proposition 5 then implies that the option of offering targeted advertising by the platform will introduce social inefficiency to the market when $\alpha < \alpha^*$, where the social waste is $I(U+V)(1-\alpha^*)$.[4]

Given the optimal design involves "noisy matching", the platform may find it sometimes unattractive to invest in data analytics technologies that produce precise targeted advertising and recommendation. It also implies that decreased data availability due to regulations on consumer privacy does not necessarily hurt platforms that much. In fact, the no-information case characterized by Proposition 2 is the market equilibrium when the consumer opts out of the platform's data collection. It is straightforward to show

[4] This finding depends on the assumption that the platform has perfect information about which seller is the match for each consumer. With imperfect match information, targeted advertising can increase match efficiency by allowing matched sellers to bid higher and win the prominent position.

that the consumer surplus is always weakly higher under the platform's optimal information design compared with that under no information. Therefore, if we endogenize consumers' opt-in decision on the platform's data collection, they will indeed choose to opt in under our model. This will also weakly benefit the platform and the seller.

Notice that α^* increases with both c and N. Proposition 5 then implies that a platform's profit under the optimal information design increases in consumer's search cost c as well as the number of sellers. A higher search cost reduce consumers' incentive to search beyond the prominent seller, leading to competitive ad bidding of the sellers. This can benefit the platform when it cannot demand a high commission fee and thus relies on persuading sellers to bid for the prominent position. Similarly, as N increases, consumers are less willing to search, and thus the platform profit increases.

The match efficiency and social welfare under the optimal information design are non-monotonic in the number of sellers, N. When $N \leq 2U/c$, the social welfare is maximized and invariant to N under the full disclosure (assuming that when the platform is indifferent among several information designs, he chooses the one that maximizes the social welfare). As N increases just above $2U/c$, the social welfare drops discretely because mismatches are introduced to discourage consumer search under the optimal information design. As N further increases, consumers are less willing to search, and thus the mismatches are gradually reduced.

5 Concluding Remarks

In this research, we highlight a looming trade-off in the two revenue streams that resulted from ubiquitous usage of big data in online retail platforms. While big data can improve match efficiency, it enhances a seller's market power. This effect, however, may reduce sellers' competition for advertising, limiting the platform's ability to extract sellers' surplus. The optimal information design that takes into account this strategic trade-off entails limiting consumer search and mixing the matched product with a long tail of unmatched ones for recommendation. Our research has a few limitations, and thus we invite for future research.

First, we consider a specific form of match value distribution among sellers such that there is one and only match for each consumer. This leads to correlated match values among sellers. This setting approximates situations in reality where upon one purchase occasion, each consumer has a favorite seller in one product category and prefers this seller much more than the others. Alternative specification such as independent and identically distributed match values of sellers implies a multi-dimensional state space which greatly complicates the information design problem. In a setting similar with ours but with two or more matches, the information design problem becomes trivial, where the optimal design is to disclose information fully, because it always ensures match and thus maximizes the commission while at the same time also results in the maximum advertising revenue.

Second, consistent with the majority of the literature on online platforms, our analysis assumes exogenous linear commission rate, α. Notice that sellers earn zero profit under the optimal information design, and thus we cannot simply endogenize α by considering an elastic demand of seller entry or platform competition. Hence, it seems to go beyond the current model setup to fully endogenize α. We leave this as an important and interesting direction for future research. More complex contracts such as two-part tariff will make our result trivial, because the platform can simply set $\alpha = 1$ while at the same time pay the sellers a lump-sum payment. This will resolve the platform's incentive to extract seller surplus via sponsored advertising. In other words, our result requires the vertical relationship between the platform and the sellers to be not fully integrated so that there is some room for sponsored advertising to play a role. This seems reasonable given casual observations from the practice, where even though two-part tariffs have been used, it is the sellers instead of the platform who pay the lump-sum fee.

