Online Business Models, Digital Ads, and User Welfare*

Daron Acemoglu[†] Daniel Huttenlocher[‡] Asuman Ozdaglar[§] James Siderius[¶]

October 1, 2025

Abstract

We present a model where online platforms, situated between firms and users, offer plans that intermix entertaining content with digital advertising ("ads") from firms. Users derive utility from entertainment and learn about their valuation for the firms' products from ads. While some users are fully rational, others naïvely perceive digital ads as more informative than they actually are. We characterize the profit-maximizing business model of the platform and show that welfare is lower when the platform monetizes through advertising instead of subscription both for naïfs (because they are targeted by intense digital advertising, which makes them over-optimistic about product quality and over-purchase the product) and for sophisticates (because the inflated demand from naïfs increases the firm's profit-maximizing price). This negative welfare effect is intensified when the platform can offer mixed business models that separate naïve and sophisticated users into different plans. Our results are robust to firm-level and platform-level competition, because digital ads soften competition between both firms and platforms. We also show how digital ad taxes can improve welfare.

JEL Classification: D83, D43, L13.

Keywords: digital advertising, welfare, online business models, digital ad taxation

*We are grateful to numerous participants at the 2024 NBER Economics of Artificial Intelligence Conference, INFORMS 2022 Annual Conference, UT Dallas Jindal School of Business, Santa Clara Leavey School of Business, Johns Hopkins Carey School of Business, University of Michigan Ross School of Business, Tuck School of Business at Dartmouth College, USC Marshall School of Business, Emory University, Chicago Booth School of Business, MIT Sloan School of Management, Columbia NYC Media Seminar, and AsuFest 2023 for their suggestions and feedback.

[†]Massachusetts Institute of Technology, NBER, and CEPR, daron@mit.edu

[‡]Massachusetts Institute of Technology, huttenlocher@mit.edu

[§]Massachusetts Institute of Technology, asuman@mit.edu

[¶]Tuck School of Business at Dartmouth College, james.siderius@tuck.dartmouth.edu

1 Introduction

Online platforms have become the dominant medium for entertainment and social interactions, overtaking traditional sources such as TV and print media (see Cinelli et al. (2020) and Sherman and Waterman (2016)). The average adult now spends over three hours a day on social media,[1] and even more time on streaming services such as Netflix, YouTube, and Hulu (Budzinski et al. (2021), Richter (2019), and Twenge et al. (2019)). There is an active debate on the costs and benefits of social media engagement, with many of the potential negative effects being related to targeted digital advertisement and its impacts on user beliefs and behavior (e.g., see Marwick and Lewis (2017), Allcott et al. (2020), and Allcott et al. (2022)).

The majority of social media platforms generate their revenue from digital advertising ("digital ads" for short).[2] Unlike traditional advertising, where the same product recommendation is broadcast to a large audience, digital advertising allows ads to be tailored and targeted to different users. Concretely, the targeted ads we have in mind include digital ads based on past browsing, app-install and free-trial offers in video and social feeds, and promotions for novelty products (e.g., personal care products or clothing). While this targeting may make such ads more informative about relevant products and services, it also opens the way to greater manipulation and enticement for other users (Bennett and Gordon (2020), Deng and Mela (2018), and De Jans et al. (2019)). Long-standing critiques dating back to Galbraith (1958) argue that advertising can inefficiently amplify demand, while Nelson (1974) emphasizes the informational role of ads in signaling quality, lowering search costs, and improving matches. Despite growing concerns on these topics, there is currently no framework in which digital ads have both informative and manipulative roles. There are also only very few analyses of online business models.

In this paper, we develop a parsimonious model where an online media platform offers both entertainment and digital ads and acts as a two-sided marketplace bringing together users that can learn from informative ads and a firm interested in advertising to users. The platform can monetize its services via advertising, subscription fees, or both. Digital ads are informative about (user-specific) quality of a product. Ads are therefore beneficial because they provide informative signals, but are also costly as they interrupt the entertaining content.

A distinguishing feature of our model is that there are both *sophisticated* users who have the correct model about the relationship between good signals from digital ads and product quality, and *naïve* users who have a misspecified model. Specifically, naïve users underestimate the likelihood of "false positives," whereby a product that is low-quality for them may nonetheless generate a positive signal via ads. This may be because of their inherent naïveté or because they underestimate the degree to which the targeting of digital ads may exaggerate the appeal of the underlying product to them. As

[1] See https://www.forbes.com/sites/petersuciu/2021/06/24/americans-spent-more-than-1300-hours-on-social-media/ and https://whatagraph.com/blog/articles/how-much-time-do-people-spend-on-social-media.

[2] For example, digital ads made up 98% of Facebook's revenue from 2017-2019 (see https://www.nasdaq.com/articles/what-facebooks-revenue-breakdown-2019-03-28-0) and about 85% of YouTube's revenue in 2020 (despite its premium ad-free subscription plan, see https://spendmenot.com/blog/youtube-revenue-statistics/). Ad-supported plans are now a large slice of video streaming services where both subscription and ad-based plans co-exist: as of 2025, 46% of subscriptions among dual-tier streamers were ad-supported, see https://www.antenna.live/insights/antenna-q225-state-of-subscriptions-report-adds-and-ads.

a special case of our framework, naïve agents may believe a "good" product signal is always perfectly informative of a good product, even though false-positives may exist. This misspecification on the part of naïve agents opens the way to manipulation—it is profitable for the firm and hence for the platform to send more ads to naïve users to boost their demand for the product.

For expositional clarity, we first restrict the platform to two simple business models: a free-of-charge advertising-based plan or an ad-free plan with a subscription fee. Our first main result (Proposition 2) is a striking one: provided that naïve users do not have a model very close to that of sophisticated users, the unique equilibrium involves an advertising-based plan designed for naïfs, which consequently fully segments the market, and sophisticates are excluded from the platform. This is because the platform chooses a high level of "ad load" (high ad intensity), which is unattractive for sophisticates and fully extracts all surplus from naïfs. As a result, the *ex post* welfare of naïve users is even less than the benchmark without the platform, as they end up *de facto* manipulated into over-consuming the product.[3] Interestingly, sophisticated users also have lower welfare than the scenario without the platform, because when digital ads inflate the demand from naïve agents, the firm prefers to charge a higher price, which sophisticates also pay.

An important comparative static is that the separating equilibrium with digital ads targeting naïve agents is more likely when the likelihood of false positive signals for naïfs is higher—implying that digital ads emerge precisely when they are *more* misleading. Intuitively, it is this misleading aspect of digital advertisement that makes them expand the demand for the product. This is also the feature that makes our welfare results nuanced: absent naïfs, informative ads are not necessarily over-provided, but as the false-positive rate or the fraction of naïve users increases, the platform re-optimizes toward menus with a higher ad load, and the firm sets a higher price. In consequence, aggregate welfare falls, even though ads retain informational value.

The results from this simplified model with just two business plans generalize directly when the platform can offer multiple entertainment plans that intermix advertising and subscription. The equilibrium typically separates naïfs and sophisticates, but this time the platform can also extract surplus from sophisticates through a subscription fee. Welfare effects are similar to our baseline result, though with additional implications. Specifically, the separating equilibrium emerges when false positives from the ad technology are not too rare (otherwise the equilibrium is pooling). Equilibrium user welfare is decreasing in the likelihood of false positives (which implies greater distortion in naïve agents' assessment of product quality) and is also decreasing in the overall informativeness of ads (because more informative ads make a separating equilibrium, which is worse for naïfs and features higher prices, more likely).

The insights from our analysis with a single platform and a single advertising firm generalize to an environment with multiple platforms and multiple firms. The fundamental reason for this is that digital ads soften the competition between both firms and platforms. For example, two firms with identical products that would otherwise engage in Bertrand competition and earn zero profits now gain market power, because users who obtain different information from the ads they see will have

[3]We refer to the use of these ads as "*de facto* manipulation," because they may not have been designed to maximize manipulation, but their use still has a manipulative effect on naïve users. We discuss later the choice of the degree of manipulation.

different (derived) willingness to pay for the products of the two firms. More generally, we show that digital ads soften firm-level competition, as they enable endogenous differentiation of products based on the signals that users receive about their quality. Consequently, provided that both false positive and true positive signals from the advertisement technology are sufficiently likely, the equilibrium is again separating and a high ad-load plan targets naïve users, while sophisticates are charged a subscription fee. As before, the separating equilibrium features higher markups and lower welfare for both types of users.

Platform-level competition also has nuanced effects on user welfare for similar reasons. All else equal, competition between platforms could reduce surplus extraction from users. Nevertheless, this offset is incomplete because naïve users overvalue ads, as they consider them to be more informative than they actually are. Consequently, when ads are sufficiently informative, we obtain a separating equilibrium where naïve agents are targeted by frequent digital ads and welfare is low, despite between-platform competition.

The pervasive market failures and the equilibrium choice of online business models that are *de facto* manipulative for naïve agents raise the question of whether feasible regulatory policies might improve welfare. We show that the first best (where a social planner controls the full allocation without any incentive compatibility constraints on the side of users) is in general not achievable, but the second best (where the social planner is subject to the "self-selection" or incentive-compatibility constraints of different user types) can be decentralized using nonlinear taxes and subsidies. We also show that linear taxes on digital ad revenues can improve consumer welfare (see Romer (2021) and Acemoglu and Johnson (2024) for other arguments in favor of digital ad taxes).

Our results are robust to a number of extensions that also underscore the richness of our model. First, we endogenize "manipulation" by letting the platform choose ad persuasiveness, which we show strengthens platform incentives to target naïfs and leads to reductions in user welfare. Second, we add naïve learning dynamics: exposure to disconfirming feedback gradually raises sophistication, makes the ad-heavy business model less likely, and attenuates product price distortions, potentially mitigating (but not overturning) the main driver of welfare loss in our baseline analysis. Third, we allow limited downstream price discrimination, under which firms sometimes observe sophistication types; prices tilt toward type-specific levels, ad-intensive menus become more likely, and naïfs are strictly worse off (as are sophisticates, often). Fourth, we also introduce a price-sensitivity dimension and show that the platform can profitably run a 2×2 menu that still separates by sophistication while preserving our main comparative statics (and that it is the distinction between sophisticates and naïfs that is at the root of our results).

Related Literature. This paper relates to several strands of work. The first strand is the industrial organization literature on informational advertising (e.g., see Nelson (1974), Tirole (1988), Grossman and Shapiro (1984) and Dixit and Norman (1978)). Most closely related to our work is Meurer and Stahl (1994), which builds on the seminal paper of Butters (1977), to construct a model in which advertising is informative about partially substitutable, horizontally-differentiated products, and consumers use the information in the ads to decide which products to buy. We differ from this paper and from all others in this literature in three important ways. First, a platform is situated in-between the users and the ads, and users make active decisions influencing how many ads they will see via their choice of plan. This type of platform intermediation makes us more closely related to the work on two-sided marketplaces (e.g., Rochet and Tirole (2006), De Reuver et al. (2018)). Second, because of the presence of naïfs, ads in our setup are simultaneously informative and (*de facto*) manipulative. This dual role of ads and the interaction between naïfs and sophisticates are at the root of all of our results. Lastly, there is no analogue of platforms' choices over business models in this literature, which is critical for our results about separating naïve and sophisticated users and the negative welfare consequences of digital ads.

The second related literature focuses on deceptive and manipulative advertising (e.g., Danciu et al. (2014) and Eyal (2014)), and is also connected to the work on behavioral manipulation on platforms (see Acemoglu et al. (2025) and Susser and Grimaldi (2021)). Within the strand, our paper is most similar to Piccolo et al. (2018), Hattori and Higashida (2012), and Gupta (2023), where ads can be potentially misleading and persuade users to take actions that benefit the advertisers but make themselves worse off. In Piccolo et al. (2018), for example, products are vertically-differentiated and the focus is on the existence of pooling equilibria where advertising obfuscates true differences in quality. In Hattori and Higashida (2012), all consumers are gullible and take misleading advertising at face value rather than make inferences about product quality from the information contained in ads. In Gupta (2023), deceptive advertising is more persuasive to naïve consumers who do not internalize the possibility of false advertising in their belief updates. Our work differs from this literature in three different dimensions as well. The first is again the presence of a platform intermediating between firms and users, and choosing business models (which is the key vehicle for separating equilibria to emerge in our model). Second, ads in our framework are both informative and manipulative, and as noted above, this dual role of advertisement is critical for our results. Third, required policy interventions are very different in our setup. While in the presence of purely deceptive advertising, it is optimal to ban or prevent advertising altogether or strictly regulate deception, in our setup the second-best includes a positive level of advertisement and this can be achieved with nonlinear taxes and subsidies.

Most related is a third strand on ad-funded media with user annoyance costs. In this strand, Anderson and Coate (2005) treat ads as a nuisance and show that ad loads can be too high or too low, depending on programming and market structure. Anderson and Renault (2006) endogenize ad content in the presence of search costs and establish that firms often prefer partial disclosure (while mandating full disclosure could lower welfare). Anderson and Gans (2011) study ad-avoidance (e.g., DVRs or blockers) and find ambiguous welfare effects because of the response of ads and media content. In Section 7.4, we formally show that our results are distinct and do not obtain when there are annoyance costs but no naïve and sophisticated users.[4]

A fourth strand of work relates to a nascent literature on online business models and monetization, to which we are also related. Sato (2019) characterize the optimal business model of a digital platform

[4] Also related are Athey and Gans (2010) and Prat and Valletti (2022) who study the allocation of scarce attention between different media. Our model also features attention scarcity, since each user has a constant time endowment which has to be allocated between ads and entertainment, and chooses which platform to join. But the key decision about the allocation of attention/time is made by the platform, which, together with the interplay between these economic forces and updating by naïfs and sophisticates, is the source of our very different focus and results.

with users who have different demand elasticities for entertainment and advertising. This paper establishes that a two-item menu, comprising a free ad-based plan and a paid-for premium plan with no ads, is profit maximizing for the platform. Building on this work, Zennyo (2020) studies a setup with multiple competing ad-based platforms. In contrast to our work, in these papers digital advertising does not play a crucial role (for example, digital ads are not informative and the quantity of ads does not impact the platform's revenues). More importantly, there is no notion of manipulative advertising (and hence no dual informative-manipulative role of ads) in these models.

A fifth literature we relate to is the one on information design and disclosure, for example, Bergemann and Morris (2019), Ottaviani and Prat (2001), Rayo and Segal (2010), Rayo (2013), Roesler and Szentes (2017), and Skreta (2011). Our contribution and results are very different, however, as we embed such information provision in a two-sided platform that sets an ad regime through its business model and faces users with heterogenous subjective models about a digital ad technology.

Finally, we are also related to empirical works that investigate the costs and effectiveness of advertising. Hundreds of experiments on Google's network document heterogenous effectiveness of ad exposure, consistent with persuasion operating for some audiences but not others (Johnson et al. (2017)). As for user costs, a nine-year Facebook experiment finds that the median disutility from ads is less than 10% of baseline benefits from participating in the platform, suggesting either small direct nuisance costs or offsetting benefits from the ads (Brynjolfsson et al. (2024)). We read this evidence to suggest that ads are indeed costly to users, but the main social costs may come from equilibrium effects—as in our framework. Additionally, our persuasion and manipulation channel also lines up with evidence that persuasive communication can shift behavior (as surveyed in DellaVigna and Gentzkow (2010)) and with quasi-experimental studies showing sizable belief and behavior changes from ad exposure and targeting (e.g., Sinkinson and Starc (2019) and Goldfarb and Tucker (2011)).

The rest of the paper is organized as follows. The next section introduces our model, describes agent payoffs, and defines user welfare. Section 3 characterizes the unique (Berk-Nash) equilibrium of the model and provides comparative statics. Section 4 generalizes the baseline model to allow the platform to adopt richer (mixed) business models. Section 5 studies the effects of introducing firm-level and/or platform-level competition. Section 6 characterizes welfare-increasing policy interventions, Section 7 develops a handful of extensions and robustness checks, while Section 8 concludes. All proofs and some additional analysis omitted from the text are available in Online Appendixes A, B, and C.

2 A Model of Content Platforms

There are three types of agents: firms, platforms, and users. Our baseline model consists of a single firm and a single platform. The firm is a monopolist who sells a single horizontally-differentiated product. The media platform supplies entertainment and (digital) ads to its users, but can intermix advertisements (from now on, simply ads) that are informative about the product.

Users. Users consume the entertaining content offered by the platform and are potential consumers for the product of the firm. There is a continuum of users who each have a two-dimensional type

$(\tau_i, \theta_i) \in \{S, N\} \times \{0, 1\}$. The first dimension corresponds to the user's sophistication level; each user i is either sophisticated ($\tau_i = S$, with probability λ) or naïve ($\tau_i = N$, with probability $1 - \lambda$). The second dimension of the user's type, $\theta_i \in \{0, 1\}$, represents whether the product offered by the firm is high or low quality for her. Specifically, the product is high-quality for user i ($\theta_i = 1$), with prior probability q. All events are independent across users and other random variables—in particular, product quality is user-specific with no common component. Users derive utility from the products they purchase and from the entertaining content on the platform, as we will describe below.

Firm. The firm is a monopolist and sells a single product at unit price p. This implies that any user i who purchases z_i pays price pz_i. The firm's marginal cost of production is constant and equal to c.

Platform. The platform operates as a two-sided marketplace, connecting firms with users. It offers engaging content, such as videos and music, while also displaying digital ads on behalf of firms. Each user spends a total time $T > 0$ on the platform, during which the platform determines the proportion of time allocated to ads versus entertaining content (where T is exogenous, see Footnote 7). Specifically, we assume that there is a single ad that may be shown multiple times to the user. The appearance of ads follows a Poisson process with a rate of α, meaning the total number of ad displays is distributed according to Poisson(αT). Each ad lasts for a normalized duration of 1. The platform selects the parameter α, which we refer to as "ad load".

The probability that a user sees the ad at least once is therefore $1 - e^{-\alpha T}$. When an ad is viewed, it provides the user with an informative signal about her type (i.e., the product's quality for her), denoted by θ_i. However, because advertising reduces the time a user spends consuming content she enjoys, more frequent advertising comes at a cost. Under our Poisson assumption, the expected time the user spends viewing entertaining content is $(1 - \alpha)T$, with the remaining time αT being allocated to ads.

Information Structure. If user i views the ad, it provides a binary signal $s_i \in \{G, B\}$ about the product, which is independent across users. However, if the same ad is shown multiple times to the same user, she does not obtain additional information from this. The signal distribution for the ad is given by

$$\begin{cases} s_i = G, & \text{with probability } \phi_1 \text{ if } \theta_i = 1\,, \\ s_i = B, & \text{with probability } 1 - \phi_1 \text{ if } \theta_i = 1\,, \\ s_i = G, & \text{with probability } \phi_0 \text{ if } \theta_i = 0\,, \\ s_i = B, & \text{with probability } 1 - \phi_0 \text{ if } \theta_i = 0\,, \end{cases} \tag{1}$$

where we assume that $\phi_1 > \phi_0$, which means that a positive ("good") signal provides information to the user that the product is *more* likely to be high quality for her ($\theta_i = 1$), while a negative ("bad") signal is bad news about the match quality. This implies that there are both type-I and type-II errors. We assume throughout that the signal distribution of (1) is the objective model or the "ground truth".[5]

[5]The assumption that a single ad is shown on the platform is for simplicity. Our analysis readily generalizes to the case of multiple different ads whereby each ad provides additional informative signals about θ_i. In this case, the number of ads seen by each user would still be given by $k \sim$ Poisson(αT), but now the ads generate incremental information about the user's preferences. This implies, in particular, that the number of ads with signal $s_i = G$ would be drawn as a binomial

Users evaluate signals according to their *subjective model*, which can differ from the objective model in (1). Specifically, we assume that the subjective model of user i of type τ_i on signal distribution is

$$\begin{cases} s_i = G, & \text{with probability } \phi_1 \text{ if } \theta_i = 1\,, \\ s_i = B, & \text{with probability } 1-\phi_1 \text{ if } \theta_i = 1\,, \\ s_i = G, & \text{with probability } \phi_{0,\tau_i} \text{ if } \theta_i = 0\,, \\ s_i = B, & \text{with probability } 1-\phi_{0,\tau_i} \text{ if } \theta_i = 0\,. \end{cases} \tag{2}$$

The subjective model summarizes the extent to which agents understand how ads are targeted and customized to their specific circumstances. We assume that sophisticated agents are aware of these marketing strategies, so for them, $\phi_{0,S} = \phi_0$. On the other hand, naïve agents are not fully aware, and we represent their parameter as $\phi_{0,N} = \omega_N \omega_P \phi_0$. Here, $\omega_N \le 1$ reflects their naïveté (which would apply even without personalization), while $\omega_P \le 1$ accounts for the personalized tailoring and targeting of ads, which may not be fully understood by naïve agents. We assume that $\omega_N \omega_P < 1$. Although the specific manner in which this bias is introduced is not crucial to our results, it may be relevant for certain informational interventions. For simplicity, throughout the rest of the paper we will treat ϕ_0 and $\phi_{0,N}$ as model primitives, suppressing the dependence on ω_N and ω_P.

