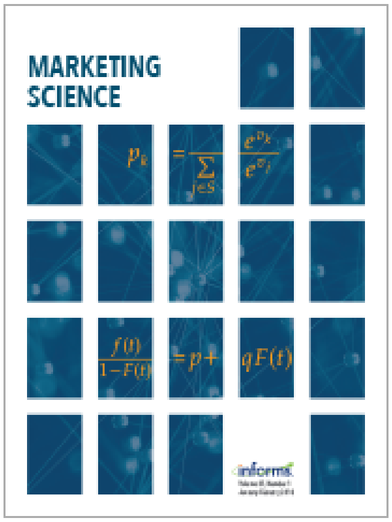


Marketing Science

Two-Sided Price Discrimination by Media Platforms

Song Lin

http://pubsonline.informs.org/journal/mksc

MARKETING SCIENCE
Vol. 39, No. 2, March–April 2020, pp. 317–338
ISSN 0732-2399 (print), ISSN 1526-548X (online)

Two-Sided Price Discrimination by Media Platforms

Song Lin[a]

[a] Department of Marketing, Hong Kong University of Science and Technology, Hong Kong
Contact: mksonglin@ust.hk, https://orcid.org/0000-0003-0393-010X (SL)

Received: July 24, 2017
Revised: July 24, 2018; January 28, 2019; May 18, 2019; August 20, 2019
Accepted: August 20, 2019
Published Online in Articles in Advance: March 3, 2020

https://doi.org/10.1287/mksc.2019.1211

Abstract. An increasingly common practice among media platforms is to provide premium content versions with fewer or even no ads. This practice leads to an intriguing question: how should ad-financed media price discriminate through versioning? I develop a two-sided media model and illustrate that price discrimination on one side can strengthen the incentive to discriminate on the other. Under this self-reinforcing mechanism, the ad allocations across different consumer types depend crucially on how much nuisance of an ad “costs” consumers relative to the value it brings to them. Interestingly, higher-type consumers, who value content and advertising quality highly, may see more ads than lower-type consumers if the nuisance cost is relatively low. Furthermore, the standard downward quality distortion generally fails to materialize in a two-sided market and may even be reversed: higher-type consumers may be exposed to too few ads that result in a lower total quality than the socially efficient level, whereas lower-type consumers may receive a socially excessive quality. The circumstances under which the self-reinforcing mechanism may be weakened and the implications for media platform design are explored.

History: Ganesh Iyer served as the senior editor and Dmitri Kuksov served as associate editor for this article.

Keywords: price discrimination • two-sided market • platform design • media pricing • advertising • ad avoidance • targeting

1. Introduction

In recent years, many digital and media platforms have introduced premium services that allow paid subscribers to see fewer or even no advertisements (ads). One notable example is the leading video-sharing platform YouTube, which offers free video streaming to users and profits from selling ad space. In October 2015, the platform introduced YouTube Red, which enables users to pay a monthly subscription fee to enjoy ad-free content, in addition to offering other premium features, such as offline video storage and original shows. Many other media platforms also offer similar subscription services, including CBS Broadcasting, Spotify, Hulu, Pandora, and the major Chinese video streaming platforms iQiyi, Youku, and Tencent Video.

At first glance, this practice seems to be a classic type of second-degree price discrimination through versioning, which enables consumers to self-select from different products: consumers who value media content highly or are averse to ads can choose to pay for the premium version with fewer or no ads, whereas others may choose to stay with the basic version that contains more ads. However, offering a premium service may reduce the profit obtained from selling consumer attention (i.e., “eyeballs”) to advertisers, which represents a standard trade-off for media platforms. If too many consumers choose the premium version, the ad revenue loss can be quite significant.[1] Furthermore, this practice may appear puzzling when juxtaposed with the phenomenon that many media publishers, such as the *Wall Street Journal* and the *New York Times*, have introduced, a paywall that involves consumers paying for premium (or simply added) content associated with *more* ads.

These phenomena call for further investigation of how media platforms should price discriminate in two-sided markets. To address this question, I cast the problem in a framework of second-degree price discrimination within a two-sided market. Unlike typical sellers, media or content providers not only sell directly to consumers but also act as two-sided platforms. By attracting consumers with free or low-priced content, they can sell consumer attention to advertisers. Because of cross-externalities between advertisers and consumers, media platforms must design pricing policies that balance incentives on both sides. This study focuses on the scenario of a media platform with monopoly power over a heterogeneous consumer market.[2] The platform offers entertainment or information content of value to two types of consumers, who differ in terms of the marginal utility of content quality (high versus low types). It can charge a lump-sum subscription fee to consumers and sell ad space to advertisers, who differ in their probability of matching consumer needs. Consumers will be induced to buy a product only if they are aware of the advertiser *and* its product matches their needs. Although advertising has informational value, consumers may dislike ads because they interrupt their content consumption.[3] Information asymmetry exists on both sides because consumers and advertisers privately know their own types, whereas the platform possesses no type information. The platform must then rely on self-selection to screen both consumers and advertisers.

In Section 3, I present a baseline analysis with the standard assumption of two-sided market analysis: agents on one side care only about the number of agents they interact with on the other side. Although content quality is fixed, price discrimination can still be implemented by the differential allocation of ads and by exploiting the negative externality of advertising on consumers. When consumers select different ad allocations, a platform can, in effect, sell three ad products to prospective advertisers: one to reach high-type consumers only, another to reach low-type consumers only, and a third to reach both. This menu of options serves as an instrument for screening advertisers, which, in turn, forms the basis for consumer-side discrimination. This self-reinforcing mechanism, in which price discrimination on one side strengthens discrimination on the other side, is a novel feature in a two-sided market. It is shown that the platform can adopt a simple advertising policy that always allocates all participating advertisers to low-type consumers and allocates fewer yet higher-matching-quality advertisers to high-type consumers. In equilibrium, the number of ads high-type consumers see depends on the degree of marginal utility, the value of advertised products, and the nuisance cost of advertising. A completely ad-free version for high-type consumers represents a corner solution when either the marginal utility or the nuisance cost is too high or the product value is too low.

On the basis of these insights, in Section 4, I explore what happens to media platforms' policies when consumers care about both the quantity and *quality* of advertisers with whom they interact. This scenario may arise if, for example, consumers prefer an advertiser that sells a better-matched product, regardless of the nuisance cost of advertising. One major implication is that the advertising externality on consumers may change with the number of ads in a nonmonotonic fashion. If the nuisance cost is not very high, the advertising externality on consumers first increases and then decreases as they see more ads. Taking into account the value of trading with an advertiser, advertising externality may even be positive if the nuisance cost is relatively low and there are relatively few ads, contrasting with the general belief that advertising on media platforms mainly exerts a negative externality on consumers.

Two notable findings emerge here. First, a media platform can still implement price discrimination through versioning; however, the consumer type that receives more ads essentially depends on how much the nuisance of an ad costs consumers relative to how much product value it brings to them. If the nuisance cost is relatively high, then the optimal versioning policy is the same as in the baseline analysis, with fewer ads shown to high-type consumers than to low-type consumers. In contrast, if the nuisance cost is relatively low, then high-type consumers may see *more* ads than low-type consumers. This mechanism may explain why some reputable newspapers, such as the *Wall Street Journal* and the *New York Times*, have introduced paid versions with unlimited content access that come with more ads. A probable reason is that newspapers tend to have a lower nuisance cost than video or music streaming services because readers can easily skip ads. Second, because a media platform must optimally balance subscription and advertising revenues, the ad allocation to each consumer type generally differs from the socially efficient level. It is possible for low-type consumers to receive fewer ads than the socially optimal level, resulting in the overprovision of total quality (i.e., the sum of content quality and advertising externality). Contrarily, high-type consumers might be exposed to too few ads, resulting in the underprovision of total quality, when they actually value more ads. Hence, the standard welfare result found in one-sided markets that quality is distorted downward for lower-type consumers does not always hold in a two-sided market and may even be reversed.

Next, in Section 5, I explore market features that can disrupt the self-reinforcing mechanism in a two-sided market. The first feature is the dependency of advertiser preferences on consumer type. In many markets, advertisers can obtain a higher profit by converting high-type consumers, for at least two reasons. First, many media platforms show advertised products that are related to the media content. For example, sports channels often show ads for sports products. Second, higher-type consumers are often less price sensitive and thus willing to pay more for both content and advertised products. Under this dependency assumption, advertisers prefer to reach high-type more than low-type consumers. A platform that successfully segments consumers based on self-selection can provide an opportunity for advertisers to target a specific segment. Although an advertiser may still choose to reach multiple consumer segments, targeted advertising can further enable it to charge different prices for different segments and thus obtain higher profits. As the profit advertisers gain from high-type consumers increases, a media platform tends to expose these consumers to more ads to extract more revenue from advertisers. The difference in ad allocation between high-type and low-type consumers then becomes smaller and eventually vanishes. Consequently, the platform cannot effectively discriminate among consumers, rendering targeted advertising and discrimination on the advertiser side infeasible.

Another market feature that can cause price discrimination to fail is informational friction between consumers and advertisers. Consumers often ignore or pay no attention to ads, and they only become informed about advertisers when they receive a sufficient amount of ad messages. Advertisers must then decide how many ads to place in the market to compete for consumers' attention. The more ads they place, the more likely it is that consumers will become informed. The incentive to place more ads is stronger for higher-type advertisers. A larger segment of high-type consumers provides advertisers with a stronger incentive to expose them to more ads. Consequently, a media platform may allocate more ads in the premium version for high-type consumers, reducing the ad intensity gap between the two consumer types. The ability to discriminate on the consumer side is then weakened, resulting in a failure to implement versioning when the market predominantly consists of high-type consumers.

Section 6 extends the baseline analysis to a broader environment in which a media platform can also vary the content quality of different versions and demonstrates how the main findings inform product-line management. Introducing an additional discriminatory tool based on content differentiation can increase subscription revenue from high-type consumers because of the high fee they are charged while also reducing the revenue from low-type consumers because of the lower price of the basic version. Thus, as long as the high-type segment is not too small, the gain outweighs the loss, and a versioning strategy with both ad and content differentiation is optimal. Furthermore, the analysis shows that content differentiation can, in turn, facilitate the implementation of ad differentiation. When price discrimination based purely on ad differentiation becomes infeasible, as demonstrated in Section 5, content differentiation can restore the feasibility of versioning. This can constitute an alternative mechanism by which versioning may enable high-type consumers to see more ads.

1.1. Related Literature

This study contributes to the burgeoning literature on two-sided media market analysis. Media platforms are widely regarded as exemplifying two-sided markets (e.g., Armstrong 2006, Rochet and Tirole 2006, Weyl 2010). They sell information or entertainment content directly to consumers and thus attract advertisers who aim to reach these consumers. The central problem is how media can balance these two sides, given the cross-externalities between advertisers and consumers, and what the implications are for content and advertising provision. Anderson and Coate (2005) examine the market provision of content and advertising and compare it with socially optimal provision. Dukes (2004) explores the equilibrium level of advertising when advertisers in a differentiated product market compete in prices and advertising and the effect of product differentiation on the media market. Godes et al. (2009) examine the impact of media competition on content pricing and find that media firms may charge higher prices for content in a duopoly than in a monopoly, in contrast with a conventional one-sided product market. Athey et al. (2016) explore how consumers' multihoming influences advertising markets and media competition and how the increasing propensity of consumers to switch affects market outcomes. A number of studies in economics and marketing have also investigated media pricing and advertising provision in media markets (e.g., Gal-Or and Dukes 2003, Peitz and Valletti 2008, Kind et al. 2009, Reisinger 2012, Ambrus et al. 2016). Most of these studies assume that media offer the same content and ads to all consumers, without versioning. In contrast, the increasingly popular practice of discriminating among consumers through versioning is examined in this study, raising the question of differential ad provision across different versions.

Prasad et al. (2003) conducted the first formal study of this phenomenon, formulating a model similar to that in my baseline analysis. A key difference is that, in my model, a media platform can design different ad products from which advertisers choose. Price discrimination among advertisers must be incentive compatible, and the media platform selects the optimal selling format of the ad space. Tåg (2009) studies the same phenomenon and analyzes media incentives and welfare implications, assuming that the advertising-based version is free and that the premium version contains no ads. Both assumptions are relaxed in my analysis. Unlike these papers, I examine situations in which consumers may have rational expectations about ex post benefits from interacting with advertisers in the product market, and advertisers may have heterogeneous preferences regarding different types of consumers. I also explore the opposing forces that can disrupt price discrimination.

The idea of granting consumers an option to remove ads is closely related to the problem of ad avoidance, which has attracted considerable research attention. Because of the nuisance cost of advertising, consumers use various means to avoid ads.

For example, they can turn the pages of a newspaper, switch television channels, or simply take a bathroom break. The advent of ad-avoidance technologies, such as the TiVo digital video recorder and many ad-blocking mobile applications, has enabled consumers to skip commercials or banner ads completely, thus threatening media firms financed by advertising. Anderson and Gans (2011) show that the greater penetration of ad-avoidance technologies may cause media to increase advertising clutter because marginal ad viewers are less sensitive to ads. They also argue that these technologies may reduce social welfare and content quality and lead to more mass-market content.[4] Johnson (2013) further examines how firms' increasing ability to target ads affects market outcomes when consumers use ad-avoidance tools. The present study suggests that media platforms can approach the problem with an alternative strategy: instead of banning consumers from using ad-avoidance technologies, they can offer them a choice between a less costly version with more ads and a more costly version with fewer ads. Such a strategy can mitigate the profit damage from ad-avoidance technologies.[5]

This study also contributes to the extensive literature on price discrimination, which typically assumes a one-sided market and focuses on how sellers can discriminate effectively among heterogeneous buyers. One question is how firms can effectively screen consumers who have private information about their own preferences (Mussa and Rosen 1978, Maskin and Riley 1984). However, in a two-sided market, price discrimination becomes more complex. Media must take into account heterogeneity on both sides and evaluate how the segmentation of one side affects the other. Limited research has been conducted in this domain. Although Liu and Serfes (2013) examine the perfect price discrimination of horizontally differentiated two-sided platforms, their study differs from the present analysis in that it neglects the interaction between the discrimination on both sides of the market; in my model, discrimination on one side can lead to further discrimination on the other side. The mechanisms in the studies conducted by Gomes and Pavan (2016) and Jeon et al. (2016) are closer to that presented here. Gomes and Pavan (2016) examine optimal price schedules when agents on either side are matched to each other. Applying the optimal matching rules to media markets results in top advertisers being matched to all prospective consumers, whereas lower-end advertisers are matched only to lower-type consumers, which is consistent with my baseline analysis. Jeon et al. (2016) similarly analyze second-degree price discrimination by a two-sided platform but focus on optimal quality–price schedules and whether price discrimination on one side substitutes for or complements the discrimination on the other side. When applying their model to media markets, they assume that ads are completely removed from the premium version, whereas I treat ad intensity in the premium version as endogenous.[6] Thus, in their model, advertisers can only advertise in one version, whereas my model allows advertising in multiple versions to reach a larger market. Unlike both of these studies, I consider the possibility of positive advertising externality on (some) consumers, with the magnitude of externality depending on product market characteristics.

The remainder of this paper is structured as follows. Section 2 introduces the model. Section 3 presents an analysis of the benchmark case when advertising externality is type independent. In Section 4, I examine a scenario in which advertising externality depends on consumer type. In Section 5, I explore two important market features that can weaken the incentive for price discrimination. In Section 6, the managerial implications are discussed, and Section 7 concludes with suggestions for future exploration. Except for the key derivations, the proofs of all the lemmas and propositions are documented in the appendix.