Third, the model assumes a simple second-price auction without a reserved price. This assumption is made partly without loss of generality, allowing us to focus on the main forces underlying the information design problem. It is easy to verify that, under the optimal information design, augmenting the auction with a reserved price (or simply offering a take-it-or-leave-it contract to the sellers) does not improve the profit of the platform. However, if the platform is allowed to set the reserved price for any auction design, it is not clear whether the same equilibrium is reached. The reason is that the reserved price only changes the sellers' bidding strategies without a direct impact on buyers' search behavior; however, in the model we consider, the platform can fine-tune the public beliefs to directly impact both the sellers' bidding strategies and buyers' search behavior.

Last, a potential direction for future research is to investigate the optimal information design that allows private signals—one to sellers and one to consumers.

APPENDIX

Proof of Lemma 1:

Proof. Consider any seller n with a positive demand. All consumers in the market must be one of the following three types: (1) those who have visited the seller and found a match, (2) those who have visited the seller and did not find a match, (3) those who decide not to visit the seller. Consumers of type (2) will never buy from seller n regardless of his price so they are irrelevant. Consumers of type (3) do not observe seller n's price, which has no impact on their visiting decision. So, the only relevant consumers are type (1), who have already paid the search cost if any and have no interest for other sellers because seller n is the only possible match, so the seller will charge the price p_n so as to maximize the monopolistic profit, $(1-\alpha)p_n D(p_n)$, which yields $p_n = p^*$. □

Proof of Lemma 2:

Proof. To simplify notation without introducing confusion, we can omit the subscript s and denote a consumer's posterior belief of product n being the match by $\mu(n)$. We only need to show that at any time point, given the posterior belief $\mu(1) > \mu(2)$, if the consumer decides to search, she weakly prefers to search seller 1 before seller 2. We prove by contradiction. Suppose it generates a strictly higher payoff for the consumer following a search order of $2, 3, \cdots, k, 1, k+1, \cdots, N$. Notice that we do not impose any constraint on $\mu(i)$ for $i = 1, \cdots, N$ except that $\mu(1) > \mu(2)$. The consumer will stop searching if she has found a match. We use z_k (for $k = 1, \cdots, N$) to represent the consumer's search decision so that if $z_k = 1$, the consumer decides to continue to search seller k given she has not found any match so far, and if $z_k = 0$, the consumer will stop searching. Under the proposed search strategy, the consumer's expected payoff is

$$U_2 = \left[z_2\mu(2) + \sum_{l=3}^{k} \left(\prod_{j=2}^{l} z_j \right) \mu(l) + \left(\prod_{j=2}^{k} z_j \right) z_1\mu(1) + \sum_{l=k+1}^{N} \left(\prod_{j=1}^{l} z_j \right) \mu(l) \right] U.$$

Now we propose an alternative search strategy, under which, the consumer follows the search order of $1, 3, \cdots, k, 2, k+1, \cdots, N$, and for each seller: applies z_1 for seller 2, z_2 for seller 1, and z_k for seller $k = \{3, \cdots, N\}$. Under the alternative search strategy, the consumer's expected payoff is

$$U_1 = \left[z_2\mu(1) + \sum_{l=3}^{k} \left(\prod_{j=2}^{l} z_j \right) \mu(l) + \left(\prod_{j=2}^{k} z_j \right) z_1\mu(2) + \sum_{l=k+1}^{N} \left(\prod_{j=1}^{l} z_j \right) \mu(l) \right] U.$$

By noting $z_k \leq 1$ for $k = 1, \cdots, N$, we have,

$$U_1 - U_2 = (\mu(1) - \mu(2))z_2 \left[1 - z_1 \left(\prod_{j=3}^{k} z_j \right) \right] U \geq 0.$$

This means that the alternative strategy generates weakly higher payoff than the proposed one. This is a contradiction, so we have proved the original statement. □

Proof of Lemma 3:

Proof. We reformulate the consumer's problem after seeing an unmatched advertised product as a dynamic programming problem. There are at most $N-1$ periods. Let W_k denote the optimal continuation value for the consumer in period k, given that no match has been found prior to and including period k. This implies that there are $N-k$ products left that are likely to be a match. We have the following Bellman equation for every $k = 1, 2, \cdots, N-1$:

$$W_k = \max\Big\{0, \underbrace{-c + \frac{1}{N-k}U + \frac{N-k-1}{N-k}W_{k+1}}_{\equiv \widehat{W}_k}\Big\}, \tag{10}$$

where the two terms in the curly brace above are respectively, the outside option and the value of continuing to search, denoted by $\widehat{W}_k$. By paying c, the consumer gets a match and thus the surplus of U with probability $1/(N-k)$ and with the remaining probability of $1 - 1/(N-k)$, the consumer does not find a match and expects the continuation value of W_{k+1}. Based on equation (10), next, we show the optimality of the proposed rule.