Throughout we use the term (*de facto*) *manipulation* as shorthand for *selective persuasion*: naïfs place excess weight on positive signals from digital ads (because $\phi_{0,N} < \phi_0$) and the platform and the monopolist firm take advantage of this proclivity (see footnote 3). In Section 7.1, we endogenize manipulation by allowing the platform to choose $\phi_{0,N}$ (for example, by modifying the characteristics of ads or other aspects of the environment), and show the profit-maximizing platform strategy is to choose the environment with the lowest $\phi_{0,N}$.[6]

The case with $\omega_N = \omega_P = 1$ (which implies $\phi_0 = \phi_{0,N}$) represents the *fully-rational benchmark*, where naïveté does not affect how ads are perceived, and digital ads are either not tailored for or not targeted to specific individuals—or if they are, their targeting is fully understood by all agents. Alternatively, the case with $\lambda = 1$, where all agents are sophisticated, also corresponds to this benchmark. Our primary focus in this paper is on the case where $\phi_{0,N} < \phi_0$, which we interpret as reflecting a common real-world scenario in which targeted *digital advertising* can mislead at least some agents. In this context, the difference between $\phi_{0,N}$ and ϕ_0 measures the platform's ability to *de facto* manipulate naïve agents.

distribution with k trials and success probability ϕ_{θ_i}. Our results go through identically under this alternative formulation, with the exception that the (post-ad) conditional probability of user i that $\theta_i = 1$ becomes $\pi_i = (\phi_1^{k_+}(1-\phi_1)^{k_-}q)(\phi_1^{k_+}(1-\phi_1)^{k_-}q + \phi_{0,\tau_i}^{k_+}(1-\phi_{0,\tau_i})^{k_-}(1-q))^{-1}$, which depends on the number of positive signals k_+ and the number of negative signals k_-, with $k_+ + k_- = k$.

[6]This is in line with the current practice on social media platforms. For instance, Facebook utilizes data on browsing history, clicks, shares, and likes to classify users into "custom audiences" for various advertisers which then adopt different strategies for micro-targeting and influence (see Tran (2017), Kruikemeier et al. (2016), Galán et al. (2019) for details on these marketing strategies). New generative AI tools have further enabled micro-targeted advertising techniques, where hyper-personalized content can be generated individually in real-time (see Simchon et al. (2024) and Golab-Andrzejak (2023)). Platforms may also be able to predict some of the "bad" signals and promote the ads with positive signals.

2.1 Actions and Timing

Next, we define the exact strategic game played by our agents. The game will consist of five stages, denoted $t = 1, 2, 3, 4$, and 5, as depicted in Figure 1.

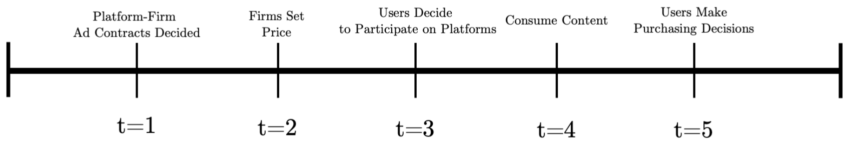

Figure 1. Timing of the Advertising Model.

(i) At $t = 1$, the platform and the firm negotiate a contract that specifies an ad load α and a monetary transfer m from the firm to the platform (the "advertising revenue"). For simplicity, we assume this takes the form of a take-it-or-leave-it offer (α, m) from the platform to the firm, which is then either accepted or rejected by the firm. If the firm rejects the contract (or the platform does not offer one), the platform can advertise at whatever rate it desires. The platform can also set a subscription fee P for users to join the platform, and the acceptance decision of the firm can be conditioned on P (since this determines participation in the platform).

(ii) At $t = 2$, the firm sets its price p^* for its product.

(iii) At $t = 3$, the platform produces content and advertises at the rate α. Each user i makes a binary decision $x_i \in \{0, 1\}$ about whether to spend time T ($x_i = 1$) or no time ($x_i = 0$) on the platform, given α and the subscription fee P.[7] The non-participation decision $x_i = 0$ gives user i an outside option $v > 0$.

(iv) At $t = 4$, if $x_i = 1$, the user digests the platform content (including any ads offered at the Poisson rate α). She receives entertainment value equal to $(1 - \alpha)T$.

(v) At $t = 5$, each user i decides how much of the product to purchase, z_i, at the price p, based on her posterior about θ_i.

2.2 Payoffs and Solution Concept

Platform. Recall that the platform can generate revenue by charging the firm for advertising and/or by charging users a subscription fee. When the firm accepts a contract (α^*, m^*) and the platform charges P^* to users, its payoff is

$$m^* + \int_0^1 P^* x_i \, di \,,$$

[7] Our results are very similar if we allow a continuous time allocation decision $x_i \in [0, 1]$, since consuming platform content (potentially with ads) has diminishing marginal utility and the outside option is constant, and thus there will be two potential candidates for consumption $x_i \in \{0, \bar{x}\}$. The only additional complication in this case would be $\bar{x}$ may change with some parameters, rather than always being equal to T.

with the convention that if the firm rejects the contract, then the monetary transfer is $m^* = 0$.

Firm. The firm generates profits by selling its product, but pays the platform for advertising. That is, the firm receives a payoff

$$\int_0^1 (p^* - c) z_i^* \, di - m^*,$$

where z_i^* is the consumption decision of agent i and m^* is the transfer to the platform (if the platform's contract is accepted and zero otherwise).

Users. Each user i receives utility both from product consumption and from content consumption on the platform. As specified above, the utility from the content on the platform is $(1-\alpha)T$, when the user is on the platform or v when she chooses her outside option. In addition, given her type θ_i and consumption level z_i, she receives a consumption utility $U(z_i; \theta_i) = \beta\theta_i z_i - z_i^2/2$. This implies that her expected utility from this consumption is

$$\max_{z_i \geq 0} \mathbb{E}^{\tau_i} \left[U(z_i; \theta_i) - p z_i \right] = \max_{z_i \geq 0} \mathbb{E}^{\tau_i} \left[\beta\theta_i z_i - z_i^2/2 - p z_i \right],$$

where $\mathbb{E}^{\tau_i}$ is the expectation according to type τ_i's subjective probably distribution.[8] Given the linear-quadratic utility, the parameter β is the slope of the demand for the product and thus determines the elasticity of demand. Linear-quadratic utility is a simplifying assumption and, as we discuss further below, it implies that for Bayesian agents with the correct probability distribution additional information does not change the expected quantity consumed, which is a convenient benchmark.[9]

Solution Concept. We use the notion of (perfect) Berk-Nash equilibrium (Esponda and Pouzo (2016)) to model agents' beliefs under misspecified signal structures. Perfection here simply means that we impose sequential rationality at each information set, given beliefs, and when this causes no confusion, we refer to our equilibrium notion as Berk-Nash equilibrium or simply as "equilibrium". This implies in particular that all agents are Bayesian, but only given their subjective model. Because the subjective model of sophisticates is the objective model, a sophisticated agent will have a standard Bayesian belief π^S about $\theta_i = 1$ conditional on ad signals:

$$\pi^S(s_i) = \begin{cases} \frac{\phi_1 q}{\phi_1 q + \phi_0(1-q)}, & \text{if } s_i = G, \\ \frac{(1-\phi_1)q}{(1-\phi_1)q + (1-\phi_0)(1-q)}, & \text{if } s_i = B. \end{cases}$$

[8]For sufficiently high values of β, we always have $z_i^* > 0$, and for simplicity, we focus on such cases and drop the non-negativity constraint from z_i. We view this as the empirically relevant configuration, since estimates in the literature suggest relatively lower demand elasticities (high β) for products as compared to entertainment (see Chyi (2005), Vock et al. (2013), and Chyi and Ng (2020)). See also Berger et al. (2015), Sherman and Waterman (2016), and Flew (2021). Appendix B relaxes the assumption that $z_i^* > 0$ and shows that our results are essentially identical with the kinked demand curves that arise without this assumption.

[9]Beyond the linear-quadratic case, expected consumption may increase or decrease depending on whether the implied demand curve is concave or convex.

In particular, $\pi^S \mid \theta_i$, conditional on viewing an ad, will be distributed as a multinomial with

$$\pi^S \mid \theta_i \sim \begin{cases} q, & \text{with probability } e^{-\alpha T}, \\ \frac{\phi_1 q}{\phi_1 q + \phi_0 (1-q)}, & \text{with probability } \phi_{\theta_i}(1 - e^{-\alpha T}), \\ \frac{(1-\phi_1) q}{(1-\phi_1) q + (1-\phi_0)(1-q)}, & \text{with probability } (1-\phi_{\theta_i})(1 - e^{-\alpha T}), \end{cases}$$

where $\alpha \in [0, 1]$ is the advertising load of the platform. We refer to F_{S0} and F_{S1} as the distributions over $\pi^S \mid \theta_i$ for $\theta_i = 0$ and $\theta_i = 1$, respectively.

Naïfs, on the other hand, update according to their subjective model and have:

$$\pi^N(s_i) = \begin{cases} \frac{\phi_1 q}{\phi_1 q + \phi_{0,N}(1-q)}, & \text{if } s_i = G, \\ \frac{(1-\phi_1) q}{(1-\phi_1) q + (1-\phi_{0,N})(1-q)}, & \text{if } s_i = B, \end{cases}$$

where $\pi^N \mid \theta_i$ is distributed as a multinomial with

$$\pi^N \mid \theta_i \sim \begin{cases} q, & \text{with probability } e^{-\alpha T}, \\ \frac{\phi_1 q}{\phi_1 q + \phi_{0,N}(1-q)}, & \text{with probability } \phi_{\theta_i}(1 - e^{-\alpha T}), \\ \frac{(1-\phi_1) q}{(1-\phi_1) q + (1-\phi_{0,N})(1-q)}, & \text{with probability } (1-\phi_{\theta_i})(1 - e^{-\alpha T}). \end{cases}$$

Importantly, because $\phi_{0,N} < \phi_0$, we have $\pi^S \mid \theta_i \preceq_{FOSD} \pi^N \mid \theta_i$ for both $\theta_i \in \{0, 1\}$. In other words, the beliefs of naïve agents that $\theta_i = 1$ are more favorable than the beliefs of sophisticated agents given their ad viewership. The special case with $\phi_{0,N} = 0$, which implies that a positive ad is always interpreted by naïve agents as evidence of high product quality, is useful for building intuition about $\pi^N \mid \theta_i$. (See Section 7.1 for other ways of modeling misperceptions.)

Note also that the distributions over $\pi^N \mid \theta_i$ (for $\theta_i = 0$ and $\theta_i = 1$), denoted by F_{N0} and F_{N1}, are governed by the objective probability distribution over signals—rather than the naïfs' subjective model. That is, while the *interim beliefs* $\pi^N(G)$ and $\pi^N(B)$ are updated using $\phi_{0,N}$ for naïfs, the induced probabilities over $\{q, \pi^N(G), \pi^N(B)\}$ when $\theta_i = 0$ are governed by the objective model using ϕ_0. Hence, naïve agents will make decisions using their *interim* beliefs, but their *ex post* utility, given those decisions, will be determined by the objective measures.

Equilibrium. We can determine the unique (perfect) Berk-Nash equilibrium of this sequential game via backward induction.

(a) At $t = 5$, each user holds belief π_i that she values the product ($\theta_i = 1$) and chooses her optimal consumption z_i^* to solve $z_i^*(\pi_i, p) \equiv \arg\max_{z_i} \pi_i U(z_i; \theta_i = 1) + (1 - \pi_i) U(z_i; \theta_i = 0) - p z_i$ given π_i, where $z_i^*(\pi_i, p)$ represents the expected consumption utility *given* belief π_i about $\theta_i = 1$ and given product price p.

(b) At $t = 4$, user i's belief π_i is determined by Bayes' rule given the realization of the signal for this user (conditional on her participation decision $x_i \in \{0, 1\}$ and the ad load α) and her subjective model, which itself depends on her sophistication type τ_i.

(c) At $t = 3$, each user decides whether to participate on the platform by solving

$$\max_{x_i \in \{0,1\}} x_i(\mathbb{E}^{\tau_i}_{\pi_i}[U(z_i^*(\pi_i, p); \theta_i = 1) - p z_i^*(\pi_i, p) \mid \alpha] + (1-\alpha)T) + (1-x_i)(U(z_i^*(q,p); \theta_i = q) - p z_i^*(q,p) + v),$$

where recall that $\mathbb{E}^{\tau_i}$ is the expectation with respect to the subjective model of a user with type τ_i. When participating on the platform and observing digital ads ($x_i = 1$), users obtain an informational value, since they believe that these ads lead to better decisions. Specifically, the perceived (interim) informational value from digital ads α for users of type τ_i is:

$$\begin{aligned} I_{\tau_i}(\alpha) =& q\mathbb{E}^{\tau_i}_{\pi_i|\theta_i=1}[U(z_i^*(\pi_i, p^*);\ \theta_i = 1) - p^* z_i^*(\pi_i, p^*) \mid \alpha] \\ &+ (1-q)\mathbb{E}^{\tau_i}_{\pi_i|\theta_i=0}[U(z_i^*(\pi_i, p^*);\ \theta_i = 0) - p^* z_i^*(\pi_i, p^*) \mid \alpha] - (U(z_i^*(q, p^*);\ \theta_i = q) - p^* z_i^*(q, p^*)). \end{aligned}$$

That this quantity does not depend on price p^* is established in Lemma A.1.

(d) At $t = 2$, given ad load α, the firm sets price p by solving

$$\Pi(\alpha) \equiv \max_{p \geq 0} \int_0^1 \left[x_i^*(\alpha, p)(p - c)\mathbb{E}_{\tau_i}[\mathbb{E}^{\tau_i}_{\pi_i}[z_i^*(\pi_i, p) \mid \alpha]] + (1 - x_i^*(\alpha, p))(p - c) z_i^*(q, p) \right] di,$$

where $\Pi(\alpha)$ is the expected profit of the firm given an advertising load α on the platform.

(e) At $t = 1$, given a contract (α, m) and subscription fee P, the firm accepts the contract if and only if $\Pi(\alpha) - m \geq \max_{p \geq 0}(p-c) z_i^*(q, p)$. The platform then selects the contract (α, m) and the subscription fee P that maximize m conditional on the acceptance rule of the firm.

We assume the outside option satisfies $v > \lim_{\alpha \to \infty} I_N(\alpha)$, so that users only participate on the platform when at least some entertainment content is shown.

2.3 Empirical Context

Before proceeding to our analysis, it is useful to map our assumptions to the relevant empirical context.

Most digital platforms offer an ad-supported tier, which mixes ads with entertaining content (including social media, streaming, and mobile games). In all of these cases, ads are targeted to specific groups of users, for example, towards those who will be more responsive. Many of these platforms also offer paid upgrades that reduce or remove ads. Also, as in our model, these platforms choose (i) ad exposure (the ad loads in each plan) and (ii) subscription prices, while advertisers pay platforms in order to deploy their ads, targeting various user categories.

Three empirical regularities are useful to note in this context. First, ad-based versus ad-free menus exhibit sizable price differences. Second, targeting and persuasion matter for outcomes: experimental studies document heterogeneous impacts from targeted ads and also show that curbs on targeting reduce effectiveness, both short-term and long-term (e.g., Goldfarb and Tucker (2011); Sinkinson and Starc (2019); surveys in DellaVigna and Gentzkow (2010); large-scale tests in Lewis and Rao (2015);

Gordon et al. (2019)). Third, ad-supported average revenue per user can rival or exceed revenues from ad-free tiers.[10]

Our framework is less well-suited to pure search settings where users primarily care about query content and have no plan choice or to cases where nuisance costs largely drive ad avoidance (see Section 7.4 for how such avoidance interacts with our mechanisms).

2.4 User Welfare

We now describe the *ex post* utilities for the two kinds of users.

When a user abstains from participating on the platform, her utility is

$$W(\tau_i, x_i = 0) = \underbrace{v}_{\text{Outside Option}} + \underbrace{qU(z^*(q,p^*);\, \theta_i = 1) + (1-q)U(z^*(q,p^*);\, \theta_i = 0) - p^* z^*(q,p^*)}_{\text{Expected Product Consumption Utility}},$$

where p^* is the product price.

Next we characterize the user's utility when she engages with the platform setting ad load $\alpha^*$and subscription fee P^*. Recall that F_{S0}, F_{S1}, F_{N0} and F_{N1} denote the distributions over π_i for the respective types, $(S,0)$, $(S,1)$, $(N,0)$, and $(N,1)$, using the objective model (which depend on the ad load α^*), which, recall, determines *ex post* utility (while subjective interim beliefs of naïve types determine their actions). Average user utilities by type can then be written as

$$\begin{aligned} W(\tau_i, x_i = 1) = \underbrace{(1-\alpha^*)T - P^*}_{\text{Content Consumption Surplus}} &+ q\, \underbrace{\mathbb{E}_{\pi \sim F_{\tau_i 1}} \left[U(z^*(\pi,p^*);\, \theta_i = 1) - p^* z^*(\pi,p^*)\right]}_{\text{Product Consumption Utility Conditional on } \theta_i = 1} \\ &+ (1-q)\, \underbrace{\mathbb{E}_{\pi \sim F_{\tau_i 0}} \left[U(z^*(\pi,p^*);\, \theta_i = 0) - p^* z^*(\pi,p^*)\right]}_{\text{Product Consumption Utility Conditional on } \theta_i = 0}. \end{aligned}$$

We denote the *ex post* utility of sophisticated and naïve agents by $W^*(S) = W(S, x_S^*)$ and $W^*(N) = W(N, x_N^*)$, respectively.

2.5 First Best

We start by characterizing the first-best allocation, which clarifies how user utility can be maximized when a planner can control ad loads separately for sophisticates and naïfs (though we still allow users to make their own purchasing decisions).

Proposition 1. *The first best involves the platform advertising at the rate α_S^{FB} to sophisticates and at the rate α_N^{FB} to naifs, where $\alpha_N^{FB} \leq \alpha_S^{FB}$. Moreover, for type τ user, $W_{FB}(\tau) > W_{base}(\tau)$, where $W_{base}(\tau)$ is the base case utility with no platform.*

Proposition 1 shows that the first-best allocation involves a small amount of advertising on the platform to both types of agents—but crucially different amounts for different types. Advertising improves sophisticated agents' decision-making and is therefore socially valuable. In addition, in the

[10]For the case of Netflix, see insidermonkey.com/blog/netflix-inc-nasdaqnflx-q2-2023-earnings-call-transcript-1170001.

first best, sophisticates enjoy the content on the platform, and hence $W_{FB}(S) > W_{\text{base}}(S)$. The same forces are present for naïfs, but the social planner prefers to send them fewer ads because they tend to misinterpret the information in the ads and thus they derive less *ex post* utility from ads. The fact that the social planner generally chooses strictly positive ads underscores their informative nature in our model.

Notice an important property, however. If naïve agents were given a choice between the two levels of advertising, α_S^{FB} and $\alpha_N^{FB} \leq \alpha_S^{FB}$, they would choose the higher one, α_S^{FB}. In fact, given the option, they would prefer even higher levels of advertising than α_S^{FB}. This is because, at the interim stage, they *erroneously* think the ads are more informative than they truly are. This is one of the reasons why the first best will never be implementable in a decentralized equilibrium.

3 Baseline Equilibrium Characterization

To build intuition, we start with a simplified version of our model, where we allow the platform to either charge a subscription fee or use digital ads, but *not both*. This has no major effect on the main insights, but simplifies our initial characterization.

3.1 Equilibrium Business Models and Digital Ads

We start by presenting a number of lemmas, which together deliver the main characterization results (for behavior and welfare) in this baseline environment.