2. The Model

Consider a monopoly—a media outlet, content provider, or publisher—that offers certain information or entertainment content to two segments of consumers, denoted by h and l. The media content has a constant quality $q > 0$, with the marginal cost of production normalized to zero. Consumers differ in their marginal utility, or taste, for the quality captured by a parameter $\theta > 0$.[7] Higher-type consumers have a higher taste such that $\theta_h > \theta_l$. The ratio $\alpha \equiv \theta_h/\theta_l$ denotes the degree of value heterogeneity. The fraction of high-type consumers is $\beta \in (0, 1)$, and low-type consumers have a share of $1 - \beta$. Consumers privately know their own types, and thus the media platform must rely on self-selection to screen consumers.

On the other side of the media market, a continuum of advertisers with mass normalized to one is interested in advertising to consumers who use the platform. Each advertiser is characterized by a matching probability $\sigma \in [0, 1]$, which is drawn from a uniform distribution and captures the quality of the advertisers in terms of providing goods or services of interest to consumers. Like consumers, advertisers privately know their own types. The platform must also rely on self-selection to segment advertisers. Next, I introduce further details of how the platform, consumers, and advertisers interact.

2.1. The Content Market

Although the quality of the media content is fixed, the platform can create differentiated products simply by varying ad allocation. For each product version, the platform can allocate a subset of the advertisers $\Sigma \subseteq [0,1]$. For example, a simple form of ad allocation is the interval assignment $\Sigma = [\sigma_0, 1]$, under which consumers are informed of all advertisers with type $\sigma > \sigma_0$. To implement versioning, the platform can determine the allocations Σ_h and Σ_l for the premium and basic versions by setting the corresponding subscription prices p^c_h and p^c_l, expecting high- and low-type consumers to self-select into the corresponding version.

Seeing an ad is assumed to be a distraction to consumers, who are mainly interested in the media content. This assumption is in accordance with the casual observations of consumers of many media outlets, including television, newspaper, and online platforms, and has gained much empirical support (Wilbur 2008). To model this assumption, each ad is assumed to cause a nuisance cost $\lambda > 0$. This uniform nuisance cost implies that consumers have the same disutility for any ad regardless of its matching quality.[8] Under an ad allocation Σ, the total number of ads that a consumer sees is denoted by $a = |\Sigma|$ (i.e., the Lebesgue measure of Σ). Then a type θ consumer obtains the utility of $\theta(q - \lambda a)$ from consuming the media content, taking into account the disutility of seeing ads. Throughout the analysis, the content quality is assumed to be sufficiently high, $q > \lambda$, to ensure that the net benefit of media consumption is always positive, even under maximal ad provision.

2.2. The Product Market

Following the literature on media markets (e.g., Dukes 2004, Anderson and Coate 2005), in this paper I consider the role of advertising as being to inform consumers of the existence of a product or a brand.[9] Upon learning about the existence of an advertiser, a consumer may purchase a product from the advertiser if there is a match. The conversion or transaction produces a surplus $v_c \geq 0$ to the consumer and a profit $v_a \geq 0$ to the advertiser. A few remarks follow. First, although all advertisers obtain the same profit per transaction, they are heterogeneous in terms of the likelihood that their products will match a consumer's needs. That is, the heterogeneity is driven by the matching probability σ. Second, I begin with the simplest assumption that advertiser profit in the product market v_a is independent of consumer type. In Section 5.1, I explore the possibility that v_a may depend on consumer type to gain more insight. Third, I assume that repeated exposure has no benefit and thus that ad intensity becomes irrelevant. In Section 5.2, I examine what happens if this assumption is relaxed.

2.3 The Advertising Market

To sell ad space, the media platform offers a menu of take-it-or-leave-it contracts to advertisers. Each contract specifies the segment(s) of consumers to which an ad is delivered and the corresponding rate. If the platform successfully segments consumers, there are three possible ad products: one reaches high-type consumers alone at rate p^a_h, another reaches low-type consumers alone at rate p^a_l, and one reaches both types at rate p^a_{hl}. An advertiser of type σ can obtain profits $\sigma\beta v_a - p^a_h$, $\sigma(1-\beta)v_a - p^a_l$, or $\sigma v_a - p^a_{hl}$ if it chooses to advertise to the high-type consumer segment, the low-type consumer segment, or both segments, respectively. Note that the decisions of advertisers have to be consistent with the ad-allocation plan (Σ_h, Σ_l) that the platform aims to implement. For example, if a type σ advertiser chooses to reach the low-type consumer segment alone, then $\sigma \in \Sigma_l$ but $\sigma \notin \Sigma_h$. If the advertiser chooses to reach both consumer segments, then $\sigma \in \Sigma_h$ and $\sigma \in \Sigma_l$.

2.4. Timing

To complete the model setup, let us summarize the timing of the game:

1. The media platform announces the subscription prices of the two versions (p^c_h, p^c_l) and the rates of the three ad products (p^a_{hl}, p^a_h, p^a_l).
2. Both consumers and advertisers observe all the prices.[10] Consumers choose which version to adopt, and advertisers choose which ad product to purchase.
3. If a consumer sees an ad, he or she becomes informed about the advertiser. The consumer is converted according to the advertiser's matching probability.
4. If converted, the consumer obtains a surplus v_c, and the advertiser obtains a profit v_a in the product market. The total payoff of a type θ consumer under ad allocation Σ and price p^c is then

$$U(\Sigma, p^c; \theta) = \theta\left(q - \lambda|\Sigma| + \int_{\sigma\in\Sigma} \sigma v_c d\sigma\right) - p^c. \quad (1)$$

Let $u_e(\Sigma)$ denote the utility component stemming from the advertising externality under an ad allocation Σ: $u_e(\Sigma) \equiv -\lambda|\Sigma| + \int_{\sigma\in\Sigma} \sigma v_c d\sigma$. In later analysis, ad allocations are often in the interval form $\Sigma = [\sigma_0, 1]$, and thus the ad volume $a = 1 - \sigma_0$ can conveniently represent an ad allocation. Hence, with a slight abuse of notation, $u_e(a)$ is used to denote the advertising externality. Equation (1) reflects that both the fixed quality of the media content and the externality as a result of advertising affect consumer utility. We can treat the sum of the two, $Q = q + u_e(\Sigma)$, as the measure of the total quality of the allocations (i.e., both media content and ad allocations) by the media platform. Table 1 summarizes the notations introduced so far.

Table 1. Summary of Notations

Parameter	Meaning
q	The fixed quality of the media content
θ_h, θ_l	The marginal utility of quality for high-type and low-type consumers
$\alpha = \theta_h/\theta_l$	The value heterogeneity of consumers
β	The fraction of high-type consumers
σ	The match probability (type) of an advertiser
λ	The nuisance cost of an ad to consumers
v_a	The advertiser profit of converting a consumer in the product market
v_c	The consumer surplus of buying from an advertiser in the product market
Choice variables	**Meaning**
p^c_h, p^c_l	The subscription fees for the premium (targeted high-type consumers) and basic (targeted low-type consumers) versions
Σ_h, Σ_l	The sets of allocated advertisers in the premium and basic versions
p^a_{hl}, p^a_h, p^a_l	The ad rates for reaching all consumers, premium version consumers, and basic version consumers
Other variables	**Meaning**
a	The total number (volume) of allocated advertisers: $a = \lvert\Sigma\rvert$
$u_e(\Sigma)$	The advertising externality on a consumer: $u_e(\Sigma) \equiv -\lambda\lvert\Sigma\rvert + \int_{\sigma\in\Sigma} \sigma v_c d\sigma$
Q	The total quality of the media allocation: $Q = q + u_e(\Sigma)$
$U(\Sigma, p^c; \theta)$	The total payoff of a type-θ consumer: $U(\Sigma, p^c; \theta) = \theta(q + u_e(\Sigma)) - p^c$

2.5. Efficient Benchmark

It will be instructive to present the efficient benchmark before the main equilibrium analysis. Consider a social planner that maximizes the joint surplus of consumers, advertisers, and the media firm. The following result shows that the (interior) efficient allocation takes the form of versioning.

Lemma 1. *The efficient allocation always entails versioning: advertisers with type $\sigma > 1 - \hat{a}_l$ reach low-type consumers, and advertisers with type $\sigma > 1 - \hat{a}_h$ reach high-type consumers. The efficient ad quantities $(\hat{a}_h, \hat{a}_l)$ are*

$$\hat{a}_h = \frac{\theta_h(v_c - \lambda) + v_a}{\theta_h v_c + v_a} \quad \text{and}$$
$$\hat{a}_l = \frac{\theta_l(v_c - \lambda) + v_a}{\theta_l v_c + v_a} \tag{2}$$

and satisfy $\hat{a}_l > \hat{a}_h > \frac{v_c - \lambda}{v_c}$.

The efficient allocation is obtained from the welfare-maximization problem:

$$\max_{a_l, a_h} \quad \beta\theta_h\big(q + u_e(a_h)\big) + \big(1-\beta\big)\theta_l\big(q + u_e(a_l)\big)$$
$$+ \beta\int_{1-a_h}^{1} \sigma v_a d\sigma + \big(1-\beta\big)\int_{1-a_l}^{1} \sigma v_a d\sigma. \tag{3}$$

Note that there is no interaction between the high- and low-type markets. The efficient allocation is then solved separately for each market using the first-order conditions. Next, I analyze how the media platform should implement a versioning policy to maximize its profit.

3. Versioning Under Type-Independent Externality

Let us first examine the elementary setting in which the externality between consumers and advertisers is independent of their types. Agents on each side care only about how many agents on the other side they will interact with, a standard assumption in the literature on two-sided markets (e.g., Armstrong 2006, Rochet and Tirole 2006). This assumption can be captured by letting $v_c = 0$. Thus, the total payoff of a consumer in Equation (1) reduces to $U(\Sigma, p^c; \theta) = \theta(q - \lambda|\Sigma|) - p^c$.

The analysis has two objectives. First, it illustrates the main mechanism by which discrimination on one side can facilitate discrimination on the other. The approach used to derive the equilibrium is introduced here and later applied in more complex settings. Second, it focuses on the scenario in which the advertising externality on consumers is always negative because only the ad nuisance cost matters. This situation can arise when consumers realize little surplus from trading with advertisers because of their strong market power or bargaining power.[11] This scenario also captures the idea that consumers may be myopic when determining to join the platform, neglecting to consider the ex post benefits of trading with advertisers in the product market.

3.1. Implementing Versioning

As mentioned earlier, discrimination among consumers based on ad allocation essentially generates three ad products in the advertising market: advertising to low-type consumers at rate p^a_l, to high-type consumers at rate p^a_h, and to both at rate p^a_{hl}. Because of the assumption of independent preferences, consumers are homogeneous from advertisers' points of view. Advertisers then only need to determine the reach of their advertising. They can expect a profit of $(1-\beta)\sigma v_a$ if advertising to low-type consumers, $\beta\sigma v_a$ if advertising to high-type consumers, and σv_a if advertising to both. The single-crossing property of advertiser preference implies that any incentive-compatible strategy should induce higher-type advertisers to choose ad products with greater reach. The following proposition summarizes the optimal implementation, which has a simple structure.

Proposition 1. *Under type-independent externality with $v_c = 0$, the media platform maximizes profit by implementing versioning so that* (a) *low-type consumers receive ads from all participating advertisers;* (b) *high-type consumers receive fewer ads with higher matching probabilities; and* (c) *if and only if $\theta_h \lambda \geq v_a$, high-type consumers receive no ads.*

Next, I derive the solution to the proposed implementation. Figure 1 illustrates the segmentation when $\theta_h \lambda < v_a$. The advertiser who is indifferent between reaching both consumer segments and reaching only low-type consumers is identified by $\sigma_h = (p^a_{hl} - p^a_l)/\beta v_a$. The marginal advertiser who is indifferent between advertising to low-type consumers only and not advertising is identified by $\sigma_l = p^a_l/(1-\beta)v_a$.

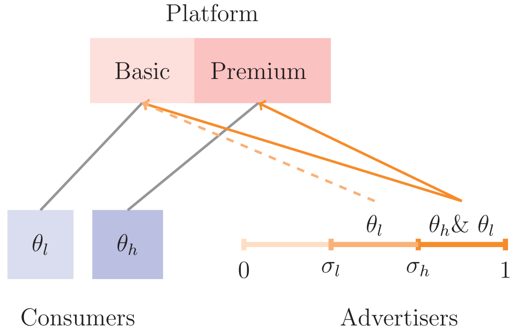
The ad allocations can be summarized by $a_h = 1 - \sigma_h$ and $a_l = 1 - \sigma_l$. Inducing the ad demands (a_h, a_l) can be achieved by setting the ad rates $p^a_l = (1-a_l)(1-\beta)v_a$ and $p^a_{hl} = p^a_l + (1-a_h)\beta v_a$. The profit-maximization problem for the platform is

$$\max_{a_l, a_h, p^c_l, p^c_h} \quad \pi = \underbrace{\beta p^c_h + (1-\beta) p^c_l}_{\text{subscription revenue}} + \underbrace{a_h(1-a_h)\beta v_a + a_l(1-a_l)(1-\beta)v_a}_{\text{advertising revenue}}, \quad (4)$$

subject to the individual rationality (*IR*), incentive compatibility (*IC*), and implementability constraints:

$$\begin{aligned}
&IR_l : \theta_l(q - \lambda a_l) - p^c_l \geq 0;\\
&IR_h : \theta_h(q - \lambda a_h) - p^c_h \geq 0;\\
&IC_l : \theta_l(q - \lambda a_l) - p^c_l \geq \theta_l(q - \lambda a_h) - p^c_h;\\
&IC_h : \theta_h(q - \lambda a_h) - p^c_h \geq \theta_h(q - \lambda a_l) - p^c_l; \text{ and}\\
&\text{Implementability} : a_l \geq a_h.
\end{aligned}$$

Figure 1. (Color online) Implementation of Versioning (When $\theta_h \lambda < v_a$)



Ignoring the IR_h, IC_l, and implementability constraints momentarily, we can use the binding IR_l and IC_h constraints to set the optimal prices $p^c_l = \theta_l(q - \lambda a_l)$ and $p^c_h = p^c_l + \theta_h \lambda(a_l - a_h)$. Substituting these price schedules back to the profit function, we can then solve for the optimal (interior) ad allocations (a^*_h, a^*_l):

$$\begin{aligned}
a^*_h &= \frac{1}{2} - \frac{\theta_h \lambda}{2v_a},\\
a^*_l &= \frac{1}{2} - \frac{(\theta_l - \beta\theta_h)\lambda}{2(1-\beta)v_a}.
\end{aligned} \quad (5)$$

The condition $\theta_h \lambda < v_a$ ensures that $a^*_h > 0$. If $\theta_h \lambda \geq v_a$, then the corner solution $a^*_h = 0$ is optimal, implying that high-type consumers do not receive any ads. To complete the derivation, we shall verify that the implementability constraint is satisfied by the proposed solution. Indeed, $a^*_h < a^*_l$ holds for any parameter value as long as $\theta_h > \theta_l$. In the appendix, I further validate the existence and uniqueness of the proposed implementation.