We first show that if $c \leq 2U/N$ or equivalently, $U \geq cN/2$, the consumer will continue to search until she finds a match. In fact, if none of the previous $N-1$ products is a match, the consumer is sure that the next and final product must be a match, so we have $\widehat{W}_{N-1} = U - c$. By applying equation (10) and the condition of $U \geq cN/2$ iteratively, we have

$$\widehat{W}_k = U - \frac{N-k+1}{2}c \text{ and } \widehat{W}_{N-1} \geq \widehat{W}_{N-2} \geq \cdots \geq \widehat{W}_1 \geq 0.$$

Therefore, the consumer will keep searching until she finds a match.

Next, we show that if $2U/N < c < U$ or equivalently, $c < U < cN/2$, the consumer will never search. In fact, given $c < U < cN/2$, there must exist $k^* \in \mathbb{N}$ and $1 \leq k^* \leq N-2$ such that $(N-k^*)c/2 \leq U < (N-k^*+1)c/2$. Starting from $\widehat{W}_{N-1} = U - c$, by applying equation (10) iteratively, we will have $W_1 = \cdots = W_{k^*} = 0$. Therefore, the consumer will never search in the first place. □

Proof of Lemma 4

Proof. Let's calculate $\Pi(\boldsymbol{\mu})$. First it is straightforward to show that seller n's bid satisfies that $0 \leq b_n^*(\boldsymbol{\mu}) \leq (1-\alpha)\mu(n)V$, because he has the outside option of zero and $(1-\alpha)\mu(n)V$ is the maximum payoff he could expect from winning the auction. Given the consumer's equilibrium search length as $L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu}))$, sellers $1, \cdots, L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu})) - 1$ will be visited if they are the match, so they will bid zero. This can be seen by noticing that they have no incentive to deviate by bidding positively to take the prominent position. This implies that the winner of the auction must be among sellers $L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu})), \cdots, N$;

that is, $n^*(\boldsymbol{\mu}) \geq L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu}))$.

Let's prove $n^*(\boldsymbol{\mu}) = L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu}))$ next. There are two cases to consider. First, if $\mu(L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu}))) = \cdots = \mu(n^*)$, the difference between sellers $L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu})), \cdots, n^*(\boldsymbol{\mu})$ is artificial and one can renumber sellers $L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu})), \cdots, n^*(\boldsymbol{\mu})$ such that $n^*(\boldsymbol{\mu}) = L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu}))$. Second, if there is a strict inequality in $\mu(L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu}))) \geq \cdots \geq \mu(n^*)$, we must have $\mu(L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu}))) > \mu(n^*)$. We prove $n^* = L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu}))$ by contradiction. Suppose $n^*(\boldsymbol{\mu}) \geq L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu}))+1$. Consider seller $L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu}))$, who will never be visited in equilibrium and thus expects zero profit. However, the seller can deviate by bidding $(1-\alpha)\mu(L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu})))V - \varepsilon > (1-\alpha)\mu(n^*)V \geq b^*_{n^*}$ to win the auction for $\varepsilon > 0$ but sufficiently small and get positive profit. This implies that it is not an equilibrium for $n^*(\boldsymbol{\mu}) \geq L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu})) + 1$. Therefore, we must have $n^* = L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu}))$. That is, seller $L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu}))$ wins the auction.