Lemma 1. *Let $\Pi_S^*(\alpha) = \int_{\tau_i = S} (p^* - c) z_i^* \, di$ be the firm's profit from the sophisticated agents under an advertising scheme with ad load $\alpha > 0$. Then, $\Pi_S^*(\alpha)$ is independent of α. In other words, the firm extracts no surplus from advertising to the sophisticated agents.*

Lemma 1 demonstrates that the firm cannot extract advertising rents from sophisticated agents' participation on the platform. This is a consequence of linear-quadratic utility and the resulting linear demand curves. Agents for whom the product is high quality receive more positive ads on average, while those who dislike the product encounter more negative ads. Although the former group is willing to purchase more and pay higher prices, the latter group's lower willingness-to-pay and reduced consumption perfectly offset these gains. This balance is a direct result of the Martingale property of Bayesian beliefs combined with linear demand. Put differently, the user's demand curve without advertising is $z_i^*(p) = \beta q - p$, whereas after advertising it becomes $z_i^*(p) = \beta \pi_i - p$. Since sophisticated agents are fully Bayesian, their expected posterior equals the prior, $\mathbb{E}^S[\pi_i] = q$. This implies that the expected demand after advertising remains the same as the demand before advertising. This does not imply that users do not benefit from information—they do, as they make more informed decisions and thus obtain greater consumer surplus. However, the firm cannot capture any of this surplus because the users' expected demand remains unchanged. As a result, there is no surplus for the platform to capture by charging the firm to display digital ads to users. Therefore, the platform also does not profit from showing digital ads to sophisticated users.

Although this result does not hold exactly with concave demand curves, it transparently illustrates why the main source of profits for the firm (and thus ad revenue for the platform) is the additional demand from naïve agents—ads will generally have a small or zero impact on the expected purchases of sophisticated agents, but potentially much larger effects on the expected purchases of naïve agents who overestimate the likelihood of a high-quality product given a positive signal.

Lemma 2. *Let $\Pi_N^*(\alpha) = \int_{\tau_i = N} (p^* - c) z_i^* \, di$ be the firm's profit from the naïve agents under an advertising scheme with ad load $\alpha > 0$. Then, $\Pi_N^*(\alpha)$ is positive and increasing in α. In other words, the firm extracts positive surplus from advertising to naïfs and this surplus is greater when there is more advertising.*

In contrast to sophisticated agents, naïfs' average demand curve drifts upward as the advertising load increases, even though they also have a linear demand curve. This result is rooted in the fact that naïfs have the wrong model and update their beliefs under the perception that low-quality products generate positive signals with probability $\phi_{0,N}$, while in truth they generate such signals with probability $\phi_0 > \phi_{0,N}$. Using our notation of F_{N0} and F_{N1} for the distribution of beliefs π_i for naïfs, we can write their demand curve as $z_i^*(p) = \beta(q\mathbb{E}_{\pi_i \sim F_{N1}}[\pi_i \,|\, \alpha^*] + (1-q)\mathbb{E}_{\pi_i \sim F_{N0}}[\pi_i \,|\, \alpha^*]) - p$, which is strictly greater than their expected purchases without ads, $\beta q - p$, and is also increasing in α. This allows the firm to charge higher prices and secure greater profits in the product market when there are naïve agents receiving digital ads. This surplus by the firm is then extracted by the platform via the monetary transfer, m^*. This comparative static aligns with existing results showing that higher profits from ads push toward greater ad intensity (e.g., Anderson and Coate (2005)). Relative to these results, our contribution is about the interplay of ads with sophisticated and naïve agents—making platforms target naïfs—and how this changes platform business models and creates additional spillovers via pricing decisions.

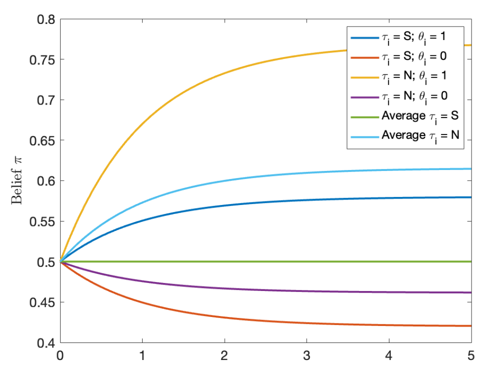

Figure 2. Expected posterior beliefs under ad load α (example with $T = 1$, $q = 0.5$, $\phi_1 = 0.7$, $\phi_0 = 0.3$, and $\phi_{0,N} = 0$).

To see the findings of Lemma 1 and 2 most transparently, consider Figure 2. In the figure, the naïf posterior belief curve with $\theta_i = 1$ rises steeply with α while the naïf posterior belief curve with $\theta_i = 0$ falls only slightly. This asymmetry is the naïf misspecification: they over-weight good signals and under-react to bad signals. The consequence is Lemma 2, that on average the naïf belief increases in α: as ad load rises, naïfs' average beliefs drift upward. This is in stark contrast to Lemma 1 and the average sophisticate posterior belief curve, which has their average posterior equal to their prior, because sophisticates correct appropriately for the arrival of good/bad signals. The takeaway from Lemma 2 is that because the average naïf posterior curve rises, more ads increase profits and shift the profit-maximizing business model toward ad-heavy plans with larger product price hikes.

The same forces underlying Lemma 2 also lead to our next result about the (interim) informational value that sophisticated and naïve agents derive from digital ads.

Lemma 3. *For any* α, $I_N(\alpha) > I_S(\alpha) > 0$ *and* $\arg\max_{\alpha\in[0,1]} I_N(\alpha) + (1-\alpha)T > \arg\max_{\alpha\in[0,1]} I_S(\alpha) + (1-\alpha)T$. *Moreover,* $I_S(\alpha)$ *and* $I_N(\alpha)$ *are concave and monotonically increasing in* α.

Because naïfs mistakenly believe that digital ads are more informative than they truly are, their subjective (interim) value from participating in the platform is greater. This implies in particular, that naïve agents are more tolerant of digital ads and in fact would choose a higher level of digital ads than sophisticated agents, as we also noted in our discussion of why the first-best allocation cannot be implemented. It also implies that they are more willing to take part in an ad-based plan. The utility of an agent of type τ_i from platform participation is $I_{\tau_i}(\alpha) + (1-\alpha)T - v \geq 0$, and thus Lemma 3 implies that the constraint for participating in the platform will always bind for sophisticates before it binds for naïfs. Put differently, whenever sophisticates participate in the platform, so do naïfs, but not vice-versa.

These three lemmas together with platform maximization yield our next result:

Lemma 4. *If the platform adopts an advertising-based business model, it sets* α^* *such that* $I_N(\alpha^*) + (1-\alpha^*)T - v = 0$ *and* $I_S(\alpha^*) + (1-\alpha^*)T - v < 0$. *In other words, the platform extracts all surplus from naïve agents, while sophisticates do not participate on an advertising-based platform.*

Intuitively, from Lemmas 1 and 2, the total advertising revenue the platform can generate is increasing in the ad load α. Moreover, from Lemma 3, the participation constraint will bind first for the sophisticates, and from Lemma 1, the platform does not collect any additional advertising revenue from serving ads to sophisticates. Therefore, the platform finds a simple form of separation profitable: sophisticates are excluded and naïfs receive a high load of digital advertising (making them indifferent between participating and not participating in the platform). Lemma 2 also shows that raising ad load α increases naïfs' average valuation, while leaving sophisticates' average valuation unchanged (per Lemma 1). As a result, total demand shifts out, which pushes up the profit-maximizing price. This creates an indirect coupling between naïve and sophisticated agents—sophisticates pay more for the same product when naïfs have an inflated demand. This hurts sophisticates even when they do not participate directly on the platform.

Whether the platform will actually choose this separating allocation depends on how much it can collect as a subscription fee from both types. This calculation leads to our main characterization result in this baseline environment:

Proposition 2. *There exists* $\hat{\phi}_0(\lambda, \phi_1, \phi_{0,N}) > \phi_{0,N}$ *such that:*

(a) If $\phi_0 < \hat{\phi}_0(\lambda, \phi_1, \phi_{0,N})$, *the platform chooses a subscription model with* $P^* = T - v$ *and the firm sets price* $p^* = \bar{p}^* \equiv (\beta q + c)/2$ *for the product;*

(b) If $\phi_0 > \hat{\phi}_0(\lambda, \phi_1, \phi_{0,N})$*, the platform chooses an advertising model with ad load* $\hat{\alpha}^* = \arg\max\{\alpha \in [0,1] : I_N(\alpha) + (1-\alpha)T - v = 0\}$ *and the firm sets price* $\hat{p}^* > \bar{p}^*$ *for the product.*

Moreover, $\hat{\phi}_0(\lambda, \phi_1, \phi_{0,N})$ *is increasing in* λ *and* $\phi_{0,N}$*, and decreasing in* ϕ_1*.*[11]

Proposition 2 follows from Lemmas 1-4, which collectively indicate that the profit-maximizing business model depends on the extent to which the platform can extract ad revenue from the firm, which, in turn, extracts surplus from naïve agents. When ϕ_0 is low (relative to $\phi_{0,N}$)—more precisely, when it is lower than the threshold $\hat{\phi}_0(\lambda, \phi_1, \phi_{0,N})$—there exists a small gap between the actual likelihood of false positives (from low-quality products) and the perceived likelihood of false positives by naïve agents. In this scenario, the equilibrium is in regime (a), because the amount of surplus that can be extracted from naïve agents is small and it is therefore more profitable to include all agents on the platform and charge a subscription fee. This fee will be chosen to be equal to the additional utility all agents derive from the entertaining content, $P^* = T - v$. In this regime, the profit-maximizing monopoly price for the firm is $\bar{p}^* \equiv (\beta q + c)/2$, and is determined by the *ex ante* linear demand curves (in the absence of digital ads there is no further information acquisition by either type of agent).

Conversely, when ϕ_0 is higher than the threshold $\hat{\phi}_0(\lambda, \phi_1, \phi_{0,N})$, it is more profitable to opt for an ad-based business model and exclude sophisticates from the platform. In this scenario (regime (b)), the price charged by the firm for its product is greater than in regime (a), $\hat{p}^* > \bar{p}^*$. This is because with digital ads, naïve agents have an inflated perception of the likelihood that they will like the product and this raises the monopoly price of the firm.

The proposition also shows that $\hat{\phi}_0(\lambda, \phi_1, \phi_{0,N})$ is decreasing in ϕ_1. This is because, in this case, ads are more informative (they are more likely to give good signals when the product is high-quality for the agent), and thus they are more valuable to users. This allows the platform to increase the ad load and extract more surplus from naïve agents, even with lower values of ϕ_0—which is the meaning of a lower cutoff $\hat{\phi}_0(\lambda, \phi_1, \phi_{0,N})$, which expands the range of ϕ_0 where the platform adopts an advertising-based business model.

Three other points are important to note. First, $\hat{\phi}_0(\lambda, \phi_1, \phi_{0,N})$ is increasing in λ for obvious reasons: when there are more sophisticated agents, excluding them is more costly, and thus a subscription-based business model becomes more likely. Put differently, the presence of sophisticates confers a positive (pecuniary) externality on naïfs: by tightening the platform's IC/participation constraints, it reduces the profitability of naïf-targeting, ad-heavy menus and thus forces the platform to be less exploitative of them. Second and more importantly, digital ads targeted at naïve agents generate a negative spillover on sophisticates: not only are they excluded from the platform, but they face a higher price for the product, $\hat{p}^* > \bar{p}^*$, than they would have done in regime (a). Third, the role of the incorrect model that naïve agents rely on for evaluating the meaning of positive signals is critical for Proposition 2, as witnessed by the fact that the equilibrium business model depends on the gap between the true

[11] In all of the comparative static results we present, we vary one parameter while holding the others fixed. In doing so, we treat ϕ_0 and $\phi_{0,N}$ as independent parameters. One could alternatively define ω_N and ω_P as independent variables and let $\phi_{0,N} = \omega_N \omega_P \phi_0$, so that varying ϕ_0 simultaneously varies $\phi_{0,N}$. This alternative formulation has no substantial impact on our results (the equilibrium would retain the same cutoff structure and comparative statics with respect to λ and ϕ_1 would remain unchanged) but the cutoff $\hat{\phi}_0$ in Proposition 2 would be higher (because we are now increasing both ϕ_0 and $\phi_{0,N}$).

likelihood of positive signals for low-quality products, ϕ_0, and their perceived likelihood, $\phi_{0,N}$. This can be seen also from considering the fully-rational benchmark, which we derive in the next subsection.

3.2 Fully-Rational Benchmark

It is useful to consider the fully-rational benchmark where $\phi_{0,N} = \phi_0$ and thus there is no misperception on the part of naïve agents.

Proposition 3. *If $\phi_{0,N} = \phi_0$, the profit-maximizing platform business model of the platform is subscription-based with $P^* = T - v$ and the firm sets price $p^* = \bar{p}^*$. The welfare of agents of type $\tau \in \{S, N\}$ of the fully-rational benchmark, $W_{\textit{fully-rational}}(\tau)$, is equal to their base case welfare, $W_{\textit{base}}(\tau)$, with no platform at all.*

In this scenario, naïve agents have an accurate understanding of how digital ads generate signals, resulting in no distinction between sophisticated and naïve agents. This implies that neither the firm nor the platform can extract any informational surplus from any of the users. Consequently, the profit-maximizing business model is subscription-based, with the same subscription fee as in Proposition 2, $P^* = T - v$. This enables the platform to capture all the surplus users would derive from consuming entertainment on the platform. The monopolist firm also sets the same price $\bar{p}^*$ as in Proposition 2. Since the platform is extracting all the surplus it helps create, the utilities of both types of agents are the same as they would have been in the hypothetical case where the platform did not exist.

3.3 Digital Advertising: Welfare Analysis

Armed with the characterization of equilibrium in Proposition 2, we next determine user welfare in the benchmark equilibrium, separately in regimes (a) and (b).

Proposition 4.

(a) When $\phi_0 < \hat{\phi}_0(\lambda, \phi_1, \phi_{0,N})$, user welfare is $\hat{W}^*(\tau) = W_{\textit{fully-rational}}(\tau) = W_{\textit{base}}(\tau)$ for both $\tau \in \{S, N\}$.

(b) When $\phi_0 > \hat{\phi}_0(\lambda, \phi_1, \phi_{0,N})$, user welfare is $\hat{W}^*(\tau) < W_{\textit{fully-rational}}(\tau) = W_{\textit{base}}(\tau)$ for both $\tau \in \{S, N\}$.

The first part of this proposition is not surprising. When $\phi_0 < \hat{\phi}_0(\lambda, \phi_1, \phi_{0,N})$, the profit maximizing business model involves a pure subscription fee and no user sees any digital ads. The platform then captures the full surplus it creates, owing to its ability to make a take-it-or-leave-it offer to users. Because there are no digital ads and thus no additional signals on quality, the equilibrium monopoly price is the same as the case without the platform as well, $\bar{p}^*$. Consequently, user welfare, for both sophisticated and naïve agents, is exactly the same as it would have been without the platform (which also coincides with the fully-rational welfare levels as we saw in Proposition 3).

The second part of the proposition is one of our main results. It shows that both sophisticates and naïfs have lower utility than the case in which the platform is not present. Naïfs, have lower *ex post* utility than the case without the platform because they receive a high number of digital ads, and given their misperception about the signal generating process, they are *de facto* manipulated and spend more on the product of the monopolist firm than they would have done with the correct beliefs. This in particular pushes them into consuming more of the product when it is low-quality for them, lowering their *ex post* utility. Their inflated demand for the product also leads to a higher monopoly price $\hat{p}^* > \bar{p}^*$ than in regime (a), and this reduces consumer surplus for both naïfs and sophisticates. Consequently, the welfare of sophisticated users is lower than in the base case without the platform. Put differently, as already noted, sophisticates' welfare is impacted purely because of the negative pecuniary externality the naïfs create on them, working through the product price. In fact, recall that in this case sophisticated users do not even participate in the platform (and thus do not enjoy the entertaining content or receive ads), so the only impact on their welfare relative to the environment without the platform is via this price effect.

Another noteworthy feature is that the negative welfare effects occur when there is a larger gap between ϕ_1 and ϕ_0, which corresponds to the informativeness of digital ads. The comparative static of increasing ϕ_0 holding ϕ_1 fixed (or reducing ϕ_1 holding ϕ_0 fixed) then gives the following paradoxical result: the ad-based business model emerges precisely when digital ads are less socially useful.

Remark — We conclude this section by reiterating the role of two assumptions in the sharp welfare result in Proposition 4. The first is the linear-quadratic utility function, which we discussed already. Recall that this utility function implies that the firm and thus the platform cannot extract any surplus from the sophisticated agents. If we adopted a different utility function, expected consumption may increase or decrease with informative digital ads. In this case, the platform may be able obtain additional revenues from sophisticates with the ad-based business model. Nevertheless, the source of our main result—the fact that digital ads generate more revenues from naïve agents—continues to hold in this case, highlighting that the linear-quadratic utility function is just a simplifying assumption. Second, it is important that naïve users become over-optimistic about product quality after seeing digital ads. If they naïvely became over-pessimistic, then the mechanism we emphasize here would not apply. We do not see this as a shortcoming, however: it is plausible for naïve agents to be *de facto* manipulated by ads into believing that advertised products are higher quality than they are in reality, and if their bias was toward thinking that they were on average lower-quality, neither the platform nor firms would have any incentive to use such ads.

4 General Platform Business Models

In this section, we relax the assumption that the platform cannot offer both subscription fees and digital ads. We will see that all of the main insights from the previous section generalize to this case. In Section 3, we allow the platform to offer a menu consisting of multiple plans, each of which is monetized either with advertisements or subscription fee. In Section 4.2, we characterize the equilibrium in the most general case where the platform can offer multiple plans some of which intermix ads and subscription.

4.1 Profit-Maximizing Menus of Business Models

We now allow the platform to offer a menu of multiple plans that specify either an ad load α_ℓ or a subscription price P_ℓ (but *not both* until the next subsection). Each user is then allowed to self select into one of these plans. We first observe that, with fully-rational users, Proposition 3 still applies and the unique profit-maximizing strategy is to offer a single subscription-based business model with $P^* = T - v$.

We next consider the case where there are both sophisticated and naïve agents. Our main result is provided in the following proposition.

Proposition 5. *When the platform is allowed to offer a menu of plans, there exists $\phi_0^*(\lambda, \phi_1, \phi_{0,N}) < \hat{\phi}_0(\lambda, \phi_1, \phi_{0,N})$ (where $\hat{\phi}_0(\lambda, \phi_1, \phi_{0,N})$ is the threshold characterized in Proposition 2) such that*

(a) If $\phi_0 < \phi_0^*(\lambda, \phi_1, \phi_{0,N})$, the profit-maximizing business model is subscription-based, with $P^* = T - v$, and the equilibrium product price is $p^* = \bar{p}^*$.

(b) If $\phi_0 > \phi_0^*(\lambda, \phi_1, \phi_{0,N})$, the profit-maximizing business model involves a menu consisting of a subscription-based plan with $P^* = T - v$, an ad-based plan with ad load $\hat{\alpha}^*$, and the equilibrium product is price $p^* = \hat{p}^*$.

Moreover, $\phi_0^*(\lambda, \phi_1, \phi_{0,N})$ is increasing in λ and $\phi_{0,N}$, and decreasing in ϕ_1.

Recall that in Proposition 2, the separating equilibrium took the form of sophisticated agents being completely excluded from the platform. This was a consequence of the restriction that the platform could offer either a subscription fee or digital ads, but not both. Now that such a menu is feasible, the platform can always attract the sophisticates with a subscription-based service. This does not impact the profit-maximizing plan offered to naïve users, which remains the same as in Proposition 2(b). The key spillover from naïfs to sophisticates identified in Proposition 2 is present here as well. When naïve agents receive the ad-based plan (with high ad load), this increases their demand for the product and the monopoly price of the firm, which then hurts sophisticates.

Interestingly, the threshold $\phi_0^*(\lambda, \phi_1, \phi_{0,N})$ is still increasing in λ, despite the platform being able to perfectly segment naïfs and sophisticates. This is because the amount of advertising revenue the platform can extract from the firm is (convex) quadratic in the fraction of naïve agents, whereas the subscription revenue is linear. Thus, as more naïve agents participate in an ad-based platform, the more profit the firm generates from *other* naïve agents, because with more naïve agents, it can charge higher prices. To see the intuition more clearly, consider the case where there are very few naïve agents, in which case the firm will set the same monopoly price as in the baseline with no advertising. As the fraction of naïve agents increases, this will have a direct positive effect on firm profits, as naïfs consume more of the product, and it will also have a positive indirect effect, because the firm can now further increase its price in order to take advantage of these naïfs. A different interpretation of this result is that a higher fraction of sophisticates in the population provides some protection for naïve agents—by discouraging higher prices—even though the two types of agents participate in different plans.