Note that equilibrium multiplicity might arise in models with network effects. However, this is not an issue here because, as Anderson and Coate (2005) point out, consumers are attracted to the media platform not by the existence of the advertisers but by the media content. More generally, the prices set for different versions (p^c_h, p^c_l) and the ad rates set for different allocations (p^a_{hl}, p^a_h, p^a_l) are what Weyl (2010) calls *insulating tariff*. He argues that if the platform charges the insulating tariff associated with its desired allocation on both sides, then the unique equilibrium is its desired equilibrium.

3.2. Discussion

Having solved the optimal ad allocations (a^*_h, a^*_l), we can substitute them back into the price schedules to obtain the optimal subscription prices and ad rates. The difference between the subscription prices of the two versions is therefore as follows:

$$p^{c*}_h - p^{c*}_l = \theta_h \lambda (a^*_l - a^*_h) = \frac{\theta_h(\theta_h - \theta_l)\lambda^2}{2(1-\beta)v_a}. \quad (6)$$

Intuitively, if consumers only treat ads as a nuisance, increasing λ can enlarge the quality difference between the two versions, leading to a larger price gap. Note that the price gap also decreases with v_a, advertisers' profit in the product market (or consumers' externality on advertisers). A higher v_a enhances advertisers' incentive to reach both consumer segments, resulting in a smaller segment of advertisers reaching only low-type consumers. That is, $\sigma_h - \sigma_l$ becomes smaller, all else equal. Hence, the difference in ad volumes between the two versions is reduced, leading to a smaller price gap.

Turning to the optimal ad rates, we have

$$p_{hl}^{a*} - p_l^{a*} = (1 - a_h^*)\beta v_a = \frac{\beta}{2}(v_a + \theta_h \lambda). \quad (7)$$

The difference in ad rates evidently increases with v_a because advertisers are more willing to reach a larger product market if they expect a higher return from it. Note that the difference in ad rates also increases with the nuisance cost λ, which captures the degree of negative externality of an ad on consumers. The platform must limit the ad exposure of high-type consumers if they become more annoyed by the interruption of ads, driving up the ad rate for reaching these consumers.

An important distinguishing feature of two-sided price discrimination is the self-reinforcing mechanism that price discrimination on one side can facilitate discrimination on the other. By segmenting consumers, the platform is able to charge advertisers differently by offering different ad products. This differentiation, in turn, provides the basis for discriminating among consumers with different versions. In contrast, under the current model setup, simple one-sided price discrimination is either infeasible or suboptimal.

Corollary 1. *One-sided price discrimination such that only one side (either consumers or advertisers) of the platform is discriminated is not strictly optimal.*

We should be cautious about interpreting the foregoing result. It does not imply that one-sided price discrimination is generally infeasible. The model has ruled out further instruments for the platform to practice price discrimination by assumption. For example, the platform cannot discriminate among advertisers based on the quantity of consumers without segmenting consumers in the first place. It can, however, still screen consumers without effectively segmenting advertisers through random ad allocation: each advertiser is randomly allocated to high-type consumers with probability $\omega \in [0,1]$ and to low-type consumers with probability $1 - \omega$. The platform can implement discrimination by imposing the constraint $\omega < 1/2$ so that low-type consumers see more ads than high-type consumers. As shown in the proof of Corollary 1, the optimal strategy is to set $\omega = 0$. Such a strategy, however, is generally suboptimal. Intuitively, the platform restricts itself to only one ad price in the advertising market, whereas in the two-sided discrimination approach, it has more flexibility to improve profit using more prices. Compared with two-sided discrimination, this one-sided discrimination strategy delivers too few ads to high-type consumers (who see no ads at all).

4. Versioning Under Type-Dependent Externality

The baseline analysis of type-independent externality illustrates that media platforms can implement versioning simply based on heterogeneous preference over the quantity of agents on the other side. The natural question to ask is what would happen if consumers cared about the quality or type of advertisers with whom they interact. There are two possible scenarios. First, the nuisance cost of seeing an ad might depend on the advertiser type. This scenario captures the idea that a better-matched advertiser is less annoying to a consumer. Under this assumption, the conclusions are qualitatively the same as those in the baseline analysis, primarily because the total advertising externality on consumers remains negative. The second and more interesting scenario is if consumers expect gains from interacting with different types of advertisers, in addition to the nuisance cost. For example, sports fans may like to see an ad featuring a new sports shoe with the latest technology. Under this scenario, the net externality of advertising on consumers may in fact be positive. To capture this idea, let us assume that $v_c > 0$ in this section.

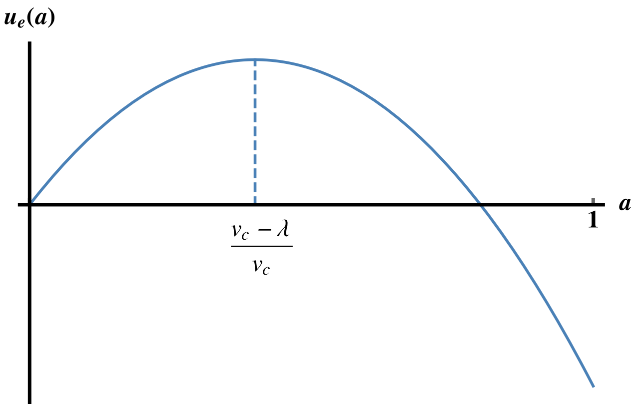
4.1. Implementing Versioning

The utility of a type θ consumer now depends not only on the number of advertisers with whom they interact but also the advertiser type. Suppose that the platform allocates the advertisers within $[\sigma_0, 1]$ to the consumer, and let $a = 1 - \sigma_0$. The total payoff the consumer expects from subscribing to the media platform then becomes

$$\begin{aligned} U(a, p^c; \theta) &= \theta\left(q - \lambda a + \int_{\sigma_0}^{1} \sigma v_c d\sigma\right) - p^c \\ &= \theta\left(\underbrace{q - \lambda a + \frac{1}{2} v_c (2 - a) a}_{u_e(a)}\right) - p^c. \end{aligned} \quad (8)$$

Equation (8) reveals a nonmonotonic relationship between the advertising externality on consumers u_e and the number of advertisers a. This pattern is novel in the literature on two-sided markets, which often assumes that cross-group externalities are linear in the number of participating agents. Figure 2 illustrates the pattern when $\lambda < v_c$. As a increases, $u_e(a)$ is first increasing and positive. It reaches a maximum when $a = (v_c - \lambda)/v_c$, after which $u_e(a)$ decreases with a and could eventually become negative if $\lambda > v_c/2$. This pattern contrasts sharply with the monotonically decreasing pattern of $u_e(a)$ under type-independent externality. Advertising can now bring a positive net gain to consumers.

Figure 2. (Color online) Illustration of Advertising Externality $u_e(a)$ (When $\lambda < v_c$)



Lemma 2. (a) *If $\lambda < v_c$, then as a increases, the advertising externality on consumers $u_e(a)$ first increases and then decreases, with the maximum reached at $a = (v_c - \lambda)/v_c$.* (b) *If $\lambda \geq v_c$, then the externality is always negative and decreasing.*

Following the baseline analysis, the media platform maximizes profit by choosing ad allocations and subscription fees as in Equation (4). The set of incentive constraints is the same except that consumer utility is replaced by $U(a, p^c; \theta)$ in Equation (8). The implementability condition now becomes $u_e(a_h) > u_e(a_l)$. Using the same approach as in the baseline analysis, we can obtain a candidate for the interior solution:

$$a_h^* = \frac{\theta_h(v_c - \lambda) + v_a}{\theta_h v_c + 2v_a},$$
$$a_l^* = \frac{(\theta_l - \beta\theta_h)(v_c - \lambda) + (1 - \beta)v_a}{(\theta_l - \beta\theta_h)v_c + 2(1 - \beta)v_a}. \tag{9}$$

By evaluating whether the solution satisfies the implementability condition, we can establish the following results regarding the implementation of the versioning policy.

Proposition 2. *Under type-dependent externality with $v_c > 0$, the platform maximizes profit by implementing versioning through the ad allocations specified in Equation (9). In addition,*

1. *If the ad nuisance cost is relatively high, $\lambda > \frac{1}{2}v_c$, then high-type consumers see fewer ads than low-type consumers, $\frac{v_c - \lambda}{v_c} < a_h^* < a_l^*$; if it is relatively low, $\lambda < \frac{1}{2}v_c$, then high-type consumers see more ads than low-type consumers, $a_l^* < a_h^* < \frac{v_c - \lambda}{v_c}$.*
2. *Suppose that high-type consumers dominate, $\alpha\beta > 1$ and $\frac{(\alpha\beta - 1)\theta_l}{2(1-\beta)} > \frac{v_a}{v_c}$. Then low-type consumers either receive all the ads, $a_l^* = 1$ if $\lambda > \frac{1}{2}v_c$, or no ads, $a_l^* = 0$ if $\lambda < \frac{1}{2}v_c$.*
3. *If $\lambda > \frac{1}{2}v_c$, then the amount of ads low-type consumers receive, a_l^*, weakly increases with the fraction of high-type consumers β. If $\lambda < \frac{1}{2}v_c$, then a_l^* weakly decreases with β. Furthermore, the advertising externality on low-type consumers, $u_e(a_l^*)$, always weakly decreases with β.*

The first part of this proposition delivers an important message: when consumers expect the informational benefit of advertising, the optimal advertising policy crucially depends on the magnitude of the ad nuisance cost relative to the informational value that ads bring to consumers. To understand this result, let us suppose that the platform initially does not differentiate offerings to consumers and sets the ad level at $a^m \equiv (v_c - \lambda)/v_c$ to maximize subscription revenue (i.e., maximize the advertising externality). When the ad nuisance cost is relatively high, consumers are more price elastic, and a^m is smaller, implying a high ad rate charged to advertisers. Now consider what happens if the platform introduces versioning to segment consumers. It can improve ad revenue by expanding the advertising market while sacrificing some subscription revenue. That is, $a_h > a^m$ and $a_l > a^m$. To screen consumers, the ad volume should be larger for low-type than for high-type consumers, $a_l > a_h$, because advertising externality decreases with ad volume when $a > a^m$. Hence, we generalize the results for type-independent externality analyzed in Section 3.

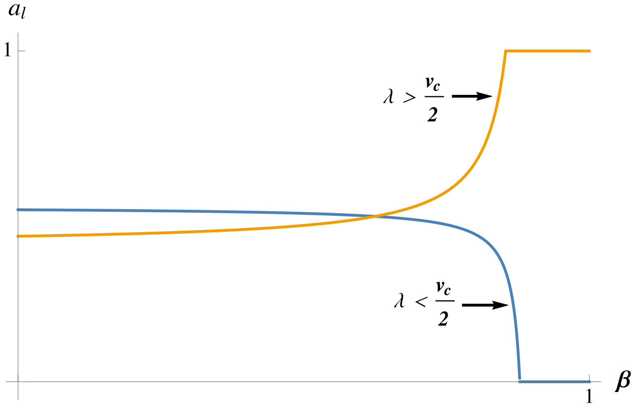
The conclusion, however, is reversed if the ad nuisance cost is relatively low, $\lambda < \frac{1}{2}v_c$. Consumers become less price elastic, and a^m is larger. The implication is that the ad rate will be lower, attracting advertisers with lower matching probabilities. To enhance the total profit, the platform has to limit ad sales by increasing the ad rate, leading to $a_h < a^m$ and $a_l < a^m$. At these ad levels, increasing ad volume can increase advertising externality, according to Lemma 2. Hence, to screen consumers, incentive compatibility requires that $a_h > a_l$.

It is worth highlighting the managerial implication of this result: it is not always optimal to associate a premium version with fewer ads. Indeed, there are markets in which media firms offer a premium version that comes with more ads, the opposite of the "paying to remove ads" phenomenon. Some leading newspapers had started to adopt a "paywall" tactic, in which a paid version is introduced after readers reach the limit of free articles. The main examples of this are the *Wall Street Journal* and the *New York Times*. Presumably, a paid version enables unlimited access to the content and thus allows for more ad exposure, whereas a free version can accommodate only a limited number of ads. This is consistent with the prediction of the model when $\lambda < \frac{1}{2}v_c$. Arguably, the nuisance cost of seeing ads in a newspaper is relatively lower than that for video or music streaming.[12] When reading newspapers, readers can easily determine whether to skip an ad. Many video or music streaming services, however, do not allow users to easily skip the ads. It is important for media platforms to evaluate the relationship between the nuisance cost and the value of an ad when designing their allocation policy.

The second part of Proposition 2 identifies the interesting possibility that the ad allocation for low-type consumers may be very extreme: they either see no ads or see all the ads. This situation arises when the high-type segment dominates (i.e., $\alpha\beta > 1$) and the consumer surplus is relatively larger than the advertiser surplus in the product market. As the basic version is loaded with more ads (i.e., a_l increases), high-type consumers become more willing to subscribe to the premium version. Hence, the platform assigns more weight to the subscription revenue gained from these consumers. The total profit is then convex in a_l. To maximize profit, a_l has to take the extreme value, either 0 or 1.

The third part of this proposition demonstrates that two-sided price discrimination by media platforms can lead to cross-segment externality, which can be either positive or negative depending on the ad nuisance cost and the value of ad exposure to consumers. As is evident from Equation (9), the optimal ad allocation to low-type consumers is affected by the number of high-type consumers. Figure 3 provides an illustration. As the high-type segment grows, the platform tends to enlarge the quality difference between the two versions. If the ad nuisance cost is relatively low, then the advertising externality increases with ad quantity according to the first part of the proposition and Lemma 2. Hence, the platform should offer a lower total quality (content plus ads) to low-type consumers to increase the perceived quality difference between the two versions. However, if the nuisance cost is relatively high, then the total quality decreases with ad quantity, and thus it is optimal for the platform to increase the ad supply to low-type consumers.

Figure 3. (Color online) Effect of the Number of High-Type Consumers on Ad Volume for Low-Type Consumers



4.2 Profit-Maximizing vs. Welfare-Maximizing Allocations

Next, I turn to the question of how the optimal allocation by the platform is different from the socially efficient allocation. Define the following quantities:

$$\lambda_h \equiv \frac{\theta_h v_c + v_a}{3\theta_h v_c + 4v_a} v_c;$$
$$\lambda_l \equiv \frac{\theta_l v_c + v_a}{3\theta_l v_c + 4v_a} v_c; \text{ and}$$
$$\beta_l' \equiv \frac{\theta_l v_c + v_a - \lambda\theta_l}{\theta_l v_c + v_a - \lambda(2\theta_l - \theta_h)}. \tag{10}$$

The following proposition summarizes the differences between the profit-maximizing allocation in Equation (9) and the efficient allocation in Lemma 1.

Proposition 3. *Compared with the welfare-maximizing benchmark $(\hat{a}_h, \hat{a}_l)$, under profit-maximizing allocation,*

(a) *high-type consumers always see less ads ($a_{h^*} < \hat{a}_h$), but low-type consumers may either see more ads ($a_{l^*} > \hat{a}_l$) if $\lambda > v_c/2$ and $\beta > \beta'$, or see less ads if otherwise; and*

(b) *both types of consumers may be subject to either higher or lower advertising externality:* (i) $u_e(a_h^*) < u_e(\hat{a}_h)$ *if $\lambda < \lambda_h$; otherwise, $u_e(a_h^*) > u_e(\hat{a}_h)$;* (ii) $u_e(a_l^*) > u_e(\hat{a}_l)$ *if $\lambda > \lambda_l$ and β is sufficiently small; otherwise, $u_e(a_l^*) < u_e(\hat{a}_l)$.*

The first part of this proposition focuses on the *quantity* of ads. The platform always prefers a less than efficient level of advertising for high-type consumers. Intuitively, the platform tends to charge a high ad rate and restrict the ad demand to increase ad revenue, whereas a social planner is interested in matching more advertisers with consumers. In contrast, under profit maximization of the media platform, low-type consumers may receive more ads than the social optimum because the platform tends to increase the "quality gap" between the two versions by allocating more ads to the basic version targeted at low-type consumers.