For sellers $n = L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu})) + 1, \cdots, N$, their profits in equilibrium is zero. If they deviate to win the auction, their profit become $(1-\alpha)\mu(n)V$. This implies that his equilibrium bid is $b^*_n = (1-\alpha)\mu(n)V$. Given seller $L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu}))$ wins the auction, the second-highest bidder must be seller $L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu})) + 1$, and thus the platform's advertising revenue is $(1-\alpha)\mu(L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu})) + 1)IV$. Given consumer's search length as $L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu}))$, the platform's commission fee is $\alpha IV \sum_{n=1}^{L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu}))} \mu(n)$. By equation (3), we have $\Pi(\boldsymbol{\mu}) = \Pi(\boldsymbol{\mu}, L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu})))$. □

Proof of Lemmas 5 and 6:

Proof. We first fix $\mu(L+1) = a$ and then optimize over a later. Problem (5) becomes

$$\begin{aligned} &\max_{\mu(1),\cdots,\mu(L),\mu(L+2),\cdots,\mu(N)} \alpha \sum_{n=1}^{L} \mu(n) + (1-\alpha)a, \qquad (11) \\ s.t.\ &\mu(1) \geq \cdots \geq \mu(L) \geq a \geq \mu(L+2) \geq \cdots \geq \mu(N), \\ &\sum_{n=1}^{L} \mu(n) + \sum_{n=L+2}^{N} \mu(n) = 1 - a, \\ &L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu})) = L. \end{aligned}$$

Notice that the objective function of problem (11) increases with $\mu(1), \cdots, \mu(L)$ but does not depend on $\mu(L+2), \cdots, \mu(N)$. The normalization constraint, $\sum_{n=1}^{L} \mu(n) + \sum_{n=L+2}^{N} \mu(n) = 1 - a$ then implies that we need to minimize $\sum_{n=L+2}^{N} \mu(n)$. There are two cases to consider depending on whether $a = 0$.

If $a = 0$, the platform's profit comes from commission solely. Full information design as shown by Proposition 1 enables all consumers to be matched with the right seller and therefore maximizes the commission, in which case we have the consumers' search length $L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu}))$ equal to 1. This implies that we do not need to analyze the cases with $L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu})) \geq 2$, because they are weakly dominated by the case with $L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu})) = 1$.

It remains to solve problem (11) given $a > 0$. We will first analyze the cases with $2 \leq L \leq N-2$. The other cases with $L = 1$ and $N-1$ will be analyzed afterwards.

Step 1: Minimizing $\sum_{n=L+2}^{N} \mu(n)$.

We shall minimize the long tail $\sum_{n=L+2}^{N} \mu(n)$ subject to the constraints (i) $a \geq \mu(L+2) \geq \cdots \geq \mu(N-1) \geq \mu(N) = b$ and (ii) $L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu})) = L$. Notice that onstraint (ii) implies that a consumer does not search beyond the first L sellers, if she has not found a match among them. That is,

$$-c + \frac{a}{a + \sum_{n=L+2}^{N} \mu(n)} U + \frac{\sum_{n=L+2}^{N} \mu(n)}{a + \sum_{n=L+2}^{N} \mu(n)} W_{L+1} \leq 0, \tag{12}$$

where W_{L+1} is the continuation value after seeing $L+1$ sellers unmatched. Next, we show that it is sufficient to consider a uniform belief allocation, $\mu(L+2) = \cdots = \mu(N) = b > 0$.

To see that, suppose we have a strict inequality in constraint (i) such that $\mu(N-1) > \mu(N)$. We can reduce $\mu(N-1)$ by a small amount of ϵ, $\mu'(N-1) = \mu(N-1) - \epsilon$, while raising $\mu(N)$ for the same amount, $\mu'(N) = \mu(N) + \epsilon$. In doing so, we can keep $\sum_{n=L+2}^{N} \mu(n)$ unchanged. However, W_{L+1} will be reduced. First,

$$\begin{aligned}
W'_{N-2} &= \max\{0, -c + \frac{\mu'(N-1)}{\mu'(N-1) + \mu'(N)} U + \frac{\mu'(N)}{\mu'(N-1) + \mu'(N)} W'_{N-1}\} \\
&= \max\{0, -c + \frac{\mu'(N-1)}{\mu'(N-1) + \mu'(N)} (U - W'_{N-1}) + W'_{N-1}\} \\
&= \max\{0, -c + \frac{\mu(N-1) - \epsilon}{\mu(N-1) + \mu(N)} (U - W_{N-1}) + W_{N-1}\} \\
&\leq W_{N-2},
\end{aligned}$$