Corollary 1.

(a) *When* $\phi_0 < \phi_0^*(\lambda, \phi_1, \phi_{0,N})$*, user welfare is* $\hat{W}^*(\tau) = W_{\textit{fully-rational}}(\tau) = W_{\textit{base}}(\tau)$ *for both* $\tau \in \{S, N\}$*.*

(b) *When* $\phi_0 > \phi_0^*(\lambda, \phi_1, \phi_{0,N})$*, user welfare is* $\hat{W}^*(\tau) < W_{\textit{fully-rational}}(\tau) = W_{\textit{base}}(\tau)$ *for both* $\tau \in \{S, N\}$*.*

Therefore, the welfare results from Section 3.3 extend immediately to the setting where the platform can offer menus to their users. Even though sophisticates are now on the platform, they are still held down to their outside option, so when $\phi_0 < \phi_0^*(\lambda, \phi_1, \phi_{0,N})$, welfare effects are analogous to those in Proposition 4. Even more importantly, when $\phi_0 > \phi_0^*(\lambda, \phi_1, \phi_{0,N})$, both types of agents are pushed to levels of welfare worse than the benchmark without the platform at all—just as in Proposition 4. As shown in Corollary 1, more informative digital ads generally lead to lower welfare. Specifically, a higher ϕ_1, while holding ϕ_0 constant, which corresponds to digital ads being more informative, reduces welfare. The intuition for this comparative static is as follows: as digital ads become more informative, the platform is more likely to adopt an ad-based plan, but for the reasons we already identified in Section 3, digital ads reduce welfare for both sophisticated and naïve agents.

A final important observation is that $\phi_0^*(\lambda, \phi_1, \phi_{0,N})$ is lower than the threshold characterized in the previous section, $\hat{\phi}_0(\lambda, \phi_1, \phi_{0,N})$. This implies that for values of ϕ_0 between these two thresholds, the equilibrium in the previous section was subscription-based, whereas in the current section, it involves digital ads targeted at naïve agents. Intuitively, since the platform could not previously segment the market between sophisticated and naïve agents without excluding the former, it was less inclined to adopt an ad-based model. However, with the ability to offer a menu of plans, the platform is now more willing to target naïve agents with ads. This means that sophisticated agents, who previously provided a form of protection for naïve agents, no longer do so when this type of segmentation is possible. This shift has significant implications for welfare, as shown in the next proposition.

Proposition 6. *Consumer welfare* $\hat{W}^*(\tau)$ *is monotonically decreasing in both* ϕ_0 *and* ϕ_1*, for both user types* $\tau \in \{S, N\}$*.*

Unsurprisingly, a higher ϕ_0—which leads to more false positives for naïfs—leads to reductions in *ex post* utilities because it gives the platform more leverage to *de facto* manipulate these users. As before, it also inflates demand and induces a higher monopoly price for the firm, which indirectly reduces the welfare of sophisticated agents as well. Perhaps more surprising is that a higher ϕ_1 also leads to an unambiguous reduction in welfare. This result works through to distinct channels. First, as we have already noted, greater ϕ_1 can induce a switch from a subscription-based business model to a mixed one where naïve users receive digital ads and thus both types of agents have lower *ex post* utilities (as in Proposition 2). Second, when the platform chooses the ad-based business model, it always extracts maximal surplus from naïfs, and this implies in particular that any surplus from the greater informativeness of ads is captured by the platform.

4.2 Mixed Business Models

The findings of Section 4.1 readily generalize to the case where the platform can offer a richer set of plans that mix subscription and advertising. In other words, we can allow the platform to either offer a single plan (α^*, P^*) to all users or to offer two plans (α_1^*, P_1^*) and (α_2^*, P_2^*)—there is never any strict gain from offering three plans. We consider these more general mixed business models throughout the remainder of the paper. The next proposition characterizes the equilibrium in this more general case.

Proposition 7. *There exists* $\tilde{\phi}_0(\lambda, \phi_1, \phi_{0,N}) < \hat{\phi}_0(\lambda, \phi_1, \phi_{0,N})$ *such that*

(a) If $\phi_0 < \tilde{\phi}_0(\lambda, \phi_1, \phi_{0,N})$*, the profit-maximizing business model offers two (not necessarily distinct) plans* (α_1^*, P_1^*) *and* (α_2^*, P_2^*)*, where* $\alpha_S^{FB} \leq \alpha_2^* < \hat{\alpha}^*$ *with the implied welfare level for users being* $\tilde{W}^*_{(a)}(\tau) \leq W^*_{\textit{fully-rational}}(\tau) = W^*_{\textit{base}}(\tau)$ *for both* $\tau \in \{S, N\}$*.*

(b) If $\phi_0 > \tilde{\phi}_0(\lambda, \phi_1, \phi_{0,N})$*, the profit-maximizing business model offers a subscription-heavy plan with* $P^* \geq T - v$*, and an ad-based plan with ad load* $\hat{\alpha}^*$ *(and no subscription fee). The welfare levels for users are* $\tilde{W}^*_{(b)}(\tau) < W^*_{\textit{fully-rational}}(\tau) = W^*_{\textit{base}}(\tau)$ *for both* $\tau \in \{S, N\}$*.*

Moreover, $\tilde{\phi}_0(\lambda, \phi_1, \phi_{0,N})$ *is increasing in* λ *and* $\phi_{0,N}$*, and is decreasing in* ϕ_1*.*

The business model choice of the platform in Proposition 7 is analogous to that Proposition 5, but also richer. Once again, like in Proposition 5, the profit-maximizing business model turns on the rate of false positives, as regulated by the parameter ϕ_0. When ϕ_0 is smaller than the threshold $\tilde{\phi}_0(\lambda, \phi_1, \phi_{0,N})$ (regime (a)), the platform offers a plan that has less advertising than $\hat{\alpha}^*$, which is the amount that makes naïve agents indifferent between participating and not participating on the platform. This is because the ad technology's power to manipulate naïve users' consumption behavior is weaker, and the platform prefers to collect more subscription fees from both types of users and consequently chooses a lower ad load. Once ϕ_0 exceeds the threshold $\tilde{\phi}_0(\lambda, \phi_1, \phi_{0,N})$ (regime (b)), then the platform can generate more revenue from maximally advertising to naïfs, capturing all of the surplus that the content on the platform generates for them (while still collecting subscription fees from sophisticates). This situation leads to the lowest user utilities for both naïve and sophisticated agents, with the impact on the latter being once more driven by the spillovers through the product price.

Furthermore, $\tilde{\phi}_0(\lambda, \phi_1, \phi_{0,N})$ is again increasing in λ, but now for very different reasons than in Proposition 2. In particular, sophisticates are no longer excluded from the platform. But the monopolist firm still charges a single price to both sophisticates and naïfs. This implies that when more of the population is sophisticated and views fewer digital ads, the firm must price more to this audience and, consequently, can extract less from naïfs who are facing a higher ad load. This makes the ad-based regime (b) equilibrium of Proposition 7 less attractive to the platform. As a result, sophisticates are again protecting naïve agents from business models that specifically target them—thanks to the feature that these ads will not lead to as high prices when only a small fraction of the population is being manipulated. Sophisticates also act as a disciplining force on the platform through the incentive-compatibility (IC) constraints: to keep naïfs from switching to the sophisticate plan, the platform can't push ads too hard on the ad-heavy plan, which then indirectly protects naïfs. If sophisticates become scarce or exit the platform, this disciplining role weakens: the platform shifts the menu toward exploiting naïfs, increasing ad load on the ad-supported plan, so naïf welfare falls.[12] Finally, $\tilde{\phi}_0(\lambda, \phi_1, \phi_{0,N})$ is

[12] This effect is even stronger if the platform can tell who is naïf and who is sophisticated: it can then offer a plan with a higher ad load to naïfs and a high-price, low-ad plan to sophisticates, without needing to satisfy individual IC constraints. With platform-level discrimination based on sophistication type, the platform no longer has to keep offers attractive enough to hold both groups with the same terms, so the ad-heavy business model becomes optimal in a wider range of environments, hurting naïfs through more ads and sophisticates through higher subscription prices.

again decreasing in ϕ_1, so that high-ad load plans are chosen for naïfs when ads are more informative. This corroborates the findings of Proposition 4, further reinforcing the idea that more informative advertising does not necessarily lead to better outcomes for the users.

5 Firm-Level and Platform-Level Competition

In this section, we show that the insights emphasized so far generalize to an environment in which there are multiple platforms and firms and that the fundamental reason for this is related to the presence of digital ads, which soften competition. We first establish that in a generalized version of our model with multiple firms and platforms, there exists a unique (robust) equilibrium. In Section 5.2, we study the case with multiple firms in greater detail, and subsequently in Section 5.3, we study competition between multiple platforms.

5.1 Existence and Uniqueness

We extend our model in Section 2 to allow for $N \geq 1$ firms and $M \geq 1$ platforms. At $t = 1$, each one of the platforms $\rho \in \{1, \ldots, M\}$ makes contract proposals to each one of the N firms (not offering a contract to a subset of these firms is a special case).

Because there can be uninteresting multiple equilibria based on coordination between firms, we adopt two less standard features. First, we assume that firms set their prices sequentially rather than simultaneously. As in standard voting models, this refinement eliminates equilibria supported by weakly-dominated strategies and the exact sequence in which firms make their offers will turn out to be irrelevant (e.g., see Moldovanu and Winter (1995)). Second, we impose a robustness refinement, which we define and explain below.[13]

The exact timing is as follows:

- At $t = 1.1$, each platform ρ simultaneously offers menus $\{(\alpha_{1,\rho}^{(j)}, P_{1,\rho}^{*}), (\alpha_{2,\rho}^{(j)}, P_{2,\rho}^{*}), m_{\rho}^{(j)}\}_{j=1}^{N}$ of entertainment plans to every firm j to advertise at load $\alpha_{\ell,\rho}^{(j)}$ for product j while charging subscription price $P_{\ell,\rho}^{*}$ if the user selects plan ℓ on its platform. The variable $m_{\rho}^{(j)}$ specifies the total amount transferred from firm j to the platform ρ, if accepted. Notice also the restriction that the subscription fees associated with each plan offered by platform ρ is the same across all firms.

- At $t = 1.2$, each of the firms $j \in \{1, \ldots, N\}$ accepts or rejects these proposals (a firm can accept multiple proposals simultaneously). If the proposal from platform ρ is accepted by firm j, advertisement rates $(\alpha_{1,\rho}^{(j)}, \alpha_{2,\rho}^{(j)})$ for firm j's product is implemented and firm j transfers $m_{\rho}^{(j)}$ to the platform ρ. The platform also charges subscription fees $P_{1,\rho}^{*}$ and $P_{2,\rho}^{*}$ as promised. If the proposal is rejected, the platform collects no transfers but can advertise at whatever rate it likes.

[13] In particular, the intuition for multiple equilibria in this case is similar to those that arise in voting models, whereby everybody else voting for a less preferred outcome renders it a weak best response for each voter to do so as well. Here, too, despite product differentiation, there can be multiple equilibria whereby each firm sets a very low price expecting the other firms to set a very low price. In voting models, sequential actions are sufficient to restore uniqueness (and do so in an order-independent manner). Here, due to the more complicated nature of the game, we need to impose one more robustness refinement in order to achieve the same objective, as explained below.

- At $t = 2$, all firms set their price p_j^* for the product sequentially. In particular, without loss of any generality, we assume that first firm 1 sets its price first, followed by firm 2, and so on.

- At $t = 3$, users decide which platform and plan to participate in, if any. Formally, user i chooses $x_{i,\ell,\rho} \in \{0, 1\}$ for all ρ and ℓ with $\sum_{\ell=1}^{2} \sum_{\rho=1}^{M} x_{i,\ell,\rho} \leq 1$, so that she can participate at most in one plan.

- At $t = 4$, users enjoy the platform content and watch the ads on the platform in which they participate. We assume that the probability of each ad appearing is independent across users and across multiple ads seen by the same user.

- At $t = 5$, user i makes purchasing decisions to maximize her utility

$$\mathbb{E}^i \left[\beta \sum_{j=1}^{N} \theta_i^{(j)} z_i^{(j)} - \frac{\left(\sum_{j=1}^{N} z_i^{(j)} \right)^2}{2} - \sum_{j=1}^{N} p_j z_i^{(j)} \right],$$

  where $\mathbb{E}^i$ denotes this user's expectation given her type and $\theta_i^{(j)}$ is drawn i.i.d. according to the distribution in Section 2.

In this section, we focus on *robust* (perfect) Berk-Nash equilibria. Berk-Nash are defined exactly analogously to before, so here we only explain our robustness notion, which is used to eliminate various equilibria that may emerge due to miscoordination. Let us consider an ε-variant of our game where prices for all firms have to belong to the discrete grid, $p_j \in \{0, \varepsilon, 2\varepsilon, \ldots\}$ for some $\varepsilon > 0$. Define $\mathcal{E}(\varepsilon)$ as the set of Berk-Nash equilibria of this discretized game. We say that equilibrium $\{p_j^*\}_{j=1}^N$ is a *limit equilibrium* if there exists a sequence $\{\varepsilon_n\}_{n=1}^{\infty}$ with $\lim_{n\to\infty} \varepsilon_n = 0$ and a sequence of (Berk-Nash) equilibria $\{p_j^*(\varepsilon_n)\}_{j=1}^N \in \mathcal{E}(\varepsilon_n)$ where $\lim_{n\to\infty} p_j^*(\varepsilon_n) = p_j^*$ for all firms j. We say that an equilibrium is *robust* if it is an equilibrium of our original game (with continuous pricing decisions) and also a limit of the equilibria of the discretized game. We say that a (robust perfect Berk-Nash) equilibrium is *essentially unique* if the resulting allocation in terms of advertisements, consumption levels and expected payoffs are uniquely pinned down (though the strategies may not be unique and, if there are any asymmetries in the equilibrium, the identity of firms or platforms taking one type of action versus another may be non-uniquely determined).

Proposition 8. *Generically, there exists an essentially unique robust (perfect) Berk-Nash equilibrium for any number of firms N and platforms M, which is independent of the sequence of pricing decisions at $t = 2$.*

Proposition 8 establishes the existence of an essentially unique robust equilibrium. There can be non-generic multiplicities when users are indifferent between two or more products, but this possibility occurs only on a set of measure zero values of the realized signals. This proposition is useful for us because it pins down the equilibrium allocation uniquely and thus simplifies our description of comparative statics and welfare with multiple firms and platforms. We use this result in the next two subsections.

Throughout the competition analysis we assume single-homing: each user chooses one platform and plan, while platforms simultaneously choose menus (plan-level ad load and subscription prices). Types are again unobserved, so the same IC constraints are present. As we will see, competition disciplines ad load choices because users can switch platforms in response to heavy ad loads. Despite this, platforms are again able to separate naïve and sophisticated users, and moreover, advertising endogenously differentiates user beliefs and this endogenous product differentiation is the reason why competition is softened and markups remain high. We conclude from this analysis that our results are due to the interplay of naïve and sophisticated agents and digital ads—not due to lack of competition.[14]

5.2 Firm Competition

To isolate the effects of firm-level competition on equilibrium and welfare, we focus on the case of $N = 2$ and $M = 1$. Each of the firms has a single product (referred to as product 1 and product 2). The two products are *ex ante* identical: they have the same (independent) probabilities of being high-quality or low-quality for each user. If competition between the two firms were at this *ex ante* stage, Bertrand competition would drive prices down to marginal cost. However, once some users start receiving additional information from advertisements, the products become horizontally differentiated—some users will have signals about the quality of either or both of the products' quality for them.

The platform can advertise for both products, and we denote the advertising rates in plan ℓ by $\alpha_\ell^{(1)}$ and $\alpha_\ell^{(2)}$ for the two products, respectively. This implies that if a user subscribes to plan ℓ, then with probability $1 - e^{-\alpha_\ell^{(1)}T}$ she views an ad for product 1 and with probability $1 - e^{-\alpha_\ell^{(2)}T}$ she views an ad for product 2. Since these events are independent, it is possible that she may view both ads or neither. At the same time, the platform charges subscription fees $\{P_\ell^*\}_{\ell=1}^2$. The platform offers simultaneous contracts $\{(\alpha_\ell^{(1)}, m_\ell^{(1)})\}_{\ell=1}^2$ and $\{(\alpha_\ell^{(2)}, m_\ell^{(2)})\}_{\ell=1}^2$ to both firms who then either accept or reject the contracts.

To build some intuition, we first discuss the fully-rational benchmark and the equilibrium prices and quantities when there is no advertising.

Fully-Rational Benchmark. Under the fully-rational benchmark ($\phi_{0,N} = \phi_0$), the platform offers a single plan, and each agent i subscribes to this plan and holds Bayesian beliefs that firm j's quality for her is $\theta_i^{(j)} = 1$ denoted by $\pi_S^{(j)}$ (given the information she has received from digital ads, if any). Each user then chooses $j^* \in \arg\max_j \beta\pi_i^{(j)} - p^{(j)*}$ and consumes a quantity $z_i^{(j^*)^*} = \beta\pi_i^{(j^*)} - p^{(j^*)*}$ of product j^* and none of the other product, i.e., $z_i^{(-j^*)^*} = 0$. We let $\bar{p}_1^*$ and $\bar{p}_2^*$ denote the prices offered under the fully-rational benchmark in equilibrium. Note that this fully-rational benchmark may or may not involve some amount of advertising in the plan. If there is no advertising in the single plan, then equilibrium prices are equal to marginal cost—$\bar{p}_1^* = \bar{p}_2^* = c$—because both products have the same *ex ante* appeal to all users.

[14] Our analysis of competition is non-cooperative: platforms choose ad intensity and menu prices as unilateral best responses, as firms do for product prices. There is no coordination device or dynamic punishment scheme. This differs from algorithmic collusion models such as Calvano et al. (2020), where algorithms learn to respond to rivals' prices and can sustain elevated markups. In our model, parallel movements in ad load and price across platforms stem from how digital ads soften price competition across products, not collusive conduct.

Benchmark with No Advertising. When there is no advertising plan offered (or selected) in general, then prices are again equal to marginal cost. With no advertisements, we have $\pi_S^{(1)} = \pi_S^{(2)} = q$, since *ex ante* the two products are identical. Therefore, the two firms will compete à la Bertrand and in equilibrium will charge prices equal to marginal cost, i.e., $p_1^* = p_2^* = c$. In this benchmark, therefore, competition from multiple firms reduces prices and increases consumer surplus relative to the monopoly scenario. We will next see that the situation is radically different in the presence of digital ads.

Equilibrium in the Full Model. We now consider our full model with both sophisticated and naïve agents, in the presence of digital ads. The first consequence of digital ads is that, in general, $\pi_S^{(1)} \neq \pi_S^{(2)}$ because (i) the user may view an ad for product 1 but not product 2 or vice-versa, and (ii) the user may receive different signals from product 1's ad versus product 2's ad. Hence, there will be (partial) differentiation between the products for different segments of the population. As a consequence of this differentiation, firms have market power and equilibrium prices will typically satisfy $p_1^* > c$ and $p_2^* > c$, as we show next.

Proposition 9. *There exist $\phi_1^{\mathcal{F}}, \phi_0^{\mathcal{F}}(\phi_1) \in [0,1]$ such that*

(i) If $\phi_1 \leq \phi_1^{\mathcal{F}}$, the platform offers a single subscription-based plan with $P^* = T - v$ and no advertising $\alpha^{(1)*} = \alpha^{(2)*} = 0$ and equilibrium prices are $p_1^* = p_2^* = c$.

(ii) If $\phi_1 > \phi_1^{\mathcal{F}}$ and $\phi_0 \geq \phi_0^{\mathcal{F}}(\phi_1)$, the platform offers a subscription-based plan with $P_1^* = T - v$ and no advertising $\alpha_1^{(1)*} = \alpha_1^{(2)*} = 0$, and an ad-based plan ($\alpha_2^{(1)*} + \alpha_2^{(2)*} > 0$). Equilibrium product prices are $\hat{p}_1^* > \bar{p}_1^*$ and $\hat{p}_2^* > \bar{p}_2^*$.

Proposition 9 extends our main characterization result, Proposition 5, to the case with multiple firms. It establishes that, in contrast to the situation without advertising, market power in the product market and prices above marginal cost can persist, despite the presence of multiple firms. This is because of endogenous horizontal differentiation created by digital ads. Digital ads thus become even more useful to firms in this environment with competition than in our benchmark setup, as the next example illustrates, and in fact, with more digital ads, equilibrium prices can be higher with competition than without.