The second part of Proposition 3 focuses on the *quality* provision affected by advertising. Recall that the total quality includes both the fixed content quality and the advertising externality $Q = q + u_e(a)$. It suffices to just focus on the externality component, $u_e(a)$. Because the advertising externality changes nonmonotonically with ad volume following Lemma 2, the distortion of quality, from the perspective of social efficiency, is also nonmonotonic and depends on the ad nuisance cost and the number of high-type consumers. The second part of this proposition illustrates that the well-known efficiency result in a one-sided market (Mussa and Rosen 1978), in which quality is distorted downward (i.e., lower-type consumers are provided with less quality than the efficient level), can fail in a two-sided market.

In particular, high-type consumers may be provided with lower quality than the efficient level if the nuisance cost is sufficiently low (i.e., $\lambda < \lambda_h$). Although these consumers are less annoyed by ads when consuming content, they also gain less from interacting with advertisers in the product market because the platform tends to charge a high ad rate and limit the ad supply. Thus, the net externality is relatively small, although it is positive. In contrast, low-type consumers may be provided with a more than efficient level of quality. This situation occurs when the nuisance cost is sufficiently high (i.e., $\lambda > \lambda_l$) and the segment of high-type consumers is relatively small. Recall that the efficient allocation is not affected by the size of a consumer segment. However, from the third part of Proposition 2 we learn that under profit maximization, $u_e(a_l^*)$ is negatively affected by the size of the high-type segment. Therefore, $u_e(a_l^*)$ can be larger than $u_e(\hat{a}_l)$ if β is sufficiently small. Note that as $\lambda_l < \lambda_h$, both $u_e(a_h^*) < u_e(\hat{a}_h)$ and $u_e(a_l^*) > u_e(\hat{a}_l)$ can occur simultaneously, implying a reversal of the standard quality-distortion result in one-sided price discrimination.

5. When Does Two-Sided Price Discrimination Fail?

The analysis thus far has focused on how price discrimination via versioning can be implemented by a profit-maximizing media platform. A major theme is that price discrimination is self-reinforcing in a two-sided market: segmentation on one side facilitates segmentation on the other. This mechanism arises regardless of whether the advertising externality on consumers depends on consumer type. A natural question to ask is therefore the following: what market features can weaken this mechanism?

Generally, if we apply the insights gained from standard one-sided markets, competition between media platforms may weaken the incentive to price discriminate. In this section, however, I focus on features pertaining to two-sided markets. The first feature to be examined is the dependency of advertiser preferences on consumer type. In particular, advertisers may prefer to convert a high-type instead of a low-type consumer. The second feature is the informational friction of matching advertisers and consumers. Advertisers must then compete for consumers' attention by choosing how much to advertise. To keep the analysis simple and focused, in both investigations, I maintain the baseline assumption that $v_c = 0$ without losing much qualitative insight.

5.1. When Advertisers Value Higher-Type Consumers More

At least two observations can motivate the assumption that advertisers' preference depends on consumer type. First, in reality, price sensitivity and marginal utility of quality are often correlated. Higher quality is typically valued by fewer rather than more price-sensitive consumers. Consumers who pay \$10 per month for the premium service YouTube Red are more likely to be willing to pay a higher price for products advertised on the platform. Second, many advertisers prefer to advertise in media outlets where consumers are most likely to buy their products. For example, manufacturers of sports products may be more likely to advertise on sports channels than on classical music channels because sports fans are likely to have a stronger preference for sports products. Recent empirical research provides supporting evidence for such a relationship (Wu 2015).

To incorporate this feature, let us assume that the ex post surplus an advertiser can extract is $v_{a,h}$ from converting a high-type consumer and is $v_{a,l}$ from converting a low-type consumer. Note, however, that advertisers cannot distinguish consumers ex ante because θ is consumers' private information. To harvest the higher surplus from high-type consumers, advertisers need to identify the type of each consumer with the help of a media platform. In particular, the platform can take three possible approaches when implementing versioning:

1. *Broad reach with targeting.* Advertisers can reach both consumer types and target a specific type.
2. *Broad reach without targeting.* Advertisers can reach both consumer types but cannot target a specific type.
3. *Narrow reach without targeting.* Advertisers can reach only one of the two consumer types.

The first approach involves targeting. If the platform can successfully discriminate among consumers, then by labeling each consumer, it can achieve targeted advertising. Thus, advertisers choosing to reach both segments and target each segment can obtain the maximum profit in the product market $\beta v_{a,h} + (1-\beta)v_{a,l}$. In essence, the platform helps the advertisers achieve perfect price discrimination. Alternatively, the platform can disable targeting when selling both consumer segments to advertisers. Then these advertisers cannot directly distinguish the two types of consumers. They could, however, practice some degree of price discrimination in the product market. Hence, their profit $\overline{v}_a$ is bounded above by the profit under perfect discrimination $\beta v_{a,h} + (1-\beta)v_{a,l}$ and bounded below by the lowest profit $v_{a,l}$. That is, $\overline{v}_a \in [v_{a,l}, \beta v_{a,h} + (1-\beta)v_{a,l}]$. By taking the last approach, the platform can let higher-type advertisers focus only on high-type consumers. Targeting becomes irrelevant here. The versioning policy now involves allocating the top advertisers only to high-type consumers and the remaining participating advertisers to low-type consumers.

Proposition 4. *If high-type consumers dominate, $\alpha\beta > 1$, then versioning with targeting is always optimal to the media platform. If $\alpha\beta < 1$, then there exist thresholds v_h' and v_h'' such that when $v_{a,h} > \max\{v_h', v_h''\}$ and $\overline{v}_a > v_h'$, none of the versioning polices are implementable.*

Proposition 4 establishes that it is not always optimal for the platform to implement versioning with targeting. The main mechanism behind it is the conflict between discriminating among consumers and targeting higher-type consumers. Intuitively, when advertisers expect more gain from targeting high-type consumers (i.e., $v_{a,h}$ becomes larger), there is more room for the platform to extract advertising revenue by allowing more advertisers to reach both segments. Then the ability to target each consumer segment allows the advertisers to extract surplus in each segment separately and thus maximize profit. Consequently, the quality gap between the two versions determined by $a_l^* - a_h^*$ shrinks and eventually disappears as $v_{a,h}$ becomes sufficiently high. There is then no basis for discriminating among consumers, rendering targeting infeasible. A similar logic applies to the case in which targeted advertising is disabled by the platform. The higher surplus from high-type consumers may allow advertisers to extract more surplus in the product market. That is, $\overline{v}_a$ is greater than $v_{a,l}$ and becomes even higher if advertisers can appropriate more rent by practicing price discrimination. Then the demand in the advertising market becomes less elastic; hence, the platform tends to allocate more top advertisers to high-type consumers, clashing with the incentive to discriminate on the consumer side.

5.2. When Advertisers Compete for Consumer Attention

To investigate the second feature, let us maintain the baseline assumption that advertisers have a homogeneous preference for consumers conditional on a match. The focus now, however, is on how advertisers behave differently if they compete for consumer attention by choosing the level of ad intensity. In reality, advertisers not only decide whether to advertise but also determine how frequently or how much they should advertise for a variety of reasons. Some consumers may simply fail to notice a particular ad given the large number of ads on a platform. They may also skip a television commercial by switching channels or taking a bathroom break. Some consumers may even forget an ad after subsequently consuming the media content. All these frictions can contribute to the need for repeated ad exposure.

I follow the approach developed by Butters (1977) (see also Bergemann and Bonatti 2011) to capture the informational frictions between advertisers and consumers. A type σ advertiser determines the number of ad messages $m(\sigma)$ to distribute on the media platform, and each message reaches a random consumer with a uniform probability. Let $I(m)$ denote the probability that a consumer becomes informed, given message volume m. We can assume a uniform random matching process to derive a convenient form of $I(m)$. In particular, suppose that a large number of messages m is distributed uniformly among a large number of consumers, denoted by n. Then the probability that a consumer receives none of the m messages is given by $(1-1/n)^m$. Taking the limit of both m and n while holding the ratio m/n constant, we have

$$\lim_{m,n\to\infty}(1-1/n)^m = e^{-\frac{m}{n}}. \tag{11}$$

It then follows that the probability of being informed is $1-e^{-m/\beta}$ if the ad messages are sent to high-type consumers only and is $1-e^{-m/(1-\beta)}$ if the recipients are low-type consumers only. In addition to its analytical convenience, this functional form has two attractive properties. First, as the segment size increases, consumers are less likely to be informed. Second, for a given segment size, the probability of being informed increases with the number of messages sent (i.e., $\partial I(m)/\partial m > 0$) but at a decreasing rate (i.e., $\partial^2 I(m)/\partial m^2 < 0$). This property captures the diminishing effectiveness of repeated advertising.

A type σ advertiser needs to determine the optimal number of messages given the ad rate per message p^a. If the messages are delivered to high-type consumers only, then the problem is

$$\max_{m} \quad \pi_\sigma = \left(1-e^{-\frac{m}{\beta}}\right)\sigma\beta v_a - mp^a. \tag{12}$$

The optimal ad intensity is $m_h^*(p^a;\sigma) = \beta\ln(\sigma v_a/p^a)$. In the same vein, restricting to low-type consumers yields the optimal strategy $m_l^*(p^a;\sigma) = (1-\beta)\ln(\sigma v_a/p^a)$. These solutions suggest that advertisers of a higher matching quality have a stronger incentive to send more ad messages and that a larger consumer market induces more ad messages. Hence, like the baseline analysis, the optimal policy for the platform is to persuade the top advertisers to reach the entire market while letting the remaining advertisers focus on the larger segment. For given ad rates (p_h^a, p_l^a), we can obtain the total ad quantity in each segment $a_h = \int_{\sigma_h}^1 m_h^*(p_h^a;\sigma)d\sigma$ and $a_l = \int_{\sigma_l}^1 m_l^*(p_l^a;\sigma)d\sigma$, where σ_h and σ_l are the marginal advertisers such that $m_h^*(\sigma_h) = 0$ and $m_l^*(\sigma_l) = 0$. Following the same approach as in the baseline analysis, we can then establish the equilibrium solution (a_h^*, a_l^*) and obtain the following result.

Proposition 5. (a) *If the fraction of high-type consumers* β *is sufficiently large and* α *is sufficiently small, then versioning based on ad intensity is not implementable.* (b) *When versioning is implementable, high-type consumers interact with fewer advertisers than low-type consumers if* $\alpha\beta > 1/2$ *but with more advertisers otherwise.*

The first part of this proposition highlights the tension between price discrimination and rent extraction from high-type consumers. Following the baseline analysis, if high-type consumers have a stronger preference for media content or avoiding ad nuisance (i.e., α becomes larger), then the platform has a stronger incentive to price discriminate, and the difference in ad allocation across the two versions becomes larger. However, advertisers' competition for consumer attention introduces an opposing effect. If there are more high-type consumers (i.e., β becomes larger), then advertisers are motivated to send even more ad messages to them. They may then expect a smaller quality difference between the two versions, clashing with the platform's incentive to discriminate on the consumer side. Hence, if β is sufficiently large and α is sufficiently small, then the competition effect can dominate, and the platform cannot implement versioning because there is no ground for separating different types of consumers.

The second part of this proposition indicates an interesting outcome: even though high-type consumers should expect fewer ads under versioning, because of the implementability condition, it is possible that they will interact with a larger set of advertisers. This result is mainly driven by the fact that the optimal number of messages sent by an advertiser depends on market size. When there are fewer high-type consumers, each advertiser tends to send fewer messages to this segment, whereas low-type consumers receive more messages per advertiser. Thus, even when high-type consumers are reached by more advertisers, they may still see fewer ad messages in total than low-type consumers.

Both Propositions 4 and 5 suggest that media platforms need to carefully assess the market environment when considering two-sided price discrimination in practice. The two propositions highlight under which circumstances two-sided price discrimination is less valuable. Proposition 4 focuses on advertisers' heterogeneity in their intrinsic preferences for contacting different types of consumer, whereas Proposition 5 concentrates on how the matching process can endogenously generate differences in advertisers' preferences. Even though the underlying mechanisms are quite different, both propositions highlight the potential conflict between media platforms and advertisers in exploiting high-type consumers. The implication for media managers is that they must be cautious about such conflict when evaluating the feasibility of two-sided price discrimination. Future work could continue to explore alternative market features that contribute to this conflict and thus provide further insights into the boundary of two-sided discrimination.

6. Implications for Product Line Management of Media Platforms

Thus far I have assumed that content quality is the same for different versions, to focus on the question of how versioning can be implemented solely with ad differentiation. In reality, media platforms can differentiate their content products for different versions. For example, YouTube's premium-version YouTube Red not only removes ads but also provides subscribers with exclusive access to offline videos and original shows. The *Wall Street Journal* differentiates its paid version from the free version by allowing unlimited access to news content. Equipped with two possible instruments for discrimination (i.e., ad and content differentiation), media platforms face the strategic question of how to use them to design their product lines. In this section, I extend the baseline analysis to examine this problem.

Suppose that a media platform can introduce two versions of its content product with different qualities (q_H, q_L) with $q_H > q_L$. For many information goods, higher quality does not come with a much higher marginal cost, although it may entail a higher fixed upfront investment cost. Motivated by this observation, I assume that the marginal costs of both versions are the same and normalized to zero. If the platform does not offer advertising and monetizes its services purely through subscription revenues, then the problem reduces to the classic second-degree price discrimination with exogenous qualities in a standard monopoly market. One can show that such one-sided price discrimination through versioning is generally inferior to selling the high-quality version alone.

Lemma 3. *If the media platform is not financed by advertising, then versioning is never strictly optimal:* (a) *if* $\alpha\beta < 1$, *then it is optimal to sell* q_H *to both types of consumers;* (b) *if* $\alpha\beta > 1$, *then it is optimal to sell* q_H *to high-type consumers only.*

Intuitively, with the assumption that the marginal costs are the same, selling the low-quality version is never profitable because it does not save any costs and consumers are only willing to pay less for it. Hence, versioning is never strictly optimal, even though consumers are heterogeneous. This intuition is quite general and does not depend on the consumer types being discrete or the content quality being exogenous.[13] Technically, with minimal difference in the marginal cost of production, the social surplus (consumers' willingness to pay less the marginal cost) becomes log submodular.[14] As Anderson and Dana (2009) show, firms may want to offer a single product quality rather than price discriminate under various settings, including endogenous product quality (either discrete or continuous).

Lemma 3 contrasts with the baseline analysis that versioning can be an optimal strategy when only ad differentiation is used as an instrument. Ad differentiation is unlike content differentiation as a price discrimination tool because it can facilitate further discrimination in the advertising market. The following proposition summarizes the optimal media policy when the media platform can use both instruments to discriminate.