where $W'_{N-1} = W_{N-1} = U - c$, and $\mu'(N-1) + \mu'(N) = \mu(N-1) + \mu(N)$. For all earlier periods $k = L+2, \cdots, N-3$, the belief allocation remains unchanged, and thus $\mu(k+1) + \cdots + \mu'(N-1) + \mu'(N) = \sum_{n=k+1}^{N} \mu(n)$. Then recursively,

$$\begin{aligned}
W'_k &= \max\{0, -c + \frac{\mu(k+1)}{\mu(k+1) + \sum_{n=k+2}^{N} \mu(n)} U + \frac{\sum_{n=k+2}^{N} \mu(n))}{\mu(k+1) + \sum_{n=k+2}^{N} \mu(n))} W'_{k+1}\} \\
&\leq \max\{0, -c + \frac{\mu(k+1)}{\mu(k+1) + \sum_{n=k+2}^{N} \mu(n)} U + \frac{\sum_{n=k+2}^{N} \mu(n))}{\mu(k+1) + \sum_{n=k+2}^{N} \mu(n))} W_{k+1}\} \\
&= W_k.
\end{aligned}$$

Hence, the LHS of inequality (12) becomes smaller because $\sum_{n=L+2}^{N} \mu(n)$ remains unchanged but $W'_{L+1} \leq W_{L+1}$. Lastly, we can turn the strict inequality $\mu(N-1) > \mu(N)$ into equality $\mu'(N-1) = \mu'(N)$ by setting $\epsilon = (\mu(N-1) - \mu(N))/2 > 0$. We can repeatedly apply the same procedure above recursively for any strict equality $\mu(k) > \mu(k+1) = \mu(k+2) = \cdots \mu(N)$. That is, we first reduce $\mu(k)$ by a small amount of ϵ, $\mu'(k) = \mu(k) - \epsilon$, and then distribute this amount uniformly to $\mu(k+1), \cdots, \mu(N)$. By the same argument above, we can keep the inequality (12) being satisfied given

$W'_{L+1} \leq W_{L+1}$. Setting $\epsilon = (\mu(k) - \mu(k+1))(N-k)/(N-k+1)$, we can ensure that $\mu'(k) = \mu'(k+1) = \mu'(k+2) = \cdots \mu'(N)$.

To summarize, applying the above argument recursively for $k = N-1, N-2, \cdots, L+1$, we can turn any original belief allocation with some strict inequalities into a uniform one $\mu(L+2) = \cdots = \mu(N)$, while keeping the sum of beliefs unchanged and satisfying the constraint (ii). Let $b = \mu(L+2) = \cdots = \mu(N)$. Minimizing $\sum_{n=L+2}^{N} \mu(n)$ is then equivalent to minimizing b. Notice that $b \neq 0$, because otherwise the LHS of inequality (12) becomes $U - c$, which is positive, violating the incentive constraint (ii).

Given the uniform belief allocation, we can apply Lemma 3 and establish that, $W_{L+1} = \max\{0, U - (N-L)c/2\}$. The LHS of inequality (12) can then be rewritten as

$$\begin{aligned} & -c + \frac{a}{a+(N-L-1)b}U + \frac{(N-L-1)b}{a+(N-L-1)b}\max\{0, U - \frac{N-L}{2}c\}, \\ = & -c + \frac{a}{a+(N-L-1)b}(U - \max\{0, U - \frac{N-L}{2}c\}) + \max\{0, U - \frac{N-L}{2}c\}, \end{aligned}$$

which is decreasing in b. Clearly, we should minimize b such that the inequality (12) holds with equality. That is,

$$b = \frac{U-c}{(N-L-1)\left(c - \max\left\{0, U - \frac{N-L}{2}c\right\}\right)}a + 0, \tag{13}$$

Lastly, we need to impose the condition that $a \geq b > 0$, which is equivalent to

$$L < N + 1 - \frac{2U}{c}, \tag{14}$$

which does not depend on a. If the condition (14) is violated, we cannot ensure the search length is equal to L. Notice that $c < U$ under Assumption 1, so condition (14) implies that $L < N-1$. Therefore, the case of $L = N-1$ can be ruled out.