Example 1. Let $\phi_{0,N} < \phi_0 = \phi_1$ and suppose $\phi_0 = \tilde{\phi}_0 - \epsilon$ for small ϵ, where $\tilde{\phi}_0$ denotes the threshold in Proposition 7. Therefore, with a monopolist firm, the equilibrium involves no digital advertising, and the equilibrium price for the product is $\bar{p}^* = (\beta q + c)/2$.

Consider next the same setup in the presence of a second firm and also suppose that a sufficiently large fraction of the population is naïve ($\lambda < \underline{\lambda}$ for some $\underline{\lambda}$, which we take to be small in this example for simplicity). In this case, firm j can capture positive market share by advertising and charging a price of $c + \beta(\pi_N^{(j)}(G) - q)$, where $\pi_N^{(j)}(G)$ is the belief of a naïve agent after receiving a positive ad signal. In essence, firm j can now generate a profit from advertising because ads relax competitive pressures. The resulting equilibrium will be separating as in part (ii) of Proposition 9: a subscription plan with $P^* = T - v$ is offered for sophisticates and an ad-based plan with $\alpha^{(1)} + \alpha^{(2)} > 0$ is offered for naïfs. This example illustrates not only the possibility that there will be more digital advertising with competition, but also shows that equilibrium prices could in fact be higher with competition than without. In particular, when $\pi_N^{(j)}(G) > 3q/2$, the equilibrium price is above the monopoly price $\bar{p}^*$, because advertising for naïve users increases firms' market power.

A major difference between Propositions 5 and 9 is also worth noting. In Proposition 9, not just the rate of false positives, ϕ_0, but also the informativeness of positive signals, ϕ_1, is critical for the equilibrium business model, advertising, and prices. In essence, ϕ_0 still matters for the same reasons (digital ads are most useful to firms and the platform because they *de facto* manipulate naïve agents). But now ϕ_1 needs to be sufficiently large as well, since otherwise digital ads do not generate sufficient differentiation between products and, as a result, do not allow firms to charge high enough markups.

When both ϕ_0 and ϕ_1 are above their respective thresholds, the equilibrium resembles part (b) of Proposition 5: sophisticates, who recognize that both parameters are large, understand that most ads give positive signals regardless of the true underlying $\theta_i^{(j)}$, which means their demand responds little to signals from ads, and this makes it profitable for the platform to segregate the market and offer a service with subscription fees and no ads to sophisticates. In contrast, naïfs continue to be responsive to positive signals from ads and are the main targets for these ads. Because firms know that naïfs will view their ads and will have differentiated willingness to pay for these products, they charge prices above marginal cost, which then creates a negative spillovers on sophisticates, as before. The platform extracts the value of entertaining content from sophisticates using a subscription fee and extracts the value of digital ads targeting naïve agents from both firms.

We next discuss utility in the presence of firm competition. Let $\bar{W}_{2,1}^*(\tau)$ denote the *ex post* utility of users of type $\tau \in \{S, N\}$ under the fully-rational benchmark with two firms and a single platform. Define $\hat{W}_{2,1}^*(\tau)$ analogously as the *ex post* welfare of the two types of users in the equilibrium characterized in Proposition 9. Our main welfare result in this section is:

Proposition 10.

(i) If $\phi_1 \leq \phi_1^{\mathcal{F}}$, then $\hat{W}_{2,1}^*(\tau) = \bar{W}_{2,1}^*(\tau)$ for both $\tau \in \{S, N\}$.

(ii) If $\phi_1 > \phi_1^{\mathcal{F}}$ and $\phi_0 \geq \phi_0^{\mathcal{F}}(\phi_1)$, then $\hat{W}_{2,1}^*(\tau) < \bar{W}_{2,1}^*(\tau)$ for both $\tau \in \{S, N\}$.

The first part of this proposition simply reiterates that when the platform opts for a subscription-based plan and there are no digital ads, Bertrand competition between the two firms will drive prices down to marginal cost. This yields the same welfare to both types of agents as in the fully-rational benchmark—which, recall, also features no digital ads and positive subscription fees. This is because, without digital ads, naïve and sophisticated agents behave identically and all products are identical in the eyes of all consumers, driving their prices down to marginal cost.

The second part of the proposition is more interesting. It shows that, analogously with Proposition 4, when both ϕ_0 and ϕ_1 are sufficiently large, the equilibrium involves digital ads targeted to naïve users, and these digital ads lead to higher prices—rather than prices being equal to marginal cost.

The intuition for higher prices is more nuanced in this case, as already hinted at in Example 1. In Proposition 4, digital ads increased equilibrium prices because they raised naïve users' willingness to pay for the product. Here, digital ads do not just inflate naïve users' valuations, but also relax competition between the two firms by generating endogenous differentiation. This is the reason why the standard Bertrand logic does not apply and competition does not protect naïve agents from *de facto* manipulation from firms and the platform. As a result of these forces, welfare is still substantially lower than that in the fully-rational benchmark.

The comparative statics of Propositions 9 and 10 are also interesting. As before, a higher rate of false positives, ϕ_0, makes digital ads more likely, which reiterates the same result that digital ads are more likely to emerge when they are less informative. But now we also have digital ads being more likely when the informativeness of positive signals from ads, ϕ_1, is higher, because with low ϕ_1, digital ads are not impactful on user valuations and the subscription-based business plan becomes more likely.

5.3 Platform Competition

We now discuss the implications of between-platform competition and to simplify the analysis, this time we focus on the case where there are two platforms and a single firm, i.e., $M = 2$ and $N = 1$. The two platforms simultaneously offer plans $\{\alpha^*_{1,\ell}, P_{1,\ell}\}_{\ell=1}^2$ with associated transfer m_1 to the firm and $\{\alpha^*_{2,\ell}, P_{2,\ell}\}_{\ell=1}^2$ with associated transfer m_2. Following this stage, the firm decides which plan(s), if any, to accept. The next proposition characterizes the equilibrium in this case. We again denote the price for the unique final good in the corresponding fully-rational benchmark by $\bar{p}^*$.

Proposition 11. *There exists $\phi_1^{\mathcal{P}}$ such that:*

(i) If $\phi_1 < \phi_1^{\mathcal{P}}$, the platforms offer competing plans with no advertising and no subscription fee, and the product is priced at $p^* = \bar{p}^*$.

(ii) If $\phi_1 > \phi_1^{\mathcal{P}}$, the platforms offer two ad-based plans $\alpha_1^* = \alpha_S^{FB}$ and $\alpha_2^* \in (\alpha_S^{FB}, \hat{\alpha}^*)$ with no subscription fees, and the product is priced at $p^* = \hat{p}^*_{\mathcal{P}} > \bar{p}^*$.

The results in this proposition are even more striking than those in Proposition 9. Our results turn on the informativeness of positive signals from ads, ϕ_1. If this parameter is below the critical threshold $\phi_1^{\mathcal{P}}$, platforms are able to extract no surplus from either the firm or the users, and completely forgo digital ads but also do not charge a subscription fee, because competition between them drives their prices down to marginal cost—which is equal to zero for the entertaining content that they offer.

In contrast, when ϕ_1 is greater than $\phi_1^{\mathcal{P}}$, the platform once again uses digital ads, targeted at naïve users, who interpret positive signals from ads as evidence of high quality, which inflates their valuation for the product. The reason why there are no subscription-based plans is that, with such plans, each platform can always undercut the other's subscription fee, and this drives down equilibrium subscription fees to zero, making them unprofitable. Equally importantly, both platforms compete for naïve users by segmenting the market between them and the sophisticates, and the naïfs again receive more frequent digital ads ($\alpha_2^* > \alpha_S^{FB}$), which further inflates their valuations for the product and the profit-maximizing monopoly price of the firm (analogously to the situation in Proposition 5). Platforms are able to extract surplus from naïfs, because these users have a greater willingness to pay for digital ads and once we are in an ad-based equilibrium, reducing digital ads would be less attractive for naïve users (subscription fees are equal to zero already). This softens competition between platforms and enables them to offer a high load of digital ads and make profits.

We next study the welfare properties of this equilibrium. Similarly to before, let $\bar{W}^*_{1,2}(\tau)$ denote *ex post* utility in the fully-rational benchmark (now with a single firm and two platforms) for $\tau \in \{S, N\}$, and $\hat{W}^*_{1,2}(\tau)$ denote *ex post* utilities in the equilibrium of Proposition 11 for $\tau \in \{S, N\}$.

Proposition 12.

(i) If $\phi_1 < \phi_1^{\mathcal{P}}$, then $\hat{W}^*_{1,2}(\tau) = \bar{W}^*_{1,2}(\tau)$ for both $\tau \in \{S, N\}$.

(ii) If $\phi_1 > \phi_1^{\mathcal{P}}$, $\hat{W}^*_{1,2}(\tau) < \bar{W}^*_{1,2}(\tau)$ for both $\tau \in \{S, N\}$.

Analogously to the case of firm-level competition analyzed in Propositions 9 and 10, when there are no digital ads, welfare is restored back to the fully-rational benchmark because platforms compete with each other and this limits their ability to extract surplus from users. In contrast, as soon as ϕ_1 is above the threshold $\phi_1^{\mathcal{P}}$, digital ads targeted at naïfs reappear, softening the competition between platforms. In particular, as explained above, once business models use digital ads, the two platforms no longer undercut each other on ad loads.

It is again interesting to note that the second regime, where welfare is low, is more likely when ϕ_1 is high. This is for the same reasons as in Proposition 11: when ϕ_1 is very low, digital ads are not sufficiently appealing to the firm, because they do not expand the demand for its product sufficiently, and only a subscription-based business model can be sustained in equilibrium.[15]

6 Digital Ad Taxes

Our analysis so far has identified a fundamental market failure in platform economies with digital ads and naïve agents who may misinterpret the signals from these ads. A natural question is whether there are regulatory or tax-based solutions to this market failure. This is the question we investigate in this section. While the first-best allocation is not implementable, we show that the second best, where user type is private information and users will choose which plain to join themselves, can be decentralized through nonlinear digital ad taxes and product price and subscription fee subsidies. We also show that a linear tax on digital advertising revenues improves welfare in the decentralized equilibrium. For simplicity, we present all results this section for a single platform and single firm. These results generalize to multiple platforms and multiple firms, as we explain in Appendix C.[16]

6.1 Second-Best User Welfare

Recall that our first-best user welfare, W_{FB}, was obtained by allowing the social planner to fully control the allocation, while observing user types. This meant that there were no self-selection or IC con-

[15] Our welfare analysis focuses on the case of exogenous platform content. As in Anderson and Coate (2005), competition can also affect content quality, in which case welfare effects could become ambiguous. Nevertheless, even with endogenous content quality, competition does not prevent digital ads from *de facto* manipulating naïve agents, is softened by digital ads, and does not necessarily remove markups.

[16] Our policy analysis is qualitative. Calibrating the form and values of second-best digital ad taxes would require much more detailed data and further functional form assumptions.

straints, and the planner could freely choose the level of digital ads different types of users would observe. This is not feasible when user type is private information, and a more natural benchmark is the one where the social planner has to obey users' IC constraints, determining in which plan they prefer to participate.

We define the *second-best* as follows: we assume that the planner can choose a menu of advertisement levels, transfers and product prices, subject to the IC constraints of different types of users. As before, it is sufficient to restrict attention to two menus, represented by (α_1, P_1, p_1) and (α_2, P_2, p_2). Here, α is the ad load, P represents a transfer from the user (equivalent to the subscription fee) and p is the price the user faces to buy the product. This formulation thus allows a user's consumption level for the product and resulting payments to depend on which menu she chooses, and is simplified by specifying that these take the form of linear prices (p_1 and p_2) at which the user can purchase as many units of the product as she desires given the realization of the signals from the ads (which are also her private information).

The IC constraint for user i with type τ_i to select plan ℓ^* can then be written as

$$
\begin{aligned}
&(1-\alpha_{\ell^*})T - P_{\ell^*} + q\mathbb{E}^{\tau_i}_{\pi_i}[U(z_i^*(\pi_i, p_{\ell^*}); \theta_i = 1) - p_{\ell^*} z_i^*(\pi_i)|\alpha_{\ell^*}] + (1-q)\mathbb{E}^{\tau_i}_{\pi_i}[U(z_i^*(\pi_i, p_{\ell^*}); \theta_i = 0) - p_{\ell^*} z_i^*(\pi_i)|\alpha_{\ell^*}] \\
\geq &(1-\alpha_{\ell})T - P_{\ell} + q\mathbb{E}^{\tau_i}_{\pi_i}[U(z_i^*(\pi_i, p_{\ell}); \theta_i = 1) - p_{\ell} z_i^*(\pi_i)|\alpha_{\ell}] + (1-q)\mathbb{E}^{\tau_i}_{\pi_i}[U(z_i^*(\pi_i, p_{\ell}); \theta_i = 0) - p_{\ell} z_i^*(\pi_i)|\alpha_{\ell}]
\end{aligned}
$$

for all other plans ℓ.

As already anticipated in the discussion following Proposition 1, assigning naïve users to the advertising load $\alpha_N^{FB} < \alpha_S^{FB}$, as the planner wishes to do, is not incentive compatible because naïfs prefer an even higher advertising load than α_S^{FB}. In the second best, the planner allows users themselves to decide among these plans. Our next result characterizes the second best.

Proposition 13. *The second best involves a single plan with advertising load* $\alpha^{SB} \in [\alpha_N^{FB}, \alpha_S^{FB}]$. *Whenever* $\alpha_N^{FB} > 0$, *second-best welfare is less than first-best welfare; that is,* $W_{FB}(\tau) > W_{SB}(\tau)$ *for both* $\tau \in \{S, N\}$. *At the same time, average welfare is higher under the second best than under the base case without the platform; that is,* $\lambda W_{SB}(S) + (1-\lambda) W_{SB}(N) > \lambda W_{base}(S) + (1-\lambda) W_{base}(N)$.

The intuition for Proposition 13 is closely related to our discussion of the first best in Proposition 1. Ideally, the planner would like to offer a menu with lower ad load for naïve agents than what will be offered to sophisticates, α_S^{FB}. However, naïfs actually prefer an even higher ad load than α_S^{FB}, because in their assessment ads are more informative than sophisticates consider them to be, and this makes it more attractive for naïfs to trade-off a little less entertainment for more ads starting in the neighborhood of α_S^{FB}. Consequently, whenever the planner offers a menu with different options, naïfs will have a stronger preference for the plan with greater ad load than do the sophisticates—which is the exact opposite of what the planner would like them to do. Hence the planner is forced to choose a single plan. Because this plan will cater to both naïve and sophisticated agents, its ad load is intermediate between α_N^{FB} and α_S^{FB}, trading off the utility of naïve and sophisticated agents.

6.2 Decentralizing the Second Best

To decentralize the second best, we consider a nonlinear tax-subsidy scheme. Let $\zeta(\alpha_1, \alpha_2)$ denote a tax on the platform as a function of digital ad quantities α_1 (in plan 1) and α_2 (in plan 2). It turns out to be sufficient to consider the separable form $\zeta(\alpha_1, \alpha_2) = \tilde{\zeta}(\alpha_1) + \tilde{\zeta}(\alpha_2)$, where each component imposes a zero tax on advertising at or below α^{SB}, but taxes advertising at intensities higher than α^{SB} at the rate $\mu > 0$. More precisely:

$$\tilde{\zeta}(\alpha) = \begin{cases} 0, & \text{if } \alpha \leq \alpha^{SB} \\ \mu(\alpha - \alpha^{SB}), & \text{if } \alpha > \alpha^{SB} \end{cases}$$

At the same time, the planner offers a per-unit product subsidy δ to the firm and a subscription-fee subsidy η to platform to undo monopoly distortions. More specifically, the planner provides a $\delta \int_0^1 z_i^* \, di$ subsidy to the firm (as a function of total quantity sold) and a subsidy the platform conditional on setting zero subscription fees, given by:

$$\begin{cases} \eta \left(\int_0^1 x_{i,1}^* \, di + \int_0^1 x_{i,2}^* \, di \right), & \text{if } P_1^* = 0 \text{ and } P_2^* = 0 \\ \eta \int_0^1 x_{i,1}^* \, di, & \text{if } P_1^* = 0 \text{ and } P_2^* > 0 \\ \eta \int_0^1 x_{i,2}^* \, di, & \text{if } P_1^* > 0 \text{ and } P_2^* = 0 \\ 0, & \text{if } P_1^* > 0 \text{ and } P_2^* > 0. \end{cases}$$

Once this tax-subsidy scheme is set, the rest of the game proceeds as before between the platform, the firm, and the users.

Our next result shows that this policy scheme decentralizes the second best as a (Berk-Nash) equilibrium.

Proposition 14. *There exists $\bar{\mu} > 0$, $\bar{\eta} > 0$, and $\delta^* > 0$ such that if the platform's digital ad tax policy satisfies $\mu > \bar{\mu}$, the firm subsidy is given by δ^*, and the platform subsidy satisfies $\eta > \bar{\eta}$, then the decentralized equilibrium implements the second best.*

In a decentralized equilibrium (with no policy), the platform prefers to advertise to naïfs at a rate higher than α^{SB} because naïfs prefer a higher advertising load than the one the planner would choose for them, and this higher ad load enables the platform to extract more revenue from the firm (which is itself extracting more surplus from naïve users). This excessive use of digital ads and the inflated demand that they induce from naïve users are at the root of the inefficiency of the equilibrium. The second best reduces the load of digital ads by imposing a tax on ad quantities larger than α^{SB}, which restores the equilibrium to the second-best advertising level. At the same time, the subsidy to the firm guarantees that equilibrium product market prices are equal to marginal cost and the subsidy to the subscription fee ensures that the platform sets zero subscription fees and has the correct trade-off between income from digital ads and subscription fees.

6.3 Flat Digital Ad Tax

In this subsection we show that the simpler intervention, consisting of a flat tax on digital ad revenues, improves welfare (though does not restore it to the second-best level). The second-best decentralization in the previous subsection requires nonlinear taxes on digital ad quantities (which may be harder to observe than digital ad revenues) and subsidies to the firm and the platform (which may be difficult to implement). A flat tax on digital ad revenues is a comparatively simpler policy.

More formally, we define a flat digital ad tax as a tax at the rate $\gamma \in (0, 1)$ imposed on total digital ad revenue, which in our model is equal to m.

Proposition 15. *Suppose that the robust Berk-Nash equilibrium without any policy features an ad-based plan. Then there exists $\bar{\gamma} < 1$ such that a flat digital ad tax with $\gamma > \bar{\gamma}$ improves welfare.*

Proposition 15 establishes that, whenever the equilibrium involves an ad-based plan, a sufficiently large flat tax on digital ad revenues improves welfare (without any other policy instrument being used). It does so by discouraging the use of digital ads and encouraging subscription-based plans. Although this flat digital ad tax does not achieve the second best, it is much simpler to implement than the nonlinear tax-subsidy scheme characterized in the previous subsection.

7 Robustness and Extensions

Here, we present four extensions that speak to the scope and robustness of our main results and to several modeling choices we have made so far. Each subsection isolates one aspect—platform manipulation, naïve learning, price discrimination, and price sensitivity—and shows that our core comparative statics and welfare conclusions are preserved. In some of the cases, we also derive additional implications. Throughout this section, we simplify the exposition by focusing on the model of Section 4.2 with a single firm and a single platform.

7.1 Endogenous Manipulation

In this extension, we endogenize manipulation by allowing the platform to choose naïfs' information-sensitivity parameter before setting ad intensity. In other words, we treat "manipulation" as a platform choice rather than a fixed feature of the environment. Concretely, before selecting ad intensity α, the platform first chooses naïfs' information-sensitivity parameter $\phi_{0,N} \in [\underline{\phi}, \phi_0]$. A lower $\phi_{0,N}$ stands in for design decisions that make digital advertising more persuasive (or manipulative) to naïfs, who imperfectly understand the ad technology. One can think of the platform as directly transforming the parameter ω_P, defined in Section 2, which accounts for how the platform's technology utilizes personalized targeting of ads to persuade naïfs. For example, concrete levers for the platform might include content, design, and placement (e.g., creative optimization, ranking, slot prominence, adjacency) and audience selection/routing (lookalike models, propensity targeting), both of which increase the likelihood that a manipulative ad is shown to or is more manipulative for a naïve agent. Formally, we introduce a time period $t = 0$ where the platform first selects $\phi_{0,N}$, before the rest of the game proceeds as in Section 2.