Proposition 6. *There exists a threshold $\bar{\beta} \in (1/\alpha, 1)$ such that* (a) *if $\beta < 1/\alpha$, then it is optimal to sell only q_H and adopt versioning with purely ad differentiation;* (b) *if $\beta \in (1/\alpha, \bar{\beta})$, then it is optimal to sell both q_H and q_L and adopt versioning using both content and ad differentiation; and* (c) *if $\beta > \bar{\beta}$, then it is optimal to sell only q_H and serve only high-type consumers.*

Under the versioning strategy purely based on ad differentiation (as in Section 3), the platform will always choose to sell the content at a high quality because a lower-quality product will only reduce subscription revenue without affecting advertising revenue. The optimal subscription fees are $p_l^c = \theta_l(q_H - \lambda a_l)$ and $p_h^c = p_l^c + \theta_h \lambda(a_l - a_h)$. The alternative versioning strategy with both ad and content differentiation allows high-type consumers to select the high-quality product and low-type consumers to choose the low-quality one. The optimal subscription fees now become $p_l^c = \theta_l(q_L - \lambda a_l)$ and $p_h^c = p_l^c + \theta_h\lambda(a_l - a_h) + \theta_h(q_H - q_L)$. For both versioning strategies, the optimal ad allocations (a_h^*, a_l^*) are exactly the same. Hence, the difference between the two strategies lies in the subscription revenues. Let π^{V1*} and π^{V2*} denote the optimal profits for the first (ad differentiation only) and second (both content and ad differentiation) strategies. It follows that

$$\pi^{V2*} - \pi^{V1*} = \underbrace{\beta(\theta_h - \theta_l)(q_H - q_L)}_{\text{gain from selling to the hightype}} - \underbrace{(1-\beta)\theta_l(q_H - q_L)}_{\text{loss from selling to the lowtype}} = (\beta\theta_h - \theta_l)(q_H - q_L). \tag{13}$$

Introducing an additional instrument based on content differentiation increases the subscription revenue from high-type consumers because of the increased quality gap but reduces the revenue from low-type consumers because of the lower price charged to them. The net difference increases with more high-type consumers in the market. As long as the high-type segment is sufficiently large, the gain outweighs the loss from the low-type segment, and thus the versioning strategy with both ad and content differentiation becomes profitable (the second part of Proposition 6). Hence, content differentiation can become useful when the platform considers advertising revenue. This result further illustrates how two-sided price discrimination can differ from one-sided discrimination. As shown in Lemma 3, when considering only the consumer side, the platform finds it challenging to discriminate among consumers using different content versions that have little difference in marginal cost. Versioning can become feasible when the platform discriminates on both sides.

However, there is a limit to this result, as indicated in the third part of Proposition 6. If the market mainly consists of high-type consumers, then it is optimal not to practice any price discrimination. The optimal strategy is to serve only high-type consumers.[15]

Another insight from Proposition 6 is that content differentiation, although it has no direct impact on ad allocation, can also facilitate the implementation of ad differentiation. As shown in Section 5, versioning based on ad differentiation can break down when there is a conflict between the media platform and advertisers in exploiting high-type consumers. One way to restore the feasibility of versioning is then to introduce content differentiation. Combining content and ad differentiation can help the platform achieve a higher profit. Then, based on the results in Section 5, it is possible that high-type consumers may see *more* ads than low-type consumers either because advertisers have a stronger preference for reaching high-type consumers or because advertisers compete for their attention.

This observation reiterates the managerial implication, first shown based on Proposition 2, that it is not always optimal to associate a premium version with fewer ads. However, the mechanism here is quite different. The result in Proposition 2 relies on type-dependent externality and the assumption that the ad nuisance cost is relatively low. It does not require content differentiation. In contrast, the mechanism here depends on advertisers either strongly preferring high-type consumers or intensively competing for the attention of these consumers. Furthermore, this mechanism does require content differentiation to enable the self-selection of consumers.

7. Concluding Remarks

In this article, I analyze the increasingly common practice of differentiating ad allocations to segment consumers on media platforms. By integrating the two-sided market and second-degree price discrimination models, I explore how platforms can implement price discrimination on both sides of the market. One key insight is that price discrimination on one side can strengthen discrimination on the other. Consumers are screened based on their heterogeneous product preferences, and advertisers are screened according to their heterogeneous preferences for reaching different consumer segments. The analysis also reveals important market features that can shape media's product policies, including the dependency of advertising externality on consumer type, the magnitude of ad nuisance cost, and the ex post surplus that consumers and advertisers expect in the product market. Furthermore, I explore two important market structures that can render price discrimination infeasible, namely the dependency of advertiser preferences on consumer type and the informational friction between advertisers and consumers. As summarized in Table 2, the analysis produces a rich set of predictions that can be relevant for managerial practice.

The model is formulated to enable tractable analysis and inevitably involves some simplifying assumptions. The first simplification is the abstract model of the product market. This basic setup can be extended to incorporate richer market structures. For example, researchers may be interested in how horizontally differentiated advertisers competing in the product market affect the pricing and advertising policies of media. Second, this study focuses on the incentives of media platforms in a monopoly market. Future research could extend this analysis to incorporate competition among media. Although competition generally weakens price discrimination, it is less clear a priori how it impacts different sides of a media market and different consumer types and how the impact depends on consumers' propensity to multihome. Third, the model assumes that a media platform can post no more than two ad rates in equilibrium to permit self-selection by heterogeneous advertisers. This list-price format implies that some advertisers can secure a positive surplus after paying for ad space. One might ask what happens if platforms can extract more surplus from advertisers by selling ad space through auctions, ad networks, or agencies. Whether a more efficient means of selling ads would change their implementation, profitability, and welfare results is worthy of investigation.

Last, although the theory provides some guidance on how equilibrium advertising policies are shaped, examining these policies in an empirical context can provide additional insights. For example, the theory indicates certain conditions in which an entirely ad-free premium is optimal. Further research could empirically test these conditions. Empirical studies can also illuminate how much externality advertising exerts on consumers in different media and whether consumers expect informational gains from advertising exposure.

Table 2. Highlights of Key Predictions Under Different Market Scenarios

1. Consumers care only about the quantity of ads.
   Low-type consumers see more ads than high-type consumers.
2. Consumers care about both the quantity and quality of ads.
   For low ad nuisance cost, high-type consumers see more ads than low-type consumers.
   For high ad nuisance cost, high-type consumers see less ads than low-type consumers.
3. Advertisers prefer high-type consumers.
   Versioning fails if high-type consumers do not dominate and advertisers extract a sufficiently high surplus from high-type consumers.
4. Advertisers compete for consumer attention.
   Versioning fails if the fraction of high-type consumers is sufficiently large and the consumer heterogeneity in marginal utility is sufficiently small.
5. Media platform can vary both content quality and ad allocations.
   It is optimal to implement versioning with both ad and content differentiations if the fraction of high-type consumers is moderately large.

Acknowledgments

The author thanks the senior editor, associate editor, and two anonymous reviewers for their constructive comments that improved the paper. The author gratefully acknowledges helpful comments from Mark Armstrong, Mengze Shi, Jiwoong Shin, Birger Wernerfelt, Juanjuan Zhang, seminar participants at Peking University and National University of Singapore, and attendees of the 2017 Marketing Science Conference, and 2017 UTD FORMS Conference.

Appendix

A.1. Proof of Lemma 1

The first-order conditions for the welfare-maximization problem in Equation (3) are

$$\beta\theta_h(-\lambda + (1 - a_h)v_c) + \beta(1 - a_h)v_a = 0,$$
$$(1 - \beta)\theta_l(-\lambda + (1 - a_l)v_c) + (1 - \beta)(1 - a_l)v_a = 0,$$

which lead to the solution expressed in the lemma. It is straightforward to check that the second-order conditions are also satisfied. Note further that

$$\hat{a}_h - \frac{v_c - \lambda}{v_c} = \frac{\lambda v_a}{(\theta_h v_c + v_a)v_c} > 0 \text{ and}$$
$$\hat{a}_l - \hat{a}_h = \frac{\lambda(\theta_h - \theta_l)v_a}{(\theta_h v_c + v_a)(\theta_l v_c + v_a)} > 0.$$

Hence, we have $\hat{a}_l > \hat{a}_h > \frac{v_c - \lambda}{v_c}$. It remains to show that there is no other alternative allocation that can improve social welfare.

This follows a similar and tedious procedure as in the proof of Proposition 1 and thus it is omitted here. □

A.2. Proof of Proposition 1

As noted in the main text, any incentive-compatible design should involve higher-type advertisers reaching more consumers because of the single-crossing property. To see this, let $x \in \{\beta, 1-\beta, 1\}$ denote the size of the consumer market. A type σ advertiser can obtain profit $x\sigma v_a$ in the product market. This profit increases with x and more so for higher values of σ (i.e., the advertiser type). Therefore, the platform can segment the participating advertisers into n groups in the order of *decreasing* type. Each group chooses a different ad product, and let $d_n \in \{h, l, hl\}$ denote the decision of the nth group. An implementation can be written as $(d_1, d_2, \ldots, d_n, \ldots, d_N)$ for a division of N groups. For example, (hl, h) means that the participating advertisers are segmented into two groups: the group of higher-type advertisers reaches both consumer segments, whereas the group of lower-type advertisers reaches high-type consumers only. Next, I validate the existence and uniqueness of the equilibrium and then show that no alternative implementation is feasible or improves profit.

A.2.1. Equilibrium Implementation (hl, l).

Using the binding IR_l and IC_h constraints, we can rewrite the platform's problem as

$$\max_{a_l, a_h} \quad \pi = \beta\theta_h(q - \lambda a_h) + (\theta_l - \beta\theta_h)(q - \lambda a_l) + a_h(1 - a_h) \times \beta v_a + a_l(1 - a_l)(1 - \beta)v_a.$$

The first-order conditions are

$$-\beta\theta_h\lambda + (1 - 2a_h)\beta v_a = 0 \text{ and}$$
$$(\theta_l - \beta\theta_h)\lambda + (1 - 2a_l)(1 - \beta)v_a = 0,$$

which lead to the optimal ad allocation (a_h^*, a_l^*) in Equation (5). The second-order conditions are also satisfied because $\partial^2\pi/\partial a_h^2 < 0$ and $\partial^2\pi/\partial a_l^2 < 0$.

A.2.2. Deviation to Implementations (hl, h) and (h).

Under these implementations, the basic version designed for low-type consumers turns out to contain fewer ads than the premium version. They will then switch to the premium version, a contradiction.

A.2.3. Deviation to Implementation (l, h).

This deviation requires that $\beta < 1/2$ so that reaching low-type consumers is more profitable to advertisers. The indifferent advertisers are $\sigma_1 = (p_l^a - p_h^a)/(1 - 2\beta)v_a$ and $\sigma_0 = p_h^a/\beta v_a$. Advertisers of type $\sigma \in [\sigma_1, 1]$ reach low-type consumers, whereas advertisers $\sigma \in [\sigma_0, \sigma_1]$ reach high-type consumers. To induce ad demands $a_l = 1 - \sigma_1$ and $a_h = \sigma_1 - \sigma_0$, the ad rates are set at $p_h^a = (1 - a_h - a_l)\beta v_a$ and $p_l^a = p_h^a + (1 - a_l)(1 - 2\beta)v_a$. The problem becomes

$$\max_{a_l, a_h, p_l^c, p_h^c} \quad \pi = \beta p_h^c + (1 - \beta)p_l^c + (a_h + a_l)(1 - a_h - a_l)\beta v_a + a_l(1 - a_l)(1 - 2\beta)v_a.$$

Let $a_{hl} = a_h + a_l$. By setting $p_l^c = \theta_l(q - \lambda a_l)$ and $p_h^c = p_l^c + \theta_h\lambda(a_l - a_h) = p_l^c + \theta_h\lambda(2a_l - a_{hl})$ and maximizing over (a_l, a_{hl}), we can obtain the optimal solution using the first-order conditions:

$$a_l^* = \frac{1}{2} - \frac{(\theta_l - 2\beta\theta_h)\lambda}{2(1 - 2\beta)v_a} \quad \text{and} \quad a_{hl}^* = \frac{1}{2} - \frac{\theta_h\lambda}{2v_a}.$$

Because $\theta_h > \frac{\theta_l - 2\beta\theta_h}{1 - 2\beta}$ implies $a_{hl}^* < a_l^*$, the implementability condition is violated. Hence, the strategy of (l, h) cannot sustain in equilibrium.

A.2.4. Deviation to Implementation (h, l).

This deviation requires that $\beta > 1/2$ so that reaching high-type consumers is more profitable to advertisers. The indifferent advertisers are $\sigma_1 = (p_h^a - p_l^a)/(2\beta - 1)v_a$ and $\sigma_0 = (p_l^a)/(1 - \beta)v_a$. To induce ad demands $a_h = 1 - \sigma_1$ and $a_l = \sigma_1 - \sigma_0$, the ad rates are set at $p_l^a = (1 - a_h - a_l)(1 - \beta)v_a$ and $p_h^a = p_l^a + (1 - a_h)(2\beta - 1)v_a$. The platform's problem becomes

$$\max_{a_l, a_h, p_l^c, p_h^c} \quad \pi = \beta p_h^c + (1 - \beta)p_l^c + (a_h + a_l)(1 - a_h - a_l)(1 - \beta) \times v_a + a_h(1 - a_h)(2\beta - 1)v_a.$$

Let $a_{hl} = a_h + a_l$. By setting $p_l^c = \theta_l(q - \lambda a_l) = \theta_l(q - \lambda a_{hl} + \lambda a_h)$ and $p_h^c = p_l^c + \theta_h\lambda(a_l - a_h) = p_l^c + \theta_h\lambda(a_{hl} - 2a_h)$ and maximizing over (a_h, a_{hl}), we can obtain the optimal conditions

$$a_h^* = \frac{1}{2} - \frac{(2\beta\theta_h - \theta_l)\lambda}{2(2\beta - 1)v_a} \quad \text{and} \quad a_{hl}^* = \frac{1}{2} - \frac{(\theta_l - \beta\theta_h)\lambda}{2(1 - \beta)v_a}.$$

Because $\beta > 1/2$, we have $\frac{2\beta\theta_h - \theta_l}{2\beta - 1} > \frac{\theta_l - \beta\theta_h}{1 - \beta}$, implying that $a_h^* < a_{hl}^*$. Thus, the proposed deviation is implementable. Next, we shall show that the strategy, however, is dominated by the implementation (hl, l). First note that a_l^* under (hl, l) is exactly the same as a_{hl}^* under (h, l). Let superscripts e and d denote the implementations under proposed equilibrium (hl, l) and under deviation (h, l), respectively. Then the difference in profit is

$$\pi^e - \pi^d = \underbrace{-\beta\theta_h\lambda a_h^e + a_h^e(1 - a_h^e)\beta v_a}_{\pi_h^e(a_h^e)} - \left(\underbrace{(\theta_l - 2\beta\theta_h)\lambda a_h^d + a_h^d(1 - a_h^d)(2\beta - 1)v_a}_{\pi_h^d(a_h^d)}\right).$$

Note that the optimal allocations satisfy $a_h^{d*} < a_h^{e*} < 1/2$. We shall show that $\pi_h^e(a_h^{e*}) \geq \pi_h^d(a_h^{d*})$. Consider the profit difference evaluated at the same a, $\Delta\pi(a) = \pi_h^e(a) - \pi_h^d(a) = (\beta\theta_h - \theta_l)\lambda a + a(1 - a)(1 - \beta)v_a$. Clearly, $\Delta\pi(0) = 0$. If $\beta\theta_h < \theta_l$, then $\Delta\pi'(0) > 0$, $\Delta\pi''(a) < 0$, and $\Delta\pi'(\frac{1}{2}) > 0$. Hence, $\Delta\pi(a) > 0$ for all $a < 1/2$. If $\beta\theta_h > \theta_l$, then $\Delta\pi'(a) > 0, \forall a < \frac{1}{2}$. Again, $\Delta\pi(a) > 0$ for all $a < 1/2$. Then $\pi_h^e(a_h^{e*}) > \pi_h^e(a_h^{d*}) > \pi_h^d(a_h^{d*})$, implying that the implementation (h, l) does not improve profit.