Step 2: Lower bounds of $\mu(1), \cdots, \mu(L)$.

Given equation (13) and condition (14), problem (11) becomes:

$$\begin{aligned} & \max_{\mu(1),\cdots,\mu(L)} \alpha\left[1 - a - (N-L-1)b\right] + (1-\alpha)a, \\ s.t.\ & \mu(1) \geq \cdots \geq \mu(L) \geq a, \\ & \sum_{n=1}^{L} \mu(n) = 1 - a - (N-L-1)b, \\ & L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu})) = L. \end{aligned} \tag{15}$$

The objective function of problem (15) does not depend on $\mu(1), \cdots, \mu(L)$, so we only need to check the feasibility of the constraints in problem (15). Next, we first identify the constraints that $\mu(1) \geq \cdots \geq \mu(L) \geq a$ and $L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu})) = L$ impose on $\mu(1), \cdots, \mu(L)$.

Given seller L wins the prominent position, we have the lower bound on $\mu(L)$ as,

$$\underline{\mu}(L) = a. \tag{16}$$

To determine the lower bound on $\mu(L-1)$, first notice that given the previous L sellers unmatched, the consumer will stop without visiting seller $L+1$, and her continuation value is zero. Given the previous $L-1$ sellers unmatched, the consumer will search seller $L-1$ if and only if

$$-c+\frac{\mu(L-1)}{\mu(L-1)+a+(N-L-1)b}U \geq 0 \Leftrightarrow \mu(L-1) \geq \frac{c}{U-c}[a+(N-L-1)b] \equiv \underline{\mu}(L-1).$$

Let's show that $\underline{\mu}(L-1) = \frac{c}{U-c}[a+(N-L-1)b] \geq a = \underline{\mu}(L)$. In fact, given the expression of b in equation (13), $\frac{c}{U-c}[a+(N-L-1)b] \geq a$ is equivalent to,

$$c^2 + (U-2c)\max\left\{0, U - \frac{N-L}{2}c\right\} \geq 0.$$

If $U \geq 2c$ or $U \leq (N-L)c/2$, obviously, the above inequality holds; otherwise when $(N-L)c/2 < U < 2c$, we have the above inequality equivalent to,

$$\left[U - \left(1+\frac{N-L}{4}\right)c\right]^2 + \frac{N-L}{2}\left(1-\frac{N-L}{8}\right)c^2 \geq 0,$$

which always holds, because $(N-L)c/2 < U < 2c$ implies that $N-L < 4$. Therefore, we have shown that $\underline{\mu}(L-1) \geq \underline{\mu}(L)$. This implies that when $\mu(L)$ takes its lower bound of $\underline{\mu}(L)$, $\mu(L-1) \geq \mu(L)$ is not binding, given $\mu(L-1) \geq \underline{\mu}(L-1)$.

By applying a similar argument that determines $\underline{\mu}(L-1)$ above iteratively, we have,

$$\begin{aligned}\underline{\mu}(n) &= \frac{c}{U-c}[a+(N-L-1)b] + \frac{c}{U-c}\sum_{k=1}^{L-n-1}\mu(L-k) \\ &= \frac{c}{U-c}\left(\frac{U}{U-c}\right)^{L-n-1}[a+(N-L-1)b], \text{ for } n = 1, \cdots, L-1.\end{aligned} \tag{17}$$

Obivously, $\underline{\mu}(1) \geq \cdots \geq \underline{\mu}(L-1)$. To summarize, $(\underline{\mu}(1), \cdots, \underline{\mu}(L))$ is the lowest values $(\mu(1), \cdots, \mu(L))$ can take given the constraints of $\mu(1) \geq \cdots \geq \mu(L) \geq a$ and $L(\boldsymbol{\mu}, n^*(\boldsymbol{\mu})) = L$.

Now, we consider the remaining constraint of $\sum_{n=1}^{L}\mu(n) = 1-a-(N-L-1)b$ in problem (15). We only need to ensure $\sum_{n=1}^{L}\underline{\mu}(n) \leq 1-a-(N-L-1)b$, which reduces to,

$$a + \left(\frac{U}{U-c}\right)^{L-1}[a+(N-L-1)b] \leq 1. \tag{18}$$

Step 3: Solving for the optimal $\mu(L+1)$.