Proposition 16. *A profit-maximizing platform always chooses $\phi_{0,N} = \underline{\phi}$. Moreover, $\hat{W}_{\underline{\phi}}(\tau) \leq \hat{W}_{\phi_0}(\tau)$ for both $\tau \in \{S, N\}$, with the inequality being strict whenever $\phi_0 > \tilde{\phi}_0$ (defined in Proposition 7).*

This result thus implies that the platform will always select the most manipulative environment possible—the lowest $\phi_{0,N}$. It thus confirms that the platform was engaging in *de facto* manipulation in our baseline model. It is also straightforward from Proposition 16 that all of our results apply to this extended environment, with $\phi_{0,N} = \underline{\phi}$.

Two forces make platform profits strictly decreasing in $\phi_{0,N}$. The first, which we refer to as the *pricing channel*, is a consequence of the fact that when $\phi_{0,N}$ is lower, a positive signal shifts naïve demand more, so the downstream firm raises its price and sales. The platform can then claw back the additional firm profit to generate higher advertising revenue. The second force, which we refer to as the *participation channel*, works through the participation margin. Lower $\phi_{0,N}$ raises the perceived interim informational value of ads for naïfs, relaxing their participation constraint and permitting a higher ad load α. These effects reinforce one another—each ad is more valuable and more ads can shown—yielding a monotone improvement in profits as $\phi_{0,N}$ falls.

The main lesson from Proposition 16 also extends to the competition setting of Section 5.3. Platform competition for users reduces the equilibrium ad load via tighter participation constraints but does not alter the choice of $\phi_{0,N} = \underline{\phi}$. Hence competition may limit how many ads are shown, but not how manipulative each ad is.

Remark — Our results extend to richer formulations of user misspecification. For example, if naïfs misperceive either priors or the signal likelihood ratios (e.g., overweighting favorable signals or under-attending to unfavorable ones), the results would be identical, since these models are equivalent to lowering $\phi_{0,N}$. A lower $\phi_{0,N}$ slackens the naïve IC constraint, enlarges the feasible set of (α_N, P_N, P_S) consistent with ad-based business models, and makes the ad-heavy regime more prevalent. As a result, *ex post* utilities for both types weakly decline. In the polar case where naïfs are "easily convinced" ($\phi_{0,N} = 0$), the platform would choose the most manipulative environment and set the highest ad load for naïfs consistent with their participation. The policy implications are also unchanged: instruments that raise $\phi_{0,N}$ (or reduce persuasion) and/or tax ad revenues mitigate the welfare losses caused by ad-based plans.

7.2 Naïve Learning Dynamics

In our baseline model, we considered a one-shot game where naïve agents held a misspecified model of the ad signal process. However, even naïve agents may learn from evidence contradicting their misspecified model, come to understand how targeted ads work, or adjust their behavior after "good-looking" ads repeatedly fail to deliver. Our purpose in this section is to explore the implications of learning by naïve agents.

To maintain tractability, we adopt a simple birth-death environment with endogenous transitions from naïve to sophisticated status. Time is continuous. Agents arrive and leave at the Poisson rate $\mu > 0$. Upon entry, a fraction $\tilde{\lambda}$ are sophisticated and the remainder are naïve. Products arrive at the Poisson rate $\Lambda > 0$. For simplicity we assume that $\phi_{0,N} = 0$ in this subsection, so naïfs believe a good signal necessarily implies that $\theta_i = 1$ (the product is a good fit for them). A naïf converts to sophisticated status when she experiences a positive ad signal that later delivers zero surplus.[17] Let $\lambda(t)$ be the share of sophisticated agents at time t. Then the law of motion is given by

$$\frac{\lambda(t)}{dt} = \mu(\tilde{\lambda} - \lambda(t)) + \alpha_N \Lambda(1-q)\phi_0(1-\lambda(t)),$$

where α_N is the ad load in the naïf's selected plan. This yields a unique steady state fraction of sophisticates:

$$\lambda^* = \frac{\mu\tilde{\lambda} + \alpha_N \Lambda(1-q)\phi_0}{\mu + \alpha_N \Lambda(1-q)\phi_0}. \tag{3}$$

Here, learning is endogenous because the upgrade rate from naïf to sophisticated is driven by the platform's own ad exposure on the naïf plan. The term $\alpha_N \Lambda(1-q)\phi_0$ in the law of motion is exactly the arrival rate of disconfirming experiences: ads arrive at intensity α_N; potential products show up at rate Λ; a false-positive occurs when the product is low match $(1-q)$ yet the ad delivers a good signal with probability ϕ_0. Thus, higher α_N (more ads), higher Λ (more new products), higher ϕ_0 (more good signals when the product is actually bad), or lower q (lower baseline quality) all accelerate learning and increase λ^*. Because α_N is chosen by the platform as part of the menu, a greater ad load simultaneously raises ad revenues today and speeds up the transition of naïfs into sophisticates tomorrow. With positive inflow ($\mu > 0$), the naïve share remains strictly positive in steady state, but is lower when the platform chooses a higher ad load for naïfs. In what follows, we simplify the discussion by assuming that the platform maximizes steady-state profits taking $\lambda = \lambda^*(\alpha_N, \phi_0)$ as a constraint (which is equivalent to the platform's discount rate going to zero). The next proposition provides the main result of this analysis:

Proposition 17.

(i) For any (α_N, ϕ_0), a unique steady-state $\lambda^*(\alpha_N, \phi_0) \in [0,1]$ exists and is given by (3).

(ii) There exists $\bar{\mu} > 0$ such that for $\mu > \bar{\mu}$, Proposition 7 holds identically under cutoff $\tilde{\phi}_0^{dynamic}(\phi_1, \phi_{0,N})$, where $\tilde{\phi}_0^{dynamic}(\phi_1, \phi_{0,N}) \geq \tilde{\phi}_0(\tilde{\lambda}, \phi_1, \phi_{0,N})$.

Learning dynamics thus create a new trade-off. Raising either the objective false-positive rate ϕ_0 or the naïfs' ad load α_N increases demand for the firm and thus platform ad revenues. But it also accelerates the transition of naïfs to sophisticates and raises the steady-state sophisticate share $\lambda^*(\alpha_N, \phi_0)$, reducing future profits. This constrains the ability of the platform to choose very high ad loads, and makes ad-based business models less profitable. Put differently, the dynamic threshold $\phi_0^{\text{dynamic}}(\phi_1, \phi_{0,N})$ is higher than the static threshold $\tilde{\phi}_0(\tilde{\lambda}, \phi_1, \phi_{0,N})$, and hence, for the same parameters, the static equilibrium (without learning) may involve an ad-based business model, while the dynamic one (with learning) leads to a more subscription-based business model.

[17] Formally, naïve agents have an ε-contamined prior over models: with probability $1-\varepsilon$ a naïve agent believes that the subjective signal process in Section 2 is true, and with probability $\varepsilon > 0$ she believes the objective process is true, and we take $\varepsilon \to 0$. She therefore behaves as if the subjective model were true except when she observes disconfirming evidence—$s_i = G$ followed by $\theta_i = 0$—which has non-zero likelihood under the objective model. Following this event, her posterior places unit mass on the objective model, after which she acts as a sophisticate.

We also note that when user turnover (μ) is high, the effects of parameters on the composition of users, $\lambda^*(\alpha_N, \phi_0)$, is small (on the order of $O(1/\mu)$), so they do not overturn the comparative statics from our baseline model.[18] Consequently, the qualitative comparative statics from Proposition 7 continue to hold when we substitute the endogenous share of sophisticates, λ^*, and the impact on prices and welfare also directly generalize.

Additionally, this extension suggests that any policy that accelerates learning by naïve agents, for example, via the use of transparency tools or better feedback, can be welfare-improving by curbing the manipulation of naïfs.

7.3 Price Discrimination

Next, we allow the downstream firm to (imperfectly) observe whether a user is naïve or sophisticated. Specifically, after the platform chooses its business model via the menu $\{(\alpha_\ell, P_\ell)\}_{\ell=1}^2$, the firm observes a buyer's sophistication type $\tau_i \in \{S, N\}$ with probability $\kappa \in [0, 1]$ and remains uninformed with probability $1 - \kappa$. After observing a user's type, the firm can post a type-specific price p_{τ_i}; otherwise it posts a pooled price p. Let $\hat{W}(\tau_i)$ be the *ex post* utility of consumer type τ_i under the equilibrium without price discrimination and $\hat{W}'(\tau_i)$ be her utility when there is price discrimination.

Proposition 18. *With price discrimination, we have:*

(i) $p_N^* \geq p^* \geq p_S^*$.

(ii) Proposition 7 applies identically except the cutoff $\tilde{\phi}_0(\lambda, \phi_1, \phi_{0,N})$ *shifts down to* $\tilde{\phi}_0^\kappa(\lambda, \phi_1, \phi_{0,N}) < \tilde{\phi}_0^0(\lambda, \phi_1, \phi_{0,N}) = \tilde{\phi}_0(\lambda, \phi_1, \phi_{0,N})$.

(iii) Naïfs are worse off with price discrimination ($\hat{W}(N) \geq \hat{W}'(N)$*).*

First, price discrimination allows the downstream firm to move prices toward the type-specific monopoly levels: $p_N^* \geq p^* \geq p_S^*$. That alone increases firm profits from advertisements—because the naïve segment has become more profitable. These profits are shared between the firm and the platform via the transfer m^*, and this makes business models relying on high ad loads more profitable for the platform. This is why the cutoff $\tilde{\phi}_0(\lambda, \phi_1, \phi_{0,N})$ shifts down: relative to the baseline without discrimination, a lower false-positive rate ϕ_0 now suffices to make the platform adopt a business model that relies on high ad loads.

On welfare, naïfs are unequivocally worse off: whenever they are identified, they face the higher type price p_N^* and a higher ad load α_N. For sophisticates, the key distinction is between partial and near-perfect discrimination. With partial observability (types observed with probability $\kappa \in (0, 1)$), sophisticates enjoy the lower type price p_S^* when observed, but they still face (i) pooled pricing when

[18] We believe the high-turnover (large μ) regime is empirically relevant for ad-supported digital platforms. App usage data consistently show steep early attrition: typical day-7 retention hovers around 10-15% and day-30 retention around 5-8% across broad app categories, implying that most users who install or try a service do not remain active one month later. These benchmarks appear in multiple independent data sets and summaries, and are broadly similar across news, games, lifestyle, and productivity apps (see Adjust (2024), OneSignal (2024), and Joshi (2024)). In other words, the pool of active users is continually refreshed, making a reduced-form "fast-churn" approximation reasonable for the timescales at which platforms set ad load and pricing.

type is not observed; and (ii) an ad-based business model that leads to higher downstream prices. As a result, the impact on sophisticate welfare is ambiguous. By contrast, when $\kappa \to 1$ and price discrimination is near-perfect, sophisticates always face price p_S^* and are never pooled with naïfs. Consequently, they benefit from price discrimination. This result also highlights that the negative externality from naïfs to sophisticates in our baseline model is entirely driven by the fact that the two types face the same product market price.

The comparative statics from Proposition 7—including those for price and welfare—continue to hold. Additionally, as anticipated, a higher κ expands the parameter region in which business models relying on high ad loads are more likely to be adopted.

7.4 Price Sensitivity

Our baseline model uses sophistication as the primary screening dimension—naïfs derive higher perceived interim informational value from ads and are therefore more tolerant of heavy ad loads, while sophisticates prefer plans with less advertising. In practice, price sensitivity (for example to avoid time-wasting ads) also varies across users. To explore the consequences of this alternative form of heterogeneity and clarify the key role of the distinction between naïve and sophisticated users, we now add an additional dimension of heterogeneity, representing (subscription) price sensitivity or preference for ad-free consumption.

Formally, each user i has a sophistication type $\tau_i \in \{S, N\}$ and a price-sensitivity parameter $\nu_i \in \{\nu_L, \nu_H\}$ with $0 < \nu_L < \nu_H \leq 1$. A plan ℓ specifies (α_ℓ, P_ℓ), where α_ℓ is the ad intensity and P_ℓ is the subscription price. Utility from plan ℓ for type (τ_i, ν_i) is $U_{\tau_i}(\alpha_\ell) + (1-\alpha_\ell)\nu_i T - P_\ell$, where $U_{\tau_i}(\alpha_\ell)$ is the expected product consumption utility under advertising load α_ℓ (noting that $U_N(\alpha) - U_S(\alpha)$ is strictly increasing in α from Lemma 3). We allow for correlation between τ_i and ν_i: in some markets, sophisticated users may also be more price-insensitive, whereas in others the two dimensions may be independent.

Proposition 19. *There exists a profit-maximizing menu for a platform with four (not necessarily distinct) plans $\{\alpha_\ell, P_\ell\}_{\ell=1}^4$ such that:*

(i) $(\alpha_{\tau,H}, P_{\tau,H})$ *and* $(\alpha_{\tau,L}, P_{\tau,L})$ *satisfy* $\alpha_{\tau,H} \leq \alpha_{\tau,L}$ *but* $P_{\tau,H} \geq P_{\tau,L}$ *for* $\tau \in \{S, N\}$*, and*

(ii) $\alpha_{N,\nu} \geq \alpha_{S,\nu}$ *for* $\nu \in \{\nu_L, \nu_H\}$.

The platform is therefore offering a full 2×2 menu, separating users both by sophistication $\tau_i \in \{S, N\}$ and by price-sensitivity $\nu_i \in \{\nu_L, \nu_H\}$. Within each sophistication stratum, the menu is monotone—the high willingness-to-pay users (ν_H) are offered fewer ads at a higher price in such a way that low willingness-to-pay users (ν_L) do not prefer to switch to these plans. Moreover, given price sensitivity, naïve agents still tolerate more ads than sophisticated agents, so that the equilibrium structure forms a lattice (in subscription prices and ad loads).

The two most notable results from this extension are as follows. First, we have a new testable prediction: within each sophistication tier there should be a premium (low-ad, high-price) and a basic (higher-ad, lower-price) plan, while cross-tier both naïf plans should have higher ad intensity than their sophisticated counterparts. Second, moving from two to four plans reallocates surplus toward high-ν users (who buy plans with fewer ads) and away from low-ν users, while, as before, *de facto* manipulation targets naïfs and has the same welfare implications as in our main analysis.

It is worth noting that when the only dimension of heterogeneity is price sensitivity, the environment mimics Anderson and Coate (2005), and there is no *de facto* manipulation—the platform either offers a single pooled plan or, at most, a basic/premium pair that screens according to price sensitivity.

To illustrate how heterogeneity in price sensitivity interacts with heterogeneity according to the degree of sophistication, we next present a simple example.

Example 2. First, consider a setting where (τ, ν) are drawn independently and we have $|\nu_H - \nu_L| \approx \beta \frac{\phi_1 q(1-q)}{(\phi_1 q + \phi_0(1-q))^2} |\phi_0 - \phi_{0,N}|$. Then the profit-maximizing menu has exactly two plans, (α_N, P_N) and (α_S, P_S) with $\alpha_N > \alpha_S$ and $P_N < P_S$, which separate by sophistication and pool ν within each plan. This can be seen by first observing that price alone cannot sort among ν-types, because for a fixed α, both ν-types prefer the lower price, so any "two prices at the same α" collapses to a single plan. At the same time, using α to sort ν-types tightens cross-IC: to separate agents with ν_H within a sophistication tier the platform must lower α. But lowering α also makes that plan more attractive to sophisticated types relative to naïfs, undermining the τ-separation and forcing compensating subscription price reductions on the sophisticated plans. Because the ν-mix inside the naïf and sophisticate pools mirrors the population (due to independence), inserting a ν-targeted subplan does not improve the composition of ad rents or demand within any sophistication class. Consequently, any incremental revenue from screening ν-types is second-order in $|\nu_H - \nu_L|$, outweighed by the first-order cross-IC/cannibalization costs on the order of $\beta \frac{\phi_1 q(1-q)}{(\phi_1 q + \phi_0(1-q))^2} |\phi_0 - \phi_{0,N}|$. Hence, the platform optimally uses α to separate naïfs from sophisticates and pools according to ν types.

Next, we keep preferences and technology the same, but let sophisticates be more likely to be price-insensitive (i.e., ν_H-type), with $\mathbb{P}[\nu_i = \nu_H \mid S] \gg \mathbb{P}[\nu_i = \nu_H \mid N]$. Then a three-plan structure (α_H, P_H), (α_M, P_M), and (α_L, P_L) with $\alpha_H > \alpha_M > \alpha_L$ and $P_H < P_M < P_L$ is profit-maximizing for the platform, targeting (N, ν_L), (N, ν_H), and (S, ν_H), respectively, and excluding (S, ν_L) from participation. First, we see that the correlation stacking ν_H on the S side means that a significantly higher price plan with low ads can be offered to capture the large (S, ν_H) segment while sacrificing just a few (S, ν_L) buyers, but these carry small volume due to the correlation. Similarly, a profitable middle plan now emerges and extracts additional surplus from (N, ν_H)-types that still prefer more ads than sophisticates, but are willing to pay some premium for fewer ads relative to (N, ν_L).

8 Conclusion

Digital advertising has become the dominant business model for online platforms, reaching revenue of nearly half a trillion dollars in 2022.[19] Many platforms have recently enriched their offerings by combining subscription-based and advertisement-based plans. Despite the growing importance of the ecosystem defined by digital ads and intensifying concerns that the "attention economy" created

[19] For exact figures, please see https://www.globenewswire.com/en/news-release/2022/09/28/2524217/28124/en/Global-Digital-Advertising-and-Marketing-Market-to-Reach-786-2-Billion-by-2026-at-a-CAGR-of-13-9.html.

by the desire of platforms to increase the profitability of digital ads has led to mental health problems, digital addiction, and polarization (Braghieri et al. (2022), Allcott et al. (2022), and Kubin and Von Sikorski (2021)), little is known about the economic consequences of digital advertisement and how it affects user beliefs and demand, and via these channels, the prices that users face for other goods and services.

To study these forces, we develop a parsimonious two-sided platform model, where an online platform brings together users and a firm wishing to sell a product to users. The platform offers both entertaining content and potentially informative ads about the match-specific quality of the product marketed by the firm. More ads mean less time for enjoying the content on the platform, and hence, all else equal, users would prefer fewer ads. Nevertheless, users also value the information that they get from the ads.

The main non-standard feature we introduce is that while some users are sophisticated and understand the exact data generating process for signals from ads, some users are naïve and underestimate the probability with which a low-quality product still generates a positive signal. We interpret this underestimation to be due to both the naïveté of some users (which can be affected by the salience of the ads) and to a lack of understanding that the ads are being specifically targeted and tailored for them on the basis of their personal data, which can make the advertised products appear more appealing than they are in reality.

This setup has a number of important and, to the best of our knowledge, novel implications. First, naïve agents will have a greater demand for digital ads than sophisticated users because they think that the ads are more informative. Second, naïve users will be *de facto* manipulated by digital ads, because a higher digital ad load means a greater likelihood that naïve users will overestimate the quality of the product. Third, as a result of these forces, targeting digital ads to naïve users is more profitable than targeting sophisticates. In fact, in our baseline model with linear-quadratic utility, expected purchases from sophisticated users do not change after they view informative ads, whereas expected purchases of naïve users increase because of their overestimation of the quality of the product from digital ads.

These observations are at the root of the systemic inefficiency of the decentralized equilibrium in this model. Unless digital ads are considered to be very uninformative by both types of agents, the equilibrium involves market segmentation between naïve and sophisticated users. Sophisticates are either left out of the platform (when we do not allow the platform to offer a menu of plans) or they sign up for a plan that involves a subscription fee, while naïfs are assigned to an ad-based plan without a subscription fee. This latter plan has very high ad load, targeting specifically these naïve users. When naïve agents sign up for such a plan, this increases the sales of the firm and also enables it to charge a higher price. The resulting greater profits are clawed back by the platform from the firm.

We evaluate user welfare by looking at their *ex post* utility, which depends on the actual quality of the product. While, at the interim stage where they see the ads, naïve users have an inflated assessment of the quality of the product, their *ex post* utility is a function of the actual quality of the product they consume. This implies that digital ads have a first-order welfare cost for naïve users. Notably, the misspecified model of the naïve users only applies to how they interpret signals from digital ads, so with a subscription-based model, these welfare costs are not present.