A.2.5. Deviation to Implementation (hl, l, h).

This implementation is consistent with the preference order of advertisers when $\beta < 1/2$. Advertisers are segmented into three groups, with thresholds $\sigma_2 = (p_{hl}^a - p_l^a)/\beta v_a$, $\sigma_1 = (p_l^a - p_h^a)/(1 - 2\beta)v_a$, and $\sigma_0 = p_h^a/\beta v_a$. The size of each segment is $a_1 = 1 - \sigma_2$, $a_2 = \sigma_2 - \sigma_1$, and $a_3 = \sigma_1 - \sigma_0$. Note that the ad allocations of the two versions are $a_h = a_1 + a_3$ and $a_l = a_1 + a_2$. Letting $a_{hl} = a_1 + a_2 + a_3$ and substituting the optimal prices as a function of advertising intensities, we can write the platform's problem as

$$\max_{a_{hl},a_l,a_1} \quad \pi = \beta\theta_h\big(q - \lambda(a_1 + a_{hl} - a_l)\big) + \big(\theta_l - \beta\theta_h\big)\big(q - \lambda a_l\big) + a_1(1-a_1)\beta v_a + a_l(1-a_l)\big(1-2\beta\big)v_a + a_{hl} \times (1-a_{hl})\beta v_a.$$

The first-order conditions lead to an interior solution that satisfies $a_1^* = a_{hl}^* < a_l^*$, violating the implementability condition. Hence, the strategy of (hl, l, h) cannot constitute an equilibrium.

A.2.6. Deviation to Implementation (hl, h, l)**.** This implementation is consistent with the preference order of advertisers when $\beta > 1/2$. Advertisers are segmented into three groups, with thresholds $\sigma_2 = (p_{hl}^a - p_l^a)/(1-\beta)v_a$, $\sigma_1 = (p_h^a - p_l^a)/(2\beta - 1)v_a$, and $\sigma_0 = (p_l^a)/(1-\beta)v_a$. The size of each segment is $a_1 = 1 - \sigma_2$, $a_2 = \sigma_2 - \sigma_1$, and $a_3 = \sigma_1 - \sigma_0$. The ad allocations of the two versions are $a_h = a_1 + a_2$ and $a_l = a_1 + a_3$. Letting $a_{hl} = a_1 + a_2 + a_3$ and substituting the optimal prices as a function of advertising intensities, we can write the platform's problem as

$$\max_{a_{hl},a_h,a_1} \quad \pi = \beta\theta_h\big(q - \lambda a_h\big) + \big(\theta_l - \beta\theta_h\big)\big(q - \lambda(a_{hl} - a_h - a_1)\big) + a_1(1-a_1)\big(1-\beta\big)v_a + a_h(1-a_h)\big(2\beta - 1\big)v_a + a_{hl}(1-a_{hl})\big(1-\beta\big)v_a.$$

The first-order conditions lead to an interior solution that satisfies $a_1^* = a_{hl}^* > a_h^*$, violating the implementability condition. Hence, the strategy of (hl, h, l) cannot sustain. □

A.3. Proof of Corollary 1

First, consider the case of discriminating among consumers without discrimination of advertisers. This can be achieved by a random ad allocation: every advertiser is randomly allocated to high-type consumers with probability $\omega \in [0, 1]$ and to the low-type with probability $1 - \omega$, with the constraint that $\omega < 1/2$. The indifferent advertiser is then given by $\sigma_0 = p_0^a/(\omega\beta + (1-\omega)(1-\beta))v_a$, which determines the ad quantity $a = 1 - \sigma_0$. After applying the incentive constraints, we can rewrite the platform's problem as follows:

$$\max_{a,\omega} \quad \pi = \beta\theta_h\big(q - \lambda\omega a\big) + \big(\theta_l - \beta\theta_h\big)\big(q - \lambda(1-\omega)a\big) + a \times (1-a)\big(\omega\beta + (1-\omega)\big(1-\beta\big)\big)v_a.$$

Because the profit function is linear in ω, the optimal strategy is $\omega = 0$ ($\omega = 1$ is ruled out because of the implementability condition). Thus, versioning under random ad assignment always leads to an ad-free premium. Substituting this quantity back to the profit function, one can solve for the optimal ad quantity $a^{r*} = \frac{1}{2} - \frac{(\theta_l - \beta\theta_h)\lambda}{2(1-\beta)v_a}$, which is exactly the same as a_l^* in Equation (5). Then the optimal profit under random allocation is no greater than that under the policy specified in Proposition 1. Hence, one-sided price discrimination under random allocation is generally inferior to two-sided price discrimination. The two become equivalent if $\theta_h\lambda \geq v_a$. Then the proposed versioning policy in the baseline analysis also leads to an ad-free premium version $a_h^* = 0$.

Second, suppose that the platform intends to discriminate among advertisers without discriminating consumers. Thus, the consumer type information cannot be used to discriminate advertisers. With other instruments unavailable, the only possible way to discriminate advertisers is to let them choose the number of consumers to reach at different rates. Consider the simplest approach that advertisers can choose to reach all consumers (size one) at price p_1^a or to reach a fraction $\rho < 1$ of consumers at price p_0^a. Then advertisers with $\sigma > \sigma_1$ choose to reach all, whereas advertisers $\sigma \in [\sigma_0, \sigma_1]$ reach the fraction of consumers, where the indifferent advertisers are given by $\sigma_1 = (p_1^a - p_0^a)/(1-\rho)v_a$ and $\sigma_0 = p_0^a/\rho v_a$. We can write $a_1 = 1 - \sigma_1$ and $a_0 = 1 - \sigma_0$ and obtain the ad rates $p_0^a = (1 - a_1)\rho v_a$ and $p_1^a = p_0^a + (1 - a_1)(1-\rho)v_a$. The platform's problem becomes

$$\max_{a_0,a_1,p^c} \quad \pi = p^c + a_0(1-a_0)\rho v_a + a_1(1-a_1)\big(1-\rho\big)v_a, \tag{A.1}$$

subject to the constraint that $p^c \leq \theta_l(q - \rho\lambda a_0 - (1-\rho)\lambda a_1)$. Clearly, the constraint has to bind, and we can substitute it back to the objective function and solve for the optimal (a_1^*, a_0^*). It follows, then, that $a_1^* = a_0^*$, which implies no segmentation of advertisers, a contradiction. Hence, the simple approach of segmenting advertisers into two segments cannot sustain. This result can be easily generalized to any segmentation plan with more than two segments. □

A.4. Proof of Lemma 2

Note that $u_e(a)$ attains maximum at $a = (v_c - \lambda)/v_c \equiv a^m$, which is positive if $\lambda < v_c$ and negative otherwise. Because $u_e(a)$ is quadratic, it first increases with a before reaching a^m and then decreases with a. If $\lambda < v_c$, then $a^m < 0$. Hence, $u_e(a)$ decreases with a and $u_e(a) < u_e(0) = 0$. □

A.5. Proof of Proposition 2

Following the baseline analysis, the indifferent advertisers are given by $\sigma_h = (p_{hl}^a - p_l^a)/\beta v_a$ and $\sigma_l = p_l^a/(1-\beta)v_a$. The platform's problem is the same as that in Equation (4), with the following constraints:

$IR_l: \theta_l\big(q + u_e(a_l)\big) - p_l^c \geq 0;$
$IR_h: \theta_h\big(q + u_e(a_h)\big) - p_h^c \geq 0;$
$IC_l: \theta_l\big(q + u_e(a_l)\big) - p_l^c \geq \theta_l\big(q + u_e(a_h)\big) - p_h^c;$
$IC_h: \theta_h\big(q + u_e(a_h)\big) - p_h^c \geq \theta_h\big(q + u_e(a_l)\big) - p_l^c;$ and
Implementability: $u_e(a_h) > u_e(a_l)$.

Ignoring the IR_h, IC_l, and implementability constraints for a moment, we can use the binding IR_l and IC_h constraints to set the optimal price-advertising relationships $p_l^c = \theta_l(q + u_e(a_l))$ and $p_h^c = p_l^c + \theta_h(u_e(a_h) - u_e(a_l))$. Substituting these prices back to the profit function and deriving the first-order conditions, we can obtain the interior solution in Equation (9).

Turning to the second-order conditions, note that $\partial^2\pi/\partial a_h^2 = -\beta\theta_h v_c - 2\beta v_a < 0$ and

$$\frac{\partial^2\pi}{\partial a_l^2} = -\big(\theta_l - \beta\theta_h\big)v_c - 2\big(1-\beta\big)v_a.$$

If $\frac{v_a}{v_c} > \frac{\beta\theta_h - \theta_l}{2(1-\beta)}$, then $\partial^2\pi/\partial a_l^2 < 0$, and thus the solution is indeed the optimal one. If contrarily $\frac{v_a}{v_c} < \frac{\beta\theta_h - \theta_l}{2(1-\beta)}$ (which requires that $\alpha\beta > 1$ because both v_c and v_a are nonnegative), then $\partial^2\pi/\partial a_l^2 > 0$. The optimal solution is either $a_l^* = 0$ or $a_l^* = 1$. Note that

$$\pi\big(a_h^*, a_l = 1\big) - \pi\big(a_h^*, a_l = 0\big) = \big(\theta_l - \beta\theta_h\big)u_e(1)\begin{cases} > 0, & \text{if } u_e(1) < 0 \Leftrightarrow \lambda > v_c/2, \\ < 0, & \text{if } u_e(1) < 0 \Leftrightarrow \lambda < v_c/2. \end{cases}$$

It follows that $a_l^* = 0$ if $\lambda < v_c/2$, whereas $a_l^* = 1$ if $\lambda > v_c/2$, establishing the second result of the proposition.

Next, let us examine the implementability of the solution and compare the externality for the two consumer segments. There are two cases. First, if $\frac{v_a}{v_c} > \frac{\beta\theta_h - \theta_l}{2(1-\beta)}$, then the interior solution holds. With some algebra, we have

$$a_h^* - \frac{v_c - \lambda}{v_c} = \frac{(2\lambda - v_c)v_a}{(\theta_h v_c + 2v_a)v_c},$$
$$a_l^* - \frac{v_c - \lambda}{v_c} = \frac{(2\lambda - v_c)v_a}{\big((\theta_l - \beta\theta_h)v_c + 2\big(1-\beta\big)v_a\big)v_c},$$

and thus

$$a_l^* - a_h^* = \frac{(2\lambda - v_c)(\theta_h - \theta_l)v_a}{\big((\theta_l - \beta\theta_h)v_c + 2\big(1-\beta\big)v_a\big)(\theta_h v_c + 2v_a)}.$$

Then $a_l^* < a_h^* < (v_c - \lambda)/v_c$ when $\lambda < v_c/2$ and $a_l^* > a_h^* > (v_c - \lambda)/v_c$ when $\lambda > v_c/2$. In either case, we have $u_e(a_h^*) > u_e(a_l^*)$, meaning that the implementability is always satisfied.

Second, if $\frac{v_a}{v_c} < \frac{\beta\theta_h - \theta_l}{2(1-\beta)}$, then a_h^* remains interior, but $a_l^* = 0$ if $\lambda < v_c/2$ and $a_l^* = 1$ if $\lambda > v_c/2$. In either case, the implementability $u_e(a_h^*) > u_e(a_l^*)$ is satisfied. The last step is to verify that no other deviation (implementation) can improve the platform's profit. The proof is very much similar to the proof of Proposition 1 and thus is omitted here. With all these results taken together, we can establish the first result of the proposition.

For the third result, with some algebra, we have

$$\frac{\partial a_l^*}{\partial \beta} = \frac{(2\lambda - v_c)(\theta_h - \theta_l)v_a}{\big((\theta_l - \beta\theta_h)v_c + 2\big(1-\beta\big)v_a\big)^2}.$$

Clearly, $\partial a_l^*/\partial\beta > 0$ if $\lambda > v_c/2$ and $\partial a_l^*/\partial\beta < 0$ if $\lambda < v_c/2$. Note that if $\lambda > v_c/2$, then the corner solution $a_l^* = 1$ occurs when $\beta > (\theta_l\lambda + v_a)/(\theta_h\lambda + v_a)$. If $\lambda < v_c/2$, the corner solution $a_l^* = 0$ applies when $\beta > (\theta_l(v_c - \lambda) + v_a)/(\theta_h(v_c - \lambda) + v_a)$. □

A.6. Proof of Proposition 3

A.6.1. The First Result: Ad Allocation. Comparing the profit-maximizing solution a_h^* and the welfare-maximizing solution $\hat{a}_h$, it is obvious that $a_h^* < \hat{a}_h$. From Lemma 1 and Proposition 2, we learn that $\hat{a}_l > \frac{v_c - \lambda}{v_c}$ and $a_l^* < \frac{v_c - \lambda}{v_c}$ if $\lambda < v_c/2$. Thus, $\hat{a}_l > a_l^*$ when $\lambda < v_c/2$. From the third result of Proposition 2, a_l^* increases with β until it reaches one. But $\hat{a}_l$ does not change with β. It then follows that $\hat{a}_l > a_l^*$ if $\beta < \beta'$ and $\hat{a}_l < a_l^*$ if $\beta > \beta'$, where β' is the point where $\hat{a}_l = a_l^*$.

A.6.2. The Second Result: Ad Externality. For the quality provision for high-type consumers, $u_e(a_h^*) > u_e(\hat{a}_h)$ holds when $\lambda > v_c/2$ because $u_e(a)$ is monotonically decreasing in a and $a_h^* < \hat{a}_h$. If $\lambda < v_c/2$, then $u_e(a_h^*) > u_e(\hat{a}_h)$ holds if and only if $a_h^* > \frac{2(v_c - \lambda)}{v_c} - \hat{a}_h$. With some algebra, this condition holds if $\lambda > \lambda_h$, where λ_h is defined in Equation (10). Note further that $\lambda_h < v_c/2$. Then $u_e(a_h^*) > u_e(\hat{a}_h)$ always holds if $\lambda > \lambda_h$ and $u_e(a_h^*) < u_e(\hat{a}_h)$ if $\lambda < \lambda_h$.