To summarize, we have identified the feasibility condition for problem (15) as in equation (18). Given $a > 0$ and $2 \leq L \leq N-2$, by substituting the expression of b in equation (13) into problem (15), we can can rewrite problem (5) in terms of a as the following:

$$\max_{0<a\leq\frac{1}{L+1}} \alpha \left[1 - \frac{U - \max\left\{0, U - \frac{N-L}{2}c\right\}}{c - \max\left\{0, U - \frac{N-L}{2}c\right\}} a\right] + (1-\alpha)a$$
$$s.t.\ L < N + 1 - \frac{2U}{c},$$
$$\left[1 + \left(\frac{U}{U-c}\right)^{L-1} \frac{U - \max\left\{0, U - \frac{N-L}{2}c\right\}}{c - \max\left\{0, U - \frac{N-L}{2}c\right\}}\right] a < 1.$$

The optimal solution to the optimization problem is:

$$a = \left[1 + \left(\frac{U}{U-c}\right)^{L-1} \frac{U - \max\left\{0, U - \frac{N-L}{2}c\right\}}{c - \max\left\{0, U - \frac{N-L}{2}c\right\}}\right]^{-1} - 0,$$
$$\text{if } L < N + 1 - \frac{2U}{c} \text{ and } \alpha < \left[1 + \frac{U - \max\left\{0, U - \frac{N-L}{2}c\right\}}{c - \max\left\{0, U - \frac{N-L}{2}c\right\}}\right]^{-1}. \quad (19)$$

Notice that if $\alpha \geq \left[1 + \frac{U - \max\left\{0, U - \frac{N-L}{2}c\right\}}{c - \max\left\{0, U - \frac{N-L}{2}c\right\}}\right]^{-1}$, it is optimal for $a = 0$, which has been discussed above. Moreover, under the optimal solution in equation (19), $\mu(n) = \underline{\mu}(n)$ for $n = 1, \cdots, L$. This also immediately implies that $a \leq 1/(L+1)$ is not binding under the optimal solution in equation (19), because we have shown that $\underline{\mu}(1) \geq \cdots \underline{\mu}(L) \geq a$, which implies that $a \leq \frac{1}{L+1} \sum_{n=1}^{L+1} \underline{\mu}(n) = [1 - (N-L-1)b]/(L+1) < 1/(L+1)$.

Completing the proof.

So far we have solved problem (11) for $2 \leq L \leq N-2$. Now, let's consider the remaining cases of $L = 1$ and $L = N-1$. Notice that the case of $L = N-1$ has been ruled out in step 1 above. If $L = 1$, then following a similar analysis as above can straightforwardly show that $\mu(3) = \cdots = \mu(N) = b$ is still given by equation (13) and $\mu(1) = \underline{\mu}(1) = 1 - a - (N-2)b$. $\mu(2) = a$ given $a > 0$ is:

$$a = \left[1 + \frac{U - \max\left\{0, U - \frac{N-1}{2}c\right\}}{c - \max\left\{0, U - \frac{N-1}{2}c\right\}}\right]^{-1} - 0, \text{ if } c > \frac{2U}{N} \text{ and } \alpha < \left[1 + \frac{U - \max\left\{0, U - \frac{N-1}{2}c\right\}}{c - \max\left\{0, U - \frac{N-1}{2}c\right\}}\right]^{-1}.$$

Lastly, by comparing the platform's profit between the two cases with $a = 0$ and $a > 0$ given by equation (19), we obtain the solution to problem (5). $\square$

Proof of Proposition 4:

Proof. If $0 < c \leq 2U/N$, Proposition 2 shows that the consumer's search length is N given $\mu(1) = \cdots \mu(N) = 1/N$. As argued in the proof of Lemmas 5 and 6 above, $\mu(1) = \cdots \mu(N) = 1/N$ is the case where the consumer is most reluctant to search because of the highest expected search cost. This implies that under any posterior belief $\boldsymbol{\mu}$, the consumer's search length will be N, and the consumer will always find the match and the platform's profit is αIV.