Even though sophisticates are not misled by digital ads and do not sign up for the high ad load plans, ads targeted at naïve users have welfare costs for sophisticates as well. This is because, when the firm knows that it will be able to target naïfs with its ads, it prefers to charge a higher price, and this price is also paid by sophisticated users. Consequently, digital ads make both naïve and sophisticated users worse off.

We show that all of our results generalize to environments in which there are multiple platforms and multiple firms. Beyond this generalization, our analysis reveals that digital ads have an important role in softening the competition between firms and platforms. Without digital ads, firms would compete a la Bertrand, driving their prices to marginal cost. Digital ads lead to endogenous differentiation of products—users that see positive signals about the quality of these products have different valuations than those who do not, and naïve users will have a particularly distorted evaluation. This endogenous differentiation breaks Bertrand competition and leads to equilibrium markups. In fact, competition increases the desire to target naïve users, because it provides a way of escaping the competitive pressure from other firms. Consequently, the demand for digital ads could be even higher under competition than monopoly. Similarly, digital ads also relax competition between platforms. Without naïve agents, the platforms would compete by making their offerings cheaper and more attractive to users. However, because digital ads appear more informative to naïve users, platforms generally have no incentive to reduce their digital ads, and the same type of market segmentation we saw with the monopoly platform occurs even when there are multiple competing platforms.

We also explored various policy options to counteract these systemic inefficiencies. The first-best allocation, where a planner can directly control the amount of digital ads served to naïve and sophisticated users, cannot generally be implemented because policy authorities do not observe who is naïve and who is sophisticated, and naïve agents have a greater willingness to consume ads than sophisticates, because they consider such ads to be more informative than they are in reality. Nevertheless, we show that a second-best allocation, where the social planner chooses entertainment and advertisement menus and resulting product demands subject to incentive compatibility, is easy to characterize and can be implemented using nonlinear taxation and subsidy schemes. Notably, the second best is pooling and offers a single level of ad load to both types of users. Even more simply, a flat digital ad tax (on digital ad revenue) can always improve welfare starting from an equilibrium in which there is an ad-based plan. Both of these results leverage the fact that the equilibrium often features excessive ad load, inflating the valuation of naïve users and inducing higher product market prices. By taxing revenues from digital ads, the planner makes it more attractive for platforms to monetize through subscription fees and thus reduce the excessive digital ad load.

We view our paper as a first step in the exploration of the positive and normative implications of new business models and information interactions that have become important over the last two decades. In this context, there are several interesting topics we have not touched on and many promising avenues for future research. Here we briefly list a few of these research directions.

1. In a first attempt to explore these issues, we abstracted from other social consequences of digital ads, including those related to mental health problems and digital addiction (Lukianoff and Haidt (2019) and Wu (2017)). An interesting direction for future research would be to model and

incorporate some of these issues and see how competition between firms and platforms and informational exchange between platforms and users influences these social consequences.

2. Relatedly, we took the content offered by platforms as given. How the platform monetizes itself may have first-order implications for the kind of content that it offers. In fact, some of the major concerns mentioned above are rooted in the fact that monetizing data via digital ads becomes more profitable when people spend more time on the platform, which can encourage the platform to offer content that is more addictive or emotionally triggering in order to increase user engagement (this is connected to the argument for digital ad taxes in Acemoglu and Johnson (2023)). One attempt to study these questions is Acemoglu et al. (2024), where platform algorithms modify the degree of homophily by political beliefs in order to affect engagement, which in turn has first-order implications for the spread of misinformation. These issues may become even more important when one considers a broader menu of content that can be offered, such as low-quality clickbait (Immorlica et al. (2024)), envy-generating content from friends (Beknazar-Yuzbashev et al. (2022)), or politically-provocative content (Mostagir and Siderius (2023)).

3. There may also be additional strategies platforms can utilize for extracting surplus from naïve agents. One such possibility is explored in Acemoglu et al. (2025), where the platform can engage in behavioral manipulation by steering users towards products where they are more likely to overestimate quality. A more general treatment of these issues in the context of two-sided platforms would be an interesting area for future research.

4. We simplified the analysis by ignoring how digital ads are constructed and targeted. A more in-depth analysis of this question requires us to study how user data is leveraged to tailor and target ads, and this opens the door to a broader discussion of how data access should be regulated, who owns the data generated in the process of social media interactions, and whether individuals can and should control their own data—especially taking into account both data externalities and other aspects of agents' naïvety or lack of information (Acemoglu et al. (2022) and Mostagir and Siderius (2022)).

5. Another major simplification was achieved by abstracting from social networks. Individuals often like to join platforms where their friends and acquaintances are active. Introducing this element in the competition between platforms and the business model choices of platforms would be another interesting direction for future research (see Bursztyn et al. (2023)).

6. Last but not least, our exploration raises a number of new empirical questions about how different platform offerings influence product market competition and prices. An important direction for future research is to explore the validity of both the key assumptions we have imposed (such as how naïve individuals process information from ads) and the new implications.

References

Acemoglu, Daron and Simon Johnson (2023), *Power and progress: Our thousand-year struggle over technology and prosperity*. Hachette UK.

Acemoglu, Daron and Simon Johnson (2024), "The urgent need to tax digital advertising." URL https://shapingwork.mit.edu/research/the-urgent-need-to-tax-digital-advertising/.

Acemoglu, Daron, Ali Makhdoumi, Azarakhsh Malekian, and Asu Ozdaglar (2022), "Too much data: Prices and inefficiencies in data markets." *American Economic Journal: Microeconomics*, 14, 218–256.

Acemoglu, Daron, Ali Makhdoumi, Azarakhsh Malekian, and Asuman Ozdaglar (2025), "When big data enables behavioral manipulation." *American Economic Review: Insights*, 7, 19–38.

Acemoglu, Daron, Asuman Ozdaglar, and James Siderius (2024), "A model of online misinformation." *Review of Economic Studies*, 91, 3117–3150.

Adjust (2024), "User retention: How to successfully keep users and grow your app." URL https://www.adjust.com/resources/guides/user-retention/. Guide with global benchmarks (e.g., D1 26%, D7 13%, D30 7%).

Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow (2020), "The welfare effects of social media." *American Economic Review*, 110, 629–676.

Allcott, Hunt, Matthew Gentzkow, and Lena Song (2022), "Digital addiction." *American Economic Review*, 112, 2424–2463.

Anderson, Simon P and Stephen Coate (2005), "Market provision of broadcasting: A welfare analysis." *The Review of Economic Studies*, 72, 947–972.

Anderson, Simon P and Joshua S Gans (2011), "Platform siphoning: Ad-avoidance and media content." *American Economic Journal: Microeconomics*, 3, 1–34.

Anderson, Simon P and Régis Renault (2006), "Advertising content." *American Economic Review*, 96, 93–113.

Athey, Susan and Joshua S Gans (2010), "The impact of targeting technology on advertising markets and media competition." *American Economic Review*, 100, 608–613.

Beknazar-Yuzbashev, George, Rafael Jiménez Durán, Jesse McCrosky, and Mateusz Stalinski (2022), "Toxic content and user engagement on social media: Evidence from a field experiment." *Available at SSRN 4307346*.

Bennett, Colin and Jesse Gordon (2020), "Understanding the "micro" in micro-targeting: An analysis of Facebook digital advertising in the 2019 federal Canadian election." *Available at SSRN 3589687*.

Bergemann, Dirk and Stephen Morris (2019), "Information design: A unified perspective." *Journal of Economic Literature*, 57, 44–95.

Berger, Benedikt, Christian Matt, Dennis M. Steininger, and Thomas Hess (2015), "It is not just about competition with "free": Differences between content formats in consumer preferences and willingness to pay." *Journal of Management Information Systems*, 32, 105–128.

Blackwell, David (1953), "Equivalent comparisons of experiments." *The annals of mathematical statistics*, 265–272.

Braghieri, Luca, Ro'ee Levy, and Alexey Makarin (2022), "Social media and mental health." *American Economic Review*, 112, 3660–3693.

Brynjolfsson, Erik, Avinash Collis, Asad Liaqat, Daley Kutzman, Haritz Garro, Daniel Deisenroth, and Nils Wernerfelt (2024), "The consumer welfare effects of online ads: Evidence from a 9-year experiment." Technical report, National Bureau of Economic Research.

Budzinski, Oliver, Sophia Gaenssle, and Nadine Lindstädt-Dreusicke (2021), "The battle of YouTube, TV and Netflix: an empirical analysis of competition in audiovisual media markets." *SN Business & Economics*, 1, 116.

Bursztyn, Leonardo, Benjamin R Handel, Rafael Jimenez, and Christopher Roth (2023), "When product markets become collective traps: The case of social media." Technical report, National Bureau of Economic Research.

Butters, Gerard R. (1977), "Equilibrium distributions of sales and advertising prices." *Review of Economic Studies*, 44, 465–491.

Calvano, Emilio, Giacomo Calzolari, Vincenzo Denicolo, and Sergio Pastorello (2020), "Artificial intelligence, algorithmic pricing, and collusion." *American Economic Review*, 110, 3267–3297.

Chyi, Hsiang Iris (2005), "Willingness to pay for online news: An empirical study on the viability of the subscription model." *Journal of Media Economics*, 18, 131–142.

Chyi, Hsiang Iris and Yee Man Margaret Ng (2020), "Still unwilling to pay: An empirical analysis of 50 U.S. newspapers' digital subscription results." *Digital Journalism*, 8, 526–547.

Cinelli, Matteo, Walter Quattrociocchi, Alessandro Galeazzi, Carlo Michele Valensise, Emanuele Brugnoli, Ana Lucia Schmidt, Paola Zola, Fabiana Zollo, and Antonio Scala (2020), "The COVID-19 social media infodemic." *Scientific Reports*, 10, 16598.

Danciu, Victor et al. (2014), "Manipulative marketing: persuasion and manipulation of the consumer through advertising." *Theoretical and Applied Economics*, 21, 591.

De Jans, Steffi, Dieneke Van de Sompel, Liselot Hudders, and Veroline Cauberghe (2019), "Advertising targeting young children: an overview of 10 years of research (2006–2016)." *International Journal of Advertising*, 38, 173–206.

De Reuver, Mark, Carsten Sørensen, and Rahul C Basole (2018), "The digital platform: a research agenda." *Journal of information technology*, 33, 124–135.

DellaVigna, Stefano and Matthew Gentzkow (2010), "Persuasion: empirical evidence." *Annu. Rev. Econ.*, 2, 643–669.

Deng, Yiting and Carl F. Mela (2018), "TV viewing and advertising targeting." *Journal of Marketing Research*, 55, 99–118.

Dixit, Avinash and Victor Norman (1978), "Advertising and welfare." *The Bell Journal of Economics*, 1–17.

Esponda, Ignacio and Demian Pouzo (2016), "Berk–Nash equilibrium: A framework for modeling agents with misspecified models." *Econometrica*, 84, 1093–1130.

Eyal, Nir (2014), *Hooked: How to build habit-forming products*. Penguin.

Flew, Terry (2021), "Willingness to pay: news media." *International Institute of Communication*.

Galán, Aritz Arrate, José González Cabañas, Ángel Cuevas, María Calderón, and Rubén Cuevas Rumin (2019), "Large-scale analysis of user exposure to online advertising on facebook." *IEEE access*, 7, 11959–11971.

Galbraith, John Kenneth (1958), *The affluent society*. Houghton Mifflin Harcourt.

Golab-Andrzejak, Edyta (2023), "The impact of generative AI and chatgpt on creating digital advertising campaigns." *Cybernetics and Systems*, 1–15.

Goldfarb, Avi and Catherine Tucker (2011), "Search engine advertising: Channel substitution when pricing ads to context." *Management Science*, 57, 458–470.

Gordon, Brett R, Florian Zettelmeyer, Neha Bhargava, and Dan Chapsky (2019), "A comparison of approaches to advertising measurement: Evidence from big field experiments at facebook." *Marketing Science*, 38, 193–225.

Grossman, Gene M. and Carl Shapiro (1984), "Informative advertising with differentiated products." *The Review of Economic Studies*, 51, 63–81.

Gupta, Aastha (2023), "Deceptive advertising, regulation and naive consumers." *International Journal of Industrial Organization*, 91, 103026.

Hattori, Keisuke and Keisaku Higashida (2012), "Misleading advertising in duopoly." *Canadian Journal of Economics/Revue canadienne d'économique*, 45, 1154–1187.

Immorlica, Nicole, Meena Jagadeesan, and Brendan Lucier (2024), "Clickbait vs. quality: How engagement-based optimization shapes the content landscape in online platforms." In *Proceedings of the ACM on Web Conference 2024*, 36–45.

Johnson, Garrett, Randall A Lewis, and Elmar Nubbemeyer (2017), "The online display ad effectiveness funnel & carryover: Lessons from 432 field experiments." *Available at SSRN 2701578*.

Joshi, Shweta (2024), "App retention benchmarks broken down by industry." URL https://sendbird.com/blog/app-retention-benchmarks-broken-down-by-industry.

Kruikemeier, Sanne, Minem Sezgin, and Sophie C Boerman (2016), "Political microtargeting: relationship between personalized advertising on facebook and voters' responses." *Cyberpsychology, Behavior, and Social Networking*, 19, 367–372.

Kubin, Emily and Christian Von Sikorski (2021), "The role of (social) media in political polarization: a systematic review." *Annals of the International Communication Association*, 45, 188–206.

Lewis, Randall A and Justin M Rao (2015), "The unfavorable economics of measuring the returns to advertising." *The Quarterly Journal of Economics*, 130, 1941–1973.

Lukianoff, Greg and Jonathan Haidt (2019), *The coddling of the American mind: How good intentions and bad ideas are setting up a generation for failure*. Penguin.

Marwick, Alice E. and Rebecca Lewis (2017), "Media manipulation and disinformation online." *Data & Society Research Institute*.

Meurer, Michael and Dale O. Stahl (1994), "Informative advertising and product match." *International Journal of Industrial Organization*, 12, 1–19.

Moldovanu, Benny and Eyal Winter (1995), "Order independent equilibria." *Games and Economic Behavior*, 9, 21–34.

Mostagir, Mohamed and James Siderius (2022), "Naive and bayesian learning with misinformation policies." Technical report, Massachusetts Institute of Technology.

Mostagir, Mohamed and James Siderius (2023), "When should platforms break echo chambers?" Technical report, Massachusetts Institute of Technology.

Nelson, Phillip (1974), "Advertising as information." *Journal of political economy*, 82, 729–754.

OneSignal (2024), "Must-know mobile app benchmarks of 2024." URL https://onesignal.com/mobile-app-benchmarks-2024.

Ottaviani, Marco and Andrea Prat (2001), "The value of public information in monopoly." *Econometrica*, 69, 1673–1683.

Piccolo, Salvatore, Piero Tedeschi, and Giovanni Ursino (2018), "Deceptive advertising with rational buyers." *Management Science*, 64, 1291–1310.

Prat, Andrea and Tommaso Valletti (2022), "Attention oligopoly." *American Economic Journal: Microeconomics*, 14, 530–557.

Rayo, Luis (2013), "Monopolistic signal provision." *The BE Journal of Theoretical Economics*, 13, 27–58.

Rayo, Luis and Ilya Segal (2010), "Optimal information disclosure." *Journal of political Economy*, 118, 949–987.

Richter, Felix (2019), "The generation gap in TV consumption." *Statista*.

Rochet, Jean-Charles and Jean Tirole (2006), "Two-sided markets: a progress report." *The RAND journal of economics*, 37, 645–667.

Roesler, Anne-Katrin and Balázs Szentes (2017), "Buyer-optimal learning and monopoly pricing." *American Economic Review*, 107, 2072–2080.

Romer, Paul (2021), "Taxing digital advertising." URL https://adtax.paulromer.net/.

Sato, Susumu (2019), "Freemium as optimal menu pricing." *International Journal of Industrial Organization*, 63, 480–510.

Sherman, Ryland and David Waterman (2016), "The economics of online video entertainment." *Handbook on the Economics of the Internet*, 458–474.

Simchon, Almog, Matthew Edwards, and Stephan Lewandowsky (2024), "The persuasive effects of political microtargeting in the age of generative artificial intelligence." *PNAS nexus*, 3, pgae035.

Sinkinson, Michael and Amanda Starc (2019), "Ask your doctor? direct-to-consumer advertising of pharmaceuticals." *The Review of Economic Studies*, 86, 836–881.

Skreta, Vasiliki (2011), "On the informed seller problem: optimal information disclosure." *Review of Economic Design*, 15, 1–36.

Susser, Daniel and Vincent Grimaldi (2021), "Measuring automated influence: between empirical evidence and ethical values." In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 242–253.

Tirole, Jean (1988), *The theory of industrial organization*, 1st edition edition. The MIT Press, Cambridge, Mass.

Tran, Trang P (2017), "Personalized ads on facebook: An effective marketing tool for online marketers." *Journal of retailing and consumer services*, 39, 230–242.

Twenge, Jean M., Gabrielle N. Martin, and Brian H. Spitzberg (2019), "Trends in U.S. adolescents' media use, 1976–2016: The rise of digital media, the decline of TV, and the (near) demise of print." *Psychology of Popular Media Culture*, 8, 329–345.

Vock, Marlene, Willemijn van Dolen, and Ko de Ruyter (2013), "Understanding willingness to pay for social network sites." *Journal of Service Research*, 16, 311–325.

Wu, Tim (2017), *The attention merchants: The epic scramble to get inside our heads*. Vintage.

Zennyo, Yusuke (2020), “Freemium competition among ad-sponsored platforms.” *Information Economics and Policy*, 50, 100848.

A Online Appendix: Omitted Proofs

Lemma A.1. *For type τ_i, the (interim) informational value is given by $I_{\tau_i}(\alpha) = q\mathbb{E}^{\tau_i}_{\pi_i|\theta_i=1}[\pi_i(\beta^2 - \beta^2\pi_i/2) \mid \alpha] - (1-q)\mathbb{E}^{\tau_i}_{\pi_i|\theta_i=0}[(\beta\pi_i)^2/2 \mid \alpha] - (\beta q)^2/2$. In particular, $I_{\tau_i}(\alpha)$ does not depend on price p.*

Proof. Note that by definition,

$$I_{\tau_i}(\alpha) = q\mathbb{E}^{\tau_i}_{\pi_i|\theta_i=1}[U(z_i^*(\pi_i, p^*);\, \theta_i = 1) - p^* z_i^*(\pi_i, p^*) \mid \alpha] \\ + (1-q)\mathbb{E}^{\tau_i}_{\pi_i|\theta_i=0}[U(z_i^*(\pi_i, p^*);\, \theta_i = 0) - p^* z_i^*(\pi_i, p^*) \mid \alpha] - (U(z_i^*(q, p^*);\, \theta_i = q) - p^* z_i^*(q, p^*)) .$$