For the quality provision for low-type consumers, we also consider two cases. First, if $\lambda > v_c/2$, then $a_l^* > \frac{(v_c - \lambda)}{v_c}$ and $u_e(a_l^*)$ monotonically decreases with β, whereas $u_e(\hat{a}_l)$ is unaffected by β. The two equate at $\beta = \beta_l'$. It then follows that $u_e(a_l^*) > u_e(\hat{a}_l)$ if $\beta < \beta_l'$ and $u_e(a_l^*) < u_e(\hat{a}_l)$ if $\beta > \beta_l'$. Second, if $\lambda < v_c/2$, then $a_l^* < \frac{(v_c - \lambda)}{v_c}$, and $u_e(a_l^*)$ also monotonically decreases with β. Note that $u_e(a_l^*) > u_e(\hat{a}_l)$ holds if and only if $a_l^* > \frac{2(v_c - \lambda)}{v_c} - \hat{a}_l$. At $\beta = 0$,

$$a_l^*\big(\beta = 0\big) = \frac{\theta_l(v_c - \lambda) + v_a}{\theta_l v_c + 2v_a}\begin{cases} > \frac{2(v_c - \lambda)}{v_c} - \hat{a}_l, & \text{if } \lambda > \lambda_l, \\ < \frac{2(v_c - \lambda)}{v_c} - \hat{a}_l, & \text{if } \lambda < \lambda_l, \end{cases}$$

where λ_l is defined in Equation (10) and $\lambda_l < v_c/2$. At $\beta = 1$, $a_l^* = 0$, and thus $u_e(a_l^* = 0) = 0 < u_e(\hat{a}_l)$. It follows, then, that $u_e(a_l^*) < u_e(\hat{a}_l)$ for all β if $\lambda < \lambda_l$. If $\lambda > \lambda_l$, then there exists β_l'' such that $u_e(a_l^*) > u_e(\hat{a}_l)$ if $\beta < \beta_l''$ and $u_e(a_l^*) < u_e(\hat{a}_l)$ if $\beta > \beta_l''$. The results of the two cases, together with the fact that $\lambda_l < v_c/2$, establish the second result of the proposition. □

A.7. Proof of Proposition 4

I first derive the solutions of the three versioning policies in Sections A.7.1 and A.7.2. In Section A.7.3, I examine the conditions under which these policies can be implemented, establishing the results in Proposition 4.

A.7.1. Broad Reach With and Without Targeting. The derivation of the solution to the first two policies with broad reach follows the same procedure as in the baseline analysis. The difference between the two is that under targeting, the profit of reaching both consumer segments is $\beta v_{a,h} + (1-\beta)v_{a,l}$, which is greater than the profit under no targeting $\overline{v}_a$. Under the first policy, it is straightforward to follow the baseline analysis and obtain the interior solution

$$a_h^* = \frac{1}{2} - \frac{\theta_h\lambda}{2v_{a,h}}, \quad a_l^* = \frac{1}{2} - \frac{(\theta_l - \beta\theta_h)\lambda}{2\big(1-\beta\big)v_{a,l}}. \tag{A.2}$$

Similarly, following the same derivation, we can obtain the solution for the second policy:

$$a_h^* = \frac{1}{2} - \frac{\beta\theta_h\lambda}{2\big(\overline{v}_a - \big(1-\beta\big)v_{a,l}\big)}, \quad a_l^* = \frac{1}{2} - \frac{\big(\theta_l - \beta\theta_h\big)\lambda}{2\big(1-\beta\big)v_{a,l}}. \tag{A.3}$$

The third policy of narrow reach without targeting is derived next.

A.7.2. Narrow Reach Without Targeting. By restricting sales to high-type consumers only, the top advertisers can obtain an expected profit of $\beta v_{a,h}$, which is greater than the profit from advertising to the whole market $\overline{v}_a$ if $\beta > \overline{v}_a/v_{a,h}$. Lower-type advertisers, however, expect profit $(1-\beta)v_{a,l}$ from reaching low-type consumers. The indifferent advertisers are then $\sigma_h = (p_h^a - p_l^a)/(\beta v_{a,h} - (1-\beta)v_{a,l})$ and $\sigma_l = p_l^a/(1-\beta)v_{a,l}$. Inducing ad demands (a_l, a_h) implies the ad rates $p_l^a = (1 - a_h - a_l)(1-\beta)v_{a,l}$ and $p_h^a = p_l^a + (1 - a_h)(\beta v_{a,h} - (1-\beta)v_{a,l})$. The platform's problem becomes

$$\max_{a_l, a_h, p_l^c, p_h^c} \pi = \beta p_h^c + \big(1-\beta\big)p_l^c + (a_h + a_l)(1 - a_h - a_l)\big(1-\beta\big) \times v_{a,l} + a_h(1 - a_h)\big(\beta v_{a,h} - \big(1-\beta\big)v_{a,l}\big).$$

Let $a_{hl} = a_h + a_l$. By setting $p_l^c = \theta_l(q - \lambda a_l)$ and $p_h^c = p_l^c + \theta_h \lambda(a_l - a_h)$ and maximizing over (a_h, a_{hl}), we can obtain the first-order conditions

$$-(2\beta\theta_h - \theta_l)\lambda + (1 - 2a_h^*)(\beta v_{a,h} - (1-\beta)v_{a,l}) = 0 \text{ and}$$
$$-(\theta_l - \beta\theta_h)\lambda + (1 - 2a_{hl}^*)(1-\beta)v_{a,l} = 0.$$

Together with the relationship $a_l^* = a_{hl}^* - a_h^*$, these conditions lead to the solution

$$\begin{aligned} a_h^* &= \frac{1}{2} - \frac{(2\beta\theta_h - \theta_l)\lambda}{2(\beta v_{a,h} - (1-\beta)v_{a,l})}, \\ a_l^* &= \frac{(2\beta\theta_h - \theta_l)\lambda}{2(\beta v_{a,h} - (1-\beta)v_{a,l})} - \frac{(\theta_l - \beta\theta_h)\lambda}{2(1-\beta)v_{a,l}}. \end{aligned} \tag{A.4}$$

There are two implementability constraints. First, we need $a_h^* < a_{hl}^*$ for the solution to exist. This condition is satisfied as long as $\alpha > v_{a,h}/v_{a,l}$. Second, to ensure that low-type consumers see more ads than high-type consumers, we need that $a_l^* > a_h^*$ or, equivalently, $a_{hl}^* > 2a_h^*$, which is a stronger condition than the first condition and is satisfied as long as

$$v_{a,h} < \frac{(1-\beta)((1-\beta)v_{a,l} + (3\beta\theta_h - \theta_l)\lambda)v_{a,l}}{\beta((1-\beta)v_{a,l} + (\theta_l - \beta\theta_h)\lambda)} \equiv v_h''.$$

A.7.3. Conditions for Versioning. If the first two policies are both feasible, then it is optimal for the platform to choose the first one (broad reach with targeting) because it can improve the profit from advertising. The question is when targeted advertising is feasible. Note that with targeting, implementability requires that $a_l^* > a_h^*$ because $u_e'(a) < 0$. It follows that

$$a_l^* > a_h^* \quad \Leftrightarrow \quad \frac{v_{a,l}}{v_{a,h}} > \frac{\theta_l - \beta\theta_h}{(1-\beta)\theta_h},$$

which always holds if $\alpha\beta > 1$. If $\alpha\beta < 1$, then the condition is equivalent to $v_{a,h} < \frac{(1-\beta)\theta_h v_{a,l}}{(\theta_l - \beta\theta_h)} \equiv v_h'$. Similarly, broad reach without targeting can be implemented only if $a_l^* > a_h^*$, which always holds if $\alpha\beta > 1$. If $\alpha\beta < 1$, then the condition is equivalent to $\overline{v}_a < v_h'$. Note that because $\overline{v}_a < v_{a,h}$, it is still feasible to implement versioning without targeting if $v_{a,h} > v_h'$. In the latter case, when both the second and third policies are feasible, which occur when $v_{a,h} < v_h''$ and $\overline{v}_a < v_h'$, then the second one is optimal as long as $\beta < \overline{v}_a/v_{a,h}$. In the case where $v_{a,h} > \max\{v_h', v_h''\}$ and $\overline{v}_a > v_h'$, none of the implementability conditions is satisfied, and thus no versioning policy is feasible. □

A.8. Proof of Proposition 5

I first derive the implementation of versioning and validate its optimality in Section A.8.1. In Section A.9.2, I then evaluate when the proposed implementation can hold to establish the first statement of Proposition 5. Section A.8.3 proves the second statement.

A.8.1. Implementing Versioning. Given ad rates (p_h^a, p_l^a), the marginal advertisers are $\sigma_h = p_h^a/v_a$ and $\sigma_l = p_l^a/v_a$, whose optimal strategy, if they advertise, is to send exactly zero messages (i.e., $m_h^*(\sigma_h) = 0$ and $m_l^*(\sigma_l) = 0$). Then the total ad intensity (a_h, a_l) for each segment can be derived as follows:

$$\begin{aligned} a_h &= \int_{\sigma_h}^1 m_h^*(p_h^a;\sigma)d\sigma = \int_{p_h^a/v_a}^1 \beta \ln\frac{\sigma p_h^a}{v_a} d\sigma \\ &= \beta\left(\frac{p_h^a}{v_a} - \ln\frac{p_h^a}{v_a} - 1\right); \end{aligned} \tag{A.5}$$

$$\begin{aligned} a_l &= \int_{\sigma_l}^1 m_l^*(p_l^a;\sigma)d\sigma = \int_{p_l^a/v_a}^1 (1-\beta) \ln\frac{\sigma p_l^a}{v_a} d\sigma \\ &= (1-\beta)\left(\frac{p_l^a}{v_a} - \ln\frac{p_l^a}{v_a} - 1\right). \end{aligned} \tag{A.6}$$

Define the normalized ad intensities, $\tilde{a}_h = a_h/\beta$ and $\tilde{a}_l = a_l/(1-\beta)$. The market-clearing ad intensities in Equations (A.5), and (A.6) can be rewritten in the form of $\tilde{a} = \frac{p^a}{v_a} - \ln\frac{p^a}{v_a} - 1 \equiv h(p^a)$. We can rewrite the ad rate as a function of ad intensity $p_h^a = h^{-1}(\tilde{a}_h)$ and $p_l^a = h^{-1}(\tilde{a}_l)$, substitute the prices p_h^c and p_l^c using the incentive constraints, and rewrite the platform's problem with $(\tilde{a}_h, \tilde{a}_l)$ as choice variables

$$\begin{aligned} \max_{\tilde{a}_h,\tilde{a}_l} \quad \pi &= \beta\theta_h(q - \lambda\beta\tilde{a}_h) + (\theta_l - \beta\theta_h)(q - \lambda(1-\beta)\tilde{a}_l) \\ &\quad + \beta\tilde{a}_h h^{-1}(\tilde{a}_h) + (1-\beta)\tilde{a}_l h^{-1}(\tilde{a}_l). \end{aligned}$$

The first-order conditions are

$$\begin{aligned} -\beta\theta_h\lambda + \tilde{a}_h(h^{-1})'(\tilde{a}_h) + h^{-1}(\tilde{a}_h) &= 0 \text{ and} \\ (\beta\theta_h - \theta_l)\lambda + \tilde{a}_l(h^{-1})'(\tilde{a}_l) + h^{-1}(\tilde{a}_l) &= 0, \end{aligned} \tag{A.7}$$

which are necessary for the interior solution $(\tilde{a}_h^*, \tilde{a}_l^*)$. The corner solutions $\tilde{a}_h^* = 0$ and $\tilde{a}_l^* = 0$ may apply if $v_a < \beta\theta_h\lambda$ and $v_a < (\theta_l - \beta\theta_h)\lambda$. It remains to check whether the solutions from the first-order conditions are unique and indeed maximize the profit. However, the function h^{-1} is hard to evaluate. It would be easier to work with the equivalent problem that maximizes over the ad rates (p_h^a, p_l^a), which are monotone transformations of the ad intensities $(\tilde{a}_h, \tilde{a}_l)$. The objective function can be rewritten as an additive sum

$$\begin{aligned} \max_{p_l^a, p_h^a} \pi = {} & \underbrace{\beta\theta_h\left(q - \lambda\beta\left(\frac{p_h^a}{v_a} - \ln\frac{p_h^a}{v_a} - 1\right)\right) + \beta\left(\frac{p_h^a}{v_a} - \ln\frac{p_h^a}{v_a} - 1\right)p_h^a}_{\pi_h(p_h^a)} \\ & + \underbrace{\begin{matrix} (\theta_l - \beta\theta_h)\left(q - \lambda(1-\beta)\left(\frac{p_l^a}{v_a} - \ln\frac{p_l^a}{v_a} - 1\right)\right) \\ +(1-\beta)\left(\frac{p_l^a}{v_a} - \ln\frac{p_l^a}{v_a} - 1\right)p_l^a \end{matrix}}_{\pi_l(p_l^a)} \end{aligned}$$

subject to the constraints $p_h^a \le v_a$ and $p_l^a \le v_a$. Let (p_l^{a*}, p_h^{a*}) denote the solution to the first-order conditions $\partial\pi_h/\partial p_h^a = 0$ and $\partial\pi_l/\partial p_l^a = 0$. The second derivative for π_h is

$$\frac{\partial^2\pi_h}{\partial p_h^{a2}} = \beta\left(\frac{2}{v_a} - \frac{1}{p_h^a} - \frac{\lambda\beta\theta_h}{p_h^{a2}}\right),$$

which is negative if $p_h^a \in (0, \overline{p})$ and positive if $p_h^a \in (\overline{p}, v_a)$, where $\overline{p} = 2\lambda\beta\theta_h/(\sqrt{1 + 8\lambda\beta\theta_h/v_a} - 1)$ (one can verify that $\overline{p} < v_a$ using the condition for interior solution $v_a > \lambda\beta\theta_h$). Thus, π_h is concave when $p_h^a < \overline{p}$ but becomes convex when $p_h^a > \overline{p}$. Instead, π_h first increases (because $\partial\pi_h/\partial p_h^a > 0$ for small values of p_h^a), reaches maximum at p_h^{a*}, and then decreases until $\bar{p}$. After this point, the decrease in π_h starts to slow down and might possibly increase at some point because of convexity. Optimality requires that at the highest value of p_h^a (i.e., $p_h^a = v_a$), the profit does not improve over the optimum defined by the first-order condition. This is indeed the case because $\pi_h(p_h^a = v_a) = 0$. Hence, p_h^{a*} is the unique optimal solution of π_h. By the same argument, p_l^{a*} is also the unique optimal solution of π_l. Therefore, equivalently, (a_h^*, a_l^*) defined by the first-order conditions in Equation (A.7) is the unique optimal solution.