On the other hand, if $2U/N < c < U$, denote the solution to problem (5) under the search length of L as $\boldsymbol{\mu}_L$. We only need to show that the platform's profit presented in Lemma 6, $\Pi(\boldsymbol{\mu}_L, L)$ is dominated by that in Lemma 5, $\Pi(\boldsymbol{\mu}_1, 1)$. In fact, by the discussion above Lemma 6, we only need to consider the case with $\alpha < (1+\kappa_L)^{-1}$, under which, we have

$$\Pi(\boldsymbol{\mu}_L, L) = \left[\alpha + \frac{1-\alpha(1+\kappa_L)}{1+\left(\frac{U}{U-c}\right)^{L-1}\kappa_L}\right] IV \leq \frac{1}{1+\kappa_1} IV \leq \Pi(\boldsymbol{\mu}_1, 1),$$

where in the second last inequality above, we have utilized that κ_L increases with L. □

References

Ajay Agrawal, Joshua Gans, and Avi Goldfarb. *Prediction machines: the simple economics of artificial intelligence*. Harvard Business Press, 2018.

Simon P Anderson and Regis Renault. Pricing, product diversity, and search costs: A bertrand-chamberlin-diamond model. *RAND Journal of Economics*, pages 719–735, 1999.

Mark Armstrong and Jidong Zhou. Consumer information and the limits to competition. *American Economic Review*, 112(2):534–577, 2022.

Susan Athey and Glenn Ellison. Position auctions with consumer search. *The Quarterly Journal of Economics*, 126(3):1213–1270, 2011.

Dirk Bergemann and Alessandro Bonatti. Targeting in advertising markets: implications for offline versus online media. *The RAND Journal of Economics*, 42(3):417–443, 2011.

Dirk Bergemann, Benjamin Brooks, and Stephen Morris. The limits of price discrimination. *American Economic Review*, 105(3):921–957, 2015.

Dirk Bergemann, Benjamin Brooks, and Stephen Morris. Search, information, and prices. *Journal of Political Economy*, 129(8):2275–2319, 2021.

Simon Board and Jay Lu. Competitive information disclosure in search markets. *Journal of Political Economy*, 126(5):1965–2010, 2018.

Yongmin Chen and Chuan He. Paid placement: Advertising and search on the internet. *The Economic Journal*, 121(556):F309–F328, 2011.

Peter A Diamond. A model of price adjustment. *Journal of Economic Theory*, 3(2):156–168, 1971.

Mustafa Dogan and Ju Hu. Consumer search and optimal information. *The Rand Journal of Economics*, 53(2):386–403, 2022.

Maarten CW Janssen and T Tony Ke. Searching for service. *American Economic Journal: Microeconomics*, 12(1):188–219, 2020.

Baojun Jiang and Tianxin Zou. Consumer search and filtering on online retail platforms. *Journal of Marketing Research*, 57(5):900–916, 2020.

Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101(6):2590–2615, 2011.

T Tony Ke and Song Lin. Informational complementarity. *Management Science*, 66(8): 3699–3716, 2020.

Jonathan Levin and Paul Milgrom. Online advertising: Heterogeneity and conflation in market design. *American Economic Review*, 100(2):603–07, 2010.

Andrew Rhodes. Multiproduct retailing. *The Review of Economic Studies*, 82(1):360–390, 2015.

Andrew Rhodes and Jidong Zhou. Consumer search and retail market structure. *Management Science*, 65(6):2607–2623, 2019.

Andrew Rhodes, Makoto Watanabe, and Jidong Zhou. Multiproduct intermediaries. *Journal of Political Economy*, 129(2):000–000, 2021.

Michael Schrage. The transformational power of recommendation. *MIT Sloan Management Review*, 62(2):17–21, 2021.

Martin L Weitzman. Optimal search for the best alternative. *Econometrica*, pages 641–654, 1979.

Mark Whitmeyer. Persuasion produces the (diamond) paradox. Working paper, 2021.

Asher Wolinsky. True monopolistic competition as a result of imperfect information. *Quarterly Journal of Economics*, 101(3):493–511, 1986.

Jidong Zhou. Multiproduct search and the joint search effect. *American Economic Review*, 104(9):2918–39, 2014.