Moreover, we have $z_i^*(\pi, p^*) = \arg\max_{z_i} \pi_i \beta z_i - z_i^2/2 - p z_i = \pi_i\beta - p^*$. Thus:

$$\begin{aligned} I_{\tau_i}(\alpha) =& q\mathbb{E}^{\tau_i}_{\pi_i|\theta_i=1}[(\pi_i\beta - p^*)\beta - (\pi_i\beta - p^*)^2/2 - p^*(\pi_i\beta - p^*)] + (1-q)\mathbb{E}^{\tau_i}_{\pi_i|\theta_i=0}[-(\pi_i\beta - p^*)^2/2 - p^*(\pi_i\beta - p^*)] \\ & - q\beta(q\beta - p^*) + (q\beta - p^*)^2/2 + p^*(q\beta - p^*) \\ =& q\mathbb{E}^{\tau_i}_{\pi_i|\theta_i=1}[\pi_i(\beta^2 - \beta^2\pi_i/2)] - q\beta p^* + q(p^*)^2/2 - (1-q)\mathbb{E}^{\tau_i}_{\pi_i|\theta_i=0}[(\beta\pi_i)^2/2] + (1-q)(p^*)^2/2 \\ & + q\beta p^* - (\beta q)^2/2 - (p^*)^2/2 \\ =& \, q\mathbb{E}^{\tau_i}_{\pi_i|\theta_i=1}[\pi_i(\beta^2 - \beta^2\pi_i/2)] - (1-q)\mathbb{E}^{\tau_i}_{\pi_i|\theta_i=0}[(\beta\pi_i)^2/2] - (\beta q)^2/2, \end{aligned}$$

where for notational simplicity, we suppressed the dependence on α. ∎

A.1 Proofs from Section 2

Proof of Proposition 1. The first-best user welfare for sophisticates occurs where $I_S(\alpha) + (1-\alpha)T$ is maximized, which is attained at some value we denote by α_S^{FB}. Let us denote by $\Delta(\alpha)$ the surplus loss associated with consumption under naïve beliefs:

$$\begin{aligned} \Delta(\alpha) =& \Bigg[q \underbrace{\mathbb{E}_{\pi\sim F_{S1}(\alpha)}[U(z^*(\pi, c);\, \theta_i = 1) - cz^*(\pi, c)]}_{\text{Sophisticate Consumption Utility Conditional on } \theta_i = 1} + (1-q) \underbrace{\mathbb{E}_{\pi\sim F_{S0}(\alpha)}[U(z^*(\pi, c);\, \theta_i = 0) - cz^*(\pi, c)]}_{\text{Sophisticate Consumption Utility Conditional on } \theta_i = 0} \Bigg] \\ & - \Bigg[q \underbrace{\mathbb{E}_{\pi\sim F_{N1}(\alpha)}[U(z^*(\pi, c);\, \theta_i = 1) - cz^*(\pi, p^*)]}_{\text{naïve Consumption Utility Conditional on } \theta_i = 1} + (1-q) \underbrace{\mathbb{E}_{\pi\sim F_{N0}(\alpha)}[U(z^*(\pi, c);\, \theta_i = 0) - cz^*(\pi, c)]}_{\text{naïve Consumption Utility Conditional on } \theta_i = 0} \Bigg] \\ =& e^{-\alpha T} \cdot 0 + (1 - e^{-\alpha T}) \Big[q\phi_1 \Big(U(z^*(\pi^S(G), c);\, \theta_i = 1) - cz^*(\pi^S(G), p^*) \\ & - U(z^*(\pi^N(G), c);\, \theta_i = 1) + cz^*(\pi^S(G), p^*) \Big) \\ & + q(1-\phi_1) \left(U(z^*(\pi^S(B), c);\, \theta_i = 1) - cz^*(\pi^S(B), p^*) - U(z^*(\pi^N(B), c);\, \theta_i = 1) + cz^*(\pi^N(B), p^*) \right) \\ & + (1-q)\phi_0 \left(U(z^*(\pi^S(G), c);\, \theta_i = 0) - cz^*(\pi^S(G), p^*) - U(z^*(\pi^N(G), c);\, \theta_i = 0) + cz^*(\pi^N(G), c) \right) \\ & + (1-q)(1-\phi_0) \left(U(z^*(\pi^S(B), c);\, \theta_i = 0) - cz^*(\pi^S(G), p^*) - U(z^*(\pi^N(B), c);\, \theta_i = 0) + cz^*(\pi^N(B), p^*) \right) \Big] \end{aligned}$$

It remains to show that $\Delta(\alpha)$ is increasing in α, which then ensures that the maximizer of $I_S(\alpha) + (1 - \alpha)T - \Delta(\alpha)$ is at some $\alpha_N^{FB} \leq \alpha_S^{FB}$. The difference in the previous expression is a linear combination with weights $\exp(-\alpha T)$ and $1-\exp(-\alpha T)$ of 0 and a strictly positive value given by the welfare loss when naïve agents see an ad relative to their sophisticated counterparts. Because this welfare loss is increasing in α, the linear combination is also increasing in α.

Finally, note that without the platform the agent gets v and faces product price $p^*_{\text{base}} \geq c$. By assumption that $T > v$, we know that there exists some value of α such that $I_S(\alpha) + (1-\alpha)T > v$, and moreover that $I_S(\alpha_S^{FB}) + (1-\alpha_S^{FB})T > v$, so $W_{FB}(S) > W_{\text{base}}(S)$. For naïve agents, we need only consider $I_S(\alpha) + (1-\alpha)T - \Delta(\alpha, c)$ as before. Because naïve agents make the same consumption decision as sophisticates when $\alpha = 0$, we know that $I_S(0) + T - \Delta(0, c) = T > v$. Because there exists a value of α where $I_S(\alpha)+(1-\alpha)T-\Delta(\alpha, c) > v$, we can conclude that $I_S(\alpha_N^{FB})+(1-\alpha_N^{FB})T-\Delta(\alpha_N^{FB}, c) > v$, and thus $W_{FB}(N) > W_{\text{base}}(N)$. ∎

A.2 Proofs from Section 3

Proof of Lemma 1. This follows immediately from the Martingale property of Bayesian beliefs. Let us denote by $\bar{\pi}_F$ the expected belief under distribution F, i.e., $\bar{\pi}_F = \mathbb{E}_{\pi\sim F}[\pi]$. Then,

$$\begin{aligned}
&q\bar{\pi}_{F_{S1}(\alpha)} + (1-q)\bar{\pi}_{F_{S0}(\alpha)} \\
&= q\left(e^{-\alpha T}q + \phi_1(1-e^{-\alpha T})\frac{\phi_1 q}{\phi_1 q + \phi_0(1-q)} + (1-\phi_1)(1-e^{-\alpha T})\frac{(1-\phi_1)q}{(1-\phi_1)q + (1-\phi_0)(1-q)}\right) \\
&+ (1-q)\left(e^{-\alpha T}q + \phi_0(1-e^{-\alpha T})\frac{\phi_1 q}{\phi_1 q + \phi_0(1-q)} + (1-\phi_0)(1-e^{-\alpha T})\frac{(1-\phi_1)q}{(1-\phi_1)q + (1-\phi_0)(1-q)}\right) \\
&= q.
\end{aligned}$$

Thus, $\int_{\tau_i=S}(p^*-c)z_i^*\,di = \int_{\tau_i=S}(p^*-c)(\beta q - p^*)\,di$, which is independent of α. ∎

Proof of Lemma 2. By the fact that $\pi_i^S|\theta_i \preceq_{FOSD} \pi_i^N|\theta_i$, we know that $q\bar{\pi}_{F_{N1}(\alpha)} + (1-q)\bar{\pi}_{F_{N0}(\alpha)} > q$ for $\alpha > 0$. To show that Π_N^* is increasing in α, we note that

$$\begin{aligned}
&q\bar{\pi}_{F_{N1}(\alpha)} + (1-q)\bar{\pi}_{F_{N0}(\alpha)} \\
&= q\left(e^{-\alpha T}q + \phi_1(1-e^{-\alpha T})\frac{\phi_1 q}{\phi_1 q + \phi_{0,N}(1-q)} + (1-\phi_1)(1-e^{-\alpha T})\frac{(1-\phi_1)q}{(1-\phi_1)q + (1-\phi_{0,N})(1-q)}\right) \\
&+ (1-q)\left(e^{-\alpha T}q + \phi_0(1-e^{-\alpha T})\frac{\phi_1 q}{\phi_1 q + \phi_{0,N}(1-q)} + (1-\phi_0)(1-e^{-\alpha T})\frac{(1-\phi_1)q}{(1-\phi_1)q + (1-\phi_{0,N})(1-q)}\right)
\end{aligned}$$

Because $\phi_{0,N} < \phi_0$, we have:

$$\begin{aligned}
&q\phi_1\frac{\phi_1 q}{\phi_1 q + \phi_{0,N}(1-q)} + q(1-\phi_1)\frac{(1-\phi_1)q}{(1-\phi_1)q + (1-\phi_{0,N})(1-q)} \\
&+ (1-q)\phi_0\frac{\phi_1 q}{\phi_1 q + \phi_{0,N}(1-q)} + (1-q)(1-\phi_0)\frac{(1-\phi_1)q}{(1-\phi_1)q + (1-\phi_{0,N})(1-q)}
\end{aligned}$$

$$>q\phi_1\frac{\phi_1 q}{\phi_1 q+\phi_0(1-q)}+q(1-\phi_1)\frac{(1-\phi_1)q}{(1-\phi_1)q+(1-\phi_0)(1-q)}$$
$$+(1-q)\phi_0\frac{\phi_1 q}{\phi_1 q+\phi_0(1-q)}+(1-q)(1-\phi_0)\frac{(1-\phi_1)q}{(1-\phi_1)q+(1-\phi_0)(1-q)}=q$$

which implies that $q\bar{\pi}_{F_{S1}(\alpha)}+(1-q)\bar{\pi}_{F_{S0}(\alpha)}=e^{-\alpha T}q+(1-e^{-\alpha T})\chi$ for some $\chi>q$. Thus, we see this expression is increasing in α and that $\int_{\tau_i=N}(p^*-c)z_i^*\,di=\int_{\tau_i=N}(p^*-c)(e^{-\alpha T}q+(1-e^{-\alpha T})\chi-p^*)$, which is increasing in α for the profit-maximizing $p^*=\arg\max_p(p-c)(\lambda\beta q+(1-\lambda)\beta(e^{-\alpha T}q+(1-e^{-\alpha T})\chi)-p)$, provided that $I_N(\alpha)+(1-\alpha)T\geq v$ (the naïve agent's participation constraint is met). ∎

Proof of Lemma 3. Since both agents are Bayesian under their subjective models, we know that, by Blackwell's theorem, we can rank $I_N(\alpha)$ and $I_S(\alpha)$ if we can show the naïve agent's subjective model is more informative than the sophisticate's subjective model in the Blackwell order (Blackwell (1953)). Consider a signal s_i generated according to the naïve agent's subjective model and consider an alternative signal generation process that is strictly less informative, constructed as follows. If $s_i=G$ given $\theta_i=0$, retain the signal $s_i'=G$. If $s_i=B$ and $\theta_i=0$, then with probability $\frac{1-\phi_0}{1-\phi_{0,N}}$, retain the signal as $s_i'=B$, otherwise switch the signal to $s_i'=G$. Clearly this construction is a garbling process that makes the sophisticates' signal structure Blackwell dominated by the naïve agents' signal structure. Consequently, $I_N(\alpha)>I_S(\alpha)$ for all $\alpha>0$.

Using Lemma A.1, we note that expanding $I_{\tau_i}(\alpha)=q\mathbb{E}^{\tau_i}_{\pi_i|\theta_i=1}[\pi_i(\beta^2-\beta^2\pi_i/2)\mid\alpha]-(1-q)\mathbb{E}^{\tau_i}_{\pi_i|\theta_i=0}[(\beta\pi_i)^2/2\mid\alpha]-(\beta q)^2/2$ implies that

$$I_N(\alpha)=(1-e^{-\alpha T})\chi_N,$$
$$I_S(\alpha)=(1-e^{-\alpha T})\chi_S,$$

for some constants $\chi_N>\chi_S$ that depend on model primitives (e.g., ϕ_0, ϕ_1) but not on α. Concavity and monotone increasing in α follow immediately. To observe $\arg\max I_N(\alpha)+(1-\alpha)T>\arg\max I_S(\alpha)+(1-\alpha)T$, we note that $I_N'(\alpha)=T$ precisely when $e^{-\alpha T}=\frac{1}{\chi_N}$, which has an intersection point that occurs later than $e^{-\alpha T}=\frac{1}{\chi_S}$, because $1/\chi_N<1/\chi_S$, and $e^{-\alpha T}$ is a decreasing function in α. ∎

Proof of Lemma 4. For $t=1$ and $t=2$, we can combine the decision problems of the platform and firm into one joint decision problem, because the platform makes a take-it-or-leave-it offer to the firm and thus effectively maximizes their joint surplus. In particular, the platform will solve a set of maximization problems, taking into account the participation constraints (PCs) of different agent types, and then compare the maximized values to select the one that gives the highest profits. These maximization problems are given as follows:

- The optimal product price p^* when no agent participates on the platform, and learns nothing about her preferences (holds prior q about $\theta_i=1$).

- The optimal product price p^* and advertising load α^* when only the sophisticated agents participate on the platform. Sophisticates learn about their preferences from ads, whereas naïves operate under their prior q.

- The optimal product price p^* and advertising load α^* when only the naïve agents participate on the platform. Naïfs learn about their preferences from ads, whereas sophisticates operate under their prior q.
- The optimal product price p^* and advertising load α^* when both sophisticates and naïfs participate on the platform. All agents learn about their preferences from ads.

We can write the maximized values in these problems as follows:

$$\mathcal{A}_1 \equiv \max_{\alpha,p} \underbrace{(p-c)}_{\text{Marginal Profit}} \cdot \underbrace{(\beta q - p)}_{\text{Consumer Demand without Ads}}, \qquad \text{(No user participation)}$$

$$\mathcal{A}_2 \equiv \max_{\alpha,p} \lambda(p-c)\underbrace{\left(\beta q \bar{\pi}_{F_{S1}(\alpha)} + \beta(1-q)\bar{\pi}_{F_{S0}(\alpha)} - p\right)}_{\text{Sophisticate Demand with Ads}} + (1-\lambda)(p-c)(\beta q - p) \qquad \text{(Sophisticates participate)}$$

$$\text{subject to} \quad \underbrace{I_S(\alpha) + (1-\alpha)T - v}_{\text{Sophisticate Surplus from Platform Participation}} \geq 0,$$

$$\mathcal{A}_3 \equiv \max_{\alpha,p} (1-\lambda)(p-c)\underbrace{\left(\beta q \bar{\pi}_{F_{N1}(\alpha)} + \beta(1-q)\bar{\pi}_{F_{N0}(\alpha)} - p\right)}_{\text{naïve Demand with Ads}} + \lambda(p-c)(\beta q - p) \qquad \text{(naïves participate)}$$

$$\text{subject to} \quad \underbrace{I_N(\alpha) + (1-\alpha)T - v}_{\text{naïve Surplus from Platform Participation}} \geq 0,$$

$$\begin{aligned}
\mathcal{A}_4 \equiv \max_{\alpha,p} \; & (p-c)(\lambda\beta q \bar{\pi}_{F_{S1}(\alpha)} + \lambda\beta(1-q)\bar{\pi}_{F_{S0}(\alpha)} \\
& + (1-\lambda)\beta q \bar{\pi}_{F_{N1}(\alpha)} + (1-\lambda)\beta(1-q)\bar{\pi}_{F_{N0}(\alpha)} - p) \qquad \text{(All users participate)} \\
\text{subject to} \quad & I_S(\alpha) + (1-\alpha)T - v \geq 0 \\
& I_N(\alpha) + (1-\alpha)T - v \geq 0.
\end{aligned}$$

We can further simplify the platform's problem by noting that it is without loss to restrict attention to just $\mathcal{A}_3$. First, one can observe that $q\bar{\pi}_{F_{S1}(\alpha)} + (1-q)\bar{\pi}_{F_{S0}(\alpha)} = q$ because sophisticated agents have a properly specified Bayesian model (by Lemma 1), and thus, $\mathcal{A}_1 \geq \mathcal{A}_2$. At the same time, $\bar{\pi}_{F_{N1}(\alpha)} > \bar{\pi}_{F_{S1}(\alpha)}$ and $\bar{\pi}_{F_{N0}(\alpha)} > \bar{\pi}_{F_{S0}(\alpha)}$ because $\pi^N \,|\, \theta_i \succeq_{FOSD} \pi^S \,|\, \theta_i$, so $q\bar{\pi}_{F_{N1}(\alpha)} + (1-q)\bar{\pi}_{F_{N0}(\alpha)} > q$ (via Lemma 2). We also note that $I_N(\alpha) + (1-\alpha)T - v \geq 0$ can be feasibly satisfied at $\alpha = 0$ (by assumption that $T > v$), so it must be that $\mathcal{A}_3 \geq \mathcal{A}_1 \geq \mathcal{A}_2$. Finally, notice that because $q\bar{\pi}_{F_{S1}(\alpha)} + (1-q)\bar{\pi}_{F_{S0}(\alpha)} = q$, the objective of $\mathcal{A}_3$ and $\mathcal{A}_4$ are identical, but $\mathcal{A}_4$ is subject to an additional constraint, implying that $\mathcal{A}_3 \geq \mathcal{A}_4$. Putting these pieces together, we observe that conditional on adopting an advertising model, the platform will advertise to attract only naïves ($\mathcal{A}_3$). Moreover, because the objective of $\mathcal{A}_3$ is increasing in α (by Lemma 2), the platform will choose an load α^* that satisfies $I_N(\alpha^*) + (1-\alpha^*)T - v = 0$. Consequently, because $I_S(\alpha) < I_N(\alpha)$ by Lemma 3, sophisticates will not participate on account of $I_S(\alpha^*) + (1-\alpha^*)T - v < 0$. ∎

Proof of Proposition 2. By Lemma 4, the only two business models possible are those in Proposition 2(a) and (b). We let $m^*(\alpha, \phi_0, \lambda, \phi_1, \phi_{0,N})$ denote the advertising revenue the platform can extract from the firm as a function of the advertising load α, false positive rate ϕ_0, the fraction of sophisticates λ, the true positive rate ϕ_1, and the naïfs' false positive rate $\phi_{0,N}$. With advertising, we know that the firm will solve the pricing problem post-advertising:

$$\max_p (p-c)(\lambda\beta q+(1-\lambda)\beta(q\bar{\pi}_{F_{N1}(\alpha)}+(1-q)\bar{\pi}_{F_{N0}(\alpha)}-p)) = \left(\frac{\lambda\beta q+(1-\lambda)\beta(q\bar{\pi}_{F_{N1}(\alpha)}+(1-q)\bar{\pi}_{F_{N0}(\alpha)})-c}{2}\right)^2,$$

by charging $\hat{p}^* = \frac{1}{2}\left(\lambda\beta q + (1-\lambda)\beta(q\bar{\pi}_{F_{N1}(\hat{\alpha}^*)} + (1-q)\bar{\pi}_{F_{N0}(\hat{\alpha}^*)}) + c\right)$. Without advertising, the platform will charge a subscription fee and the firm will make a pre-advertising profit of $\left(\frac{\beta q - c}{2}\right)^2$ by charging $\bar{p}^*$. The difference in the firm's two profit expressions, pre- and post-advertising, corresponds to the maximum transfer m^* the platform can extract from the firm.

By Lemmas 1 and 2, we know that m^* is increasing in α. Thus, if the platform chooses an advertising model, it does so at the rate $\hat{\alpha}^*$ with $I_N(\hat{\alpha}^*) + (1-\hat{\alpha}^*)T - v = 0$. For comparative statics on other primitives, we can also note that

$$\begin{aligned} &q\bar{\pi}_{F_{N1}(\alpha)} + (1-q)\bar{\pi}_{F_{N0}(\alpha)} \\ &= q\left(e^{-\alpha T}q + \phi_1(1-e^{-\alpha T})\frac{\phi_1 q}{\phi_1 q + \phi_{0,N}(1-q)} + (1-\phi_1)(1-e^{-\alpha T})\frac{(1-\phi_1)q}{(1-\phi_1)q + (1-\phi_{0,N})(1-q)}\right) \\ &+ (1-q)\left(e^{-\alpha T}q + \phi_0(1-e^{-\alpha T})\frac{\phi_1 q}{\phi_1 q + \phi_{0,N}(1-q)} + (1-\phi_0)(1-e^{-\alpha T})\frac{(1-\phi_1)q}{(1-\phi_1)q + (1-\phi_{0,N})(1-q)}\right), \end{aligned}$$

is monotone in profit for ϕ_0, ϕ_1, and $\phi_{0,N}$, that it is linear in ϕ_0, and that ϕ_0 does not affect the platform's choice of $\hat{\alpha}^*$ (because ϕ_0 does not factor into the naïfs' participation constraint). Moreover, $\frac{\phi_1 q}{\phi_1 q+\phi_{0,N}(1-q)} > \frac{(1-\phi_1)q}{(1-\phi_1)q+(1-\phi_{0,N})(1-q)}$ by assumption that $\phi_1 > \phi_0 > \phi_{0,N}$. Thus, m^* is increasing in ϕ_0 and the platform trades off the advertising revenue m^* with $T - v$. This observation establishes the existence ofa cutoff strategy in ϕ_0, $\hat{\phi}_0(\lambda, \phi_1, \phi_{0,N})$, which we simply refer to as $\hat{\phi}_0$ in the rest of this proof for notational simplicity.

To determine the dependence of $\hat{\phi}_0$ on other model primitives, we perform comparative statics. For λ, because we know $q\bar{\pi}_{F_{N1}(\alpha)} + (1-q)\bar{\pi}_{F_{N0}(\alpha)} > q$, we see that m^* is decreasing in λ. Thus, the advertising-based business model becomes less attractive as λ increases, so $\hat{\phi}_0$ is increasing in λ. For ϕ_1, we observe that holding α fixed,

$$\frac{\partial(q\bar{\pi}_{F_{N1}(\alpha)} + (1-q)\bar{\pi}_{F_{N0}(\alpha)})}{\partial\phi_1} = (1-e^{-\alpha T})\frac{(1-q)^2 q(\phi_