A.8.2. First Statement. I now evaluate the implementability condition $a_h^* < a_l^*$ that low-type consumers see more ads than high-type consumers using the optimal solution $(\tilde{a}_h^*, \tilde{a}_l^*)$. Define $G(\tilde{a}) \equiv \tilde{a}(h^{-1})'(\tilde{a}) + h^{-1}(\tilde{a})$. Then the first-order conditions in Equation (A.7) can be rewritten as $G(\tilde{a}_h) = \beta\theta_h\lambda$ and $G(\tilde{a}_l) = (\theta_l - \beta\theta_h)\lambda$. Note that $(h^{-1})'(\tilde{a}) = 1/h'(h^{-1}(\tilde{a})) = v_a h^{-1}(\tilde{a})/(h^{-1}(\tilde{a}) - v_a)$ and

$$\begin{aligned} G'(\tilde{a}) &= 2\left(h^{-1}\right)'(\tilde{a}) + \tilde{a}\left(h^{-1}\right)''(\tilde{a}) \\ &= \frac{2}{h'(h^{-1}(\tilde{a}))} - \frac{h''\left(h^{-1}(\tilde{a})\right)}{(h'(h^{-1}(\tilde{a})))^3} \\ &= \underbrace{\frac{v_a h^{-1}(\tilde{a})}{v_a - h^{-1}(\tilde{a})}}_{>0}\left[\frac{\tilde{a}v_a^2}{(h^{-1}(\tilde{a}))^2} - 2\right], \end{aligned}$$

which is negative as long as $\tilde{a} < \tilde{a}_0$ for some positive threshold $\tilde{a}_0$. Thus, $G(\tilde{a})$ first decreases in $\tilde{a}$ when $\tilde{a} < \tilde{a}_0$ and then increases again when $\tilde{a} > \tilde{a}_0$. Because the optimal solutions a_h^* and a_l^* are found by intersecting the line $\beta\theta_h\lambda$ with the curve $G(\tilde{a}_h)$ and the line $(\theta_l - \beta\theta_h)\lambda$ with the curve $G(\tilde{a}_l)$, the intersections must occur when $\tilde{a} < \tilde{a}_0$ so that $G(\cdot)$ is monotonically decreasing. Otherwise, there will be two intersections that satisfy the first-order conditions, violating the uniqueness according to Section A.8.1. This observation will be useful when comparing a_h^* and a_l^* later. First, if $\alpha\beta > 1/2$, then $\beta\theta_h > (\theta_l - \beta\theta_h)$, implying that

$$\tilde{a}_h^* < \tilde{a}_l^* \Leftrightarrow \frac{a_h^*}{\beta} < \frac{a_l^*}{1-\beta} \Leftrightarrow \frac{a_h^*}{a_l^*} < \frac{\beta}{1-\beta}.$$

If $\beta < 1/2$, then the last inequality implies that $\frac{a_h^*}{a_l^*} < \frac{\beta}{1-\beta} < 1$, and thus, $a_h^* < a_l^*$ holds. However, this may not hold if $\beta > 1/2$ because $\frac{\beta}{1-\beta} > 1$. Because $G(\tilde{a})$ is decreasing in $\tilde{a}$, fixing θ_l, $\tilde{a}_h^*$ decreases with α, whereas $\tilde{a}_l^*$ increases with α. This implies that, for a given β, the difference in ad intensity $a_l^* - a_h^*$ becomes larger as α increases. Then, for sufficiently small α, we have $a_h^* > a_l^*$, meaning that versioning is not implementable. Second, if $\alpha\beta < 1/2$, we have $\tilde{a}_h^* > \tilde{a}_l^*$, implying that $\frac{a_h^*}{a_l^*} > \frac{\beta}{1-\beta}$. If $\beta > 1/2$, then it follows that $\frac{a_h^*}{a_l^*} > \frac{\beta}{1-\beta} > 1$, violating the implementability condition. If $\beta < 1/2$ instead, then it is still possible that the implementability can break down as long as α is sufficiently small. This is again because $\tilde{a}_h^*$ decreases with α, whereas $\tilde{a}_l^*$ increases with α. Hence, $a_h^* - a_l^*$ decreases with α. For sufficiently small α, a_h^* can be larger than a_l^*. The results under both cases together can establish the first result that implementability can be violated when β is sufficiently large and α is sufficiently small. A sufficient condition is that $\beta > 1/2$ and $\alpha < 1/2\beta$.

A.8.3. Second Statement. Note that if $\tilde{a}_h < \tilde{a}_l$, then $p_h^a > p_l^a$ because $p^a = h^{-1}(\tilde{a})$ decreases in $\tilde{a}$. Using the market-clearing conditions $\sigma_h = p_h^a/v_a$ and $\sigma_l = p_l^a/v_a$, it follows that $\sigma_h > \sigma_l$. Thus, there are fewer advertisers reaching high-type consumers. The conclusion is reversed if $\tilde{a}_h > \tilde{a}_l$. Because $\tilde{a}_h^* < \tilde{a}_l^*$ if $\alpha\beta > 1/2$ and $\tilde{a}_h^* > \tilde{a}_l^*$ if otherwise, the second statement then follows. □

A.9. Proof of Lemma 3

If the platform sells q_H to both types, the optimal price is $p^c = \theta_l q_H$ so that the low type is just willing to buy, leading to profit $\pi^M = \theta_l q_H$. If the platform discriminates, then the optimal price charged for type θ_l is $p_l^c = \theta_l q_L$, whereas the price for type θ_h is $p_h^c = p_l^c + \theta_h(q_H - q_L)$. Profit in this case becomes $\pi^V = \theta_l q_L + \beta\theta_h(q_H - q_L)$. If the platform sells q_H only to the high-type segment, then the optimal price is $p^c = \theta_h q_H$ with profit $\pi^P = \beta\theta_h q_H$. It follows that if $\alpha\beta < 1$, then $\pi^M > \pi^V > \pi^P$. Conversely, if $\alpha\beta > 1$, then $\pi^P > \pi^V > \pi^M$. □

A.10. Proof of Proposition 6

First note that whenever the media platform sells only one product, the optimal strategy is to sell q_H instead of q_L. This is because the platform can extract more surplus by replacing a low-quality product with a high-quality one with a slightly higher price. Then there are four remaining strategies to evaluate: selling product q_H to both segments (henceforth, mass market), selling product q_H to the high-type segment only (henceforth, premium market), versioning with product q_H only and ad differentiation, and versioning with both q_H and q_L together with ad differentiation. Clearly, the first three are exactly the same as those analyzed in the main model by replacing q with q_H.

A.10.1. Mass-Market Strategy. With this strategy, the indifferent advertiser who is just willing to advertise to both segments is determined by $\sigma_0 = p_{hl}^a/v_a$. To sell a total ad quantity of a, the ad rate should be set at $p_{hl}^a = (1-a)v_a$. Charging the low-type segment the reservation price, the platform's problem is

$$\max_a \quad \pi = \theta_l(q_H - \lambda a) + a(1-a)v_a.$$

From the first-order condition, we obtain $a^{M*} = \frac{1}{2} - \frac{\theta_l\lambda}{2v_a}$. Let π^{M*} denote the resulting profit.

A.10.2. Premium-Market Strategy. With this strategy, the indifferent advertiser who is just willing to advertise to both segments is determined by $\sigma_0 = p_h^a/\beta v_a$. To sell a total ad quantity of a, the ad rate should be set at $p_h^a = (1-a)\beta v_a$. Charging the high-type segment the reservation price, the platform's problem is

$$\max_a \quad \pi = \beta\theta_h(q_H - \lambda a) + a(1-a)\beta v_a.$$

From the first-order condition, we obtain $a^{M*} = \frac{1}{2} - \frac{\theta_h\lambda}{2v_a}$, which is the same as the high-type segment's ad allocation under versioning. Let π^{P*} denote the resulting profit.

A.10.3. Comparison. The last strategy involves both product and ad differentiations. However, the advertising revenue remains the same as the case where only ad differentiation is involved. Therefore, the optimal advertising allocation does not change. The only difference is that now the optimal content prices are $p_l^c = \theta_l(q_L - \lambda a_l)$ and $p_h^c = p_l^c + \theta_h(q_H - q_L + \lambda(a_l - a_h))$. Let the superscript $V2$ denote the versioning strategy with both ad and product differentiations. The difference in the impact of β on the equilibrium profit is

$$\frac{d\pi^{V2*}}{d\beta} - \frac{d\pi^{V*}}{d\beta} = \theta_h(q_H - q_L) > 0.$$

Note further that $\pi^{V2*} - \pi^{V*} = (\beta\theta_h - \theta_l)(q_H - q_L)$, which becomes negative as $\beta \to 0$, positive as $\beta \to 1$, and equals zero at $\beta = 1/\alpha$.

Note also that $d\pi^{V*}/d\beta > 0$ and $d\pi^{M*}/d\beta = 0$. In addition, as $\beta \to 0$, $\pi^{V*} = \pi^{M*}$. Hence, $\pi^{V*} > \pi^{M*}$ for all β. It then remains to compare the premium-market strategy with versioning with a product line when $\beta > 1/\alpha$. Note that

$$\frac{d\pi^{P*}}{d\beta} - \frac{d\pi^{V2*}}{d\beta} = \theta_h(q_L - \lambda a_l) + a_l^{V*}(1 - a_l^{V*})v_a > 0.$$

At $\beta \to 1$, $\pi^{P*} - \pi^{V2*} = (\theta_h - \theta_l)(q_L - \lambda a_l) > 0$. Thus, there exists some cutoff $\bar{\beta} > 1/\alpha$ such that $\pi^{P*} < \pi^{V2*}$ if $\beta < \bar{\beta}$ and $\pi^{P*} > \pi^{V2*}$ if otherwise. □

Endnotes

[1] An informal interview with a senior manager at a leading Chinese online video platform validates this concern. As increasingly more consumers have joined the platform's VIP membership service, which offers an ad-free experience, its advertising business has been shrinking; the platform now faces pressure from advertisers, especially high-profile ones like Procter & Gamble and Mercedes-Benz, to display at least some ads to VIP members.

[2] A monopoly setup permits a tractable analysis that informs the study questions and approximately reflects the position of real-world platforms, such as YouTube, in their markets.

[3] In reality, it is not uncommon for consumers to like ads because they have information or entertainment value (e.g., it is fun to watch a creative commercial). Information value has been incorporated into the analysis, whereas entertainment value is not examined here.

[4] Building on a structural model of television markets, Wilbur (2008) conducts a counterfactual experiment showing that ad-avoidance technologies can increase advertising quantities and reduce media revenues.

[5] In 2015, in an effort to mitigate the increasing use of ad-blockers, Google launched a program called Google Contributor that allows users to pay a monthly fee to avoid ads in its network of content sites.

[6] There are examples in which a premium version still allows for limited ads. For example, CBS Broadcasting's All Access streaming service offers a premium package that delivers fewer ads at a higher subscription fee. In my analysis, a completely ad-free version can endogenously arise in the boundary solution under certain conditions.

[7] An equivalent interpretation is that consumers differ in their marginal disutility of seeing an ad.

[8] The main conclusions do not change qualitatively if we relax this assumption by allowing that higher-type advertisers annoy consumers less. A formal analysis under a general distribution of adviser type and type-dependent nuisance cost is available upon request.

[9] There are, of course, other purposes of advertising, such as signaling product quality. For a comprehensive review, see Bagwell (2007).

[10] Although the assumption that consumers observe the ad rates is commonly made in the literature of two-sided analysis of media markets (e.g., Anderson and Coate 2005), it may appear strong. An alternative approach is to assume that the platform can announce the allocation plan (Σ_h, Σ_l) induced by the ad rates and that there is a high exogenous "reputation" cost if it misinforms consumers about the allocation outcome. This approach would lead to the equivalent results.

[11] Anderson and Coate (2005), for example, assume that each advertiser informs consumers of a new product that has no competitor and thus is able to charge a monopoly price. Consumers expect zero surplus from the product purchase.

[12] Kaiser and Song (2009), for example, find that there is little evidence for readers disliking advertising in print media markets.

[13] Even if consumers are continuously distributed, there is no incentive to price discriminate. Suppose that consumer type θ is continuous and uniformly distributed within $[0, 1]$. All other assumptions remain unchanged. Assuming that the platform can segment consumers so that the higher types $\theta \in [\theta_H, 1]$ buy q_H, whereas the lower types $\theta \in [\theta_L, \theta_H)$ buy q_L. The optimal prices will be $p_L = q_L/2$ and $p_H = p_L + (q_H - q_L)/2$. However, the solution implies $\theta_L = \theta_H = 1/2$, a contradiction.

[14] If the consumer's value function is given by $u(q, \theta) = \theta V(q)$ and thus the social surplus is $s(q, \theta) = \theta V(q) - c$, then $s(q, \theta)$ satisfies log submodularity because $s_{q\theta} \cdot s - s_q \cdot s_\theta = -cV_q(q) \leq 0$.

[15] This conclusion is in part driven by the two-type setup on the consumer side, which is not able to capture the market expansion effect. If there are more consumer types and the market is not fully covered, then versioning can still dominate because introducing a lower-quality version can expand the market.

References

Ambrus A, Calvano E, Reisinger M (2016) Either or both competition: A" two-sided" theory of advertising with overlapping viewerships. *Amer. Econom. J. Microeconom.* 8(3):189–222.

Anderson ET, Dana JD (2009) When is price discrimination profitable? *Management Sci.* 55(6):980–989.

Anderson SP, Coate S (2005) Market provision of broadcasting: A welfare analysis. *Rev. Econom. Stud.* 72(4):947–972.

Anderson SP, Gans JS (2011) Platform siphoning: Ad-avoidance and media content. *Amer. Econom. J. Microeconom.* 3(4):1–34.

Armstrong M (2006) Competition in two-sided markets. *RAND J. Econom.* 37(3):668–691.

Athey S, Calvano E, Gans JS (2016) The impact of consumer multihoming on advertising markets and media competition. *Management Sci.* 64(4):1477–1973.

Bagwell K (2007) The economic analysis of advertising. Armstrong M, Porter R, eds. *Handbook of Industrial Organization*, vol. 3 (Elsevier, Amsterdam), 1701–1844.

Bergemann D, Bonatti A (2011) Targeting in advertising markets: implications for offline vs. online media. *RAND J. Econom.* 42(3): 417–443.

Butters GR (1977) Equilibrium distributions of sales and advertising prices. *Rev. Econom. Stud.* 44(3):465–491.

Dukes A (2004) The advertising market in a product oligopoly. *J. Indust. Econom.* 52(3):327–348.

Gal-Or E, Dukes A (2003) Minimum differentiation in commercial media markets. *J. Econom. Management Strategy* 12(3):291–325.

Godes D, Ofek E, Sarvary M (2009) Content vs. advertising: The impact of competition on media firm strategy. *Marketing Sci.* 28(1):20–35.

Gomes R, Pavan A (2016) Many-to-many matching and price discrimination. *Theoret. Econom.* 11(3):1005–1052.

Jeon D-S, Kim B-C, Menicucci D (2016). Second-degree price discrimination by a two-sided monopoly platform. Working paper, Toulouse School of Economics, Toulouse, France.

Johnson JP (2013) Targeted advertising and advertising avoidance. *RAND J. Econom.* 44(1):128–144.

Kaiser U, Song M (2009) Do media consumers really dislike advertising? an empirical assessment of the role of advertising in print media markets. *Internat. J. Indust. Organ.* 27(2):292–301.

Kind HJ, Nilssen T, Sørgard L (2009) Business models for media firms: Does competition matter for how they raise revenue? *Marketing Sci.* 28(6):1112–1128.

Liu Q, Serfes K (2013) Price discrimination in two-sided markets. *J. Econom. Management Strategy* 22(4):768–786.

Maskin E, Riley J (1984) Monopoly with incomplete information. *RAND J. Econom.* 15(2):171–196.

Mussa M, Rosen S (1978) Monopoly and product quality. *J. Econom. Theory* 18(2):301–317.

Peitz M, Valletti TM (2008) Content and advertising in the media: Pay-tv vs. free-to-air. *Internat. J. Indust. Organ.* 26(4):949–965.

Prasad A, Mahajan V, Bronnenberg B (2003) Advertising vs. pay-per-view in electronic media. *Internat. J. Res. Marketing* 20(1):13–30.

Reisinger M (2012) Platform competition for advertisers and users in media markets. *Internat. J. Indust. Organ.* 30(2):243–252.

Rochet J-C, Tirole J (2006) Two-sided markets: a progress report. *RAND J. Econom.* 37(3):645–667.

Tåg J (2009) Paying to remove advertisements. *Inform. Econom. Policy* 21(4):245–252.

Weyl EG (2010) A price theory of multi-sided platforms. *Amer. Econom. Rev.* 100(4):1642–1672.

Wilbur KC (2008) A two-sided, empirical model of television advertising and viewing markets. *Marketing Sci.* 27(3):356–378.

Wu C (2015) Matching value and market design in online advertising networks: An empirical analysis. *Marketing Sci.* 34(6): 906–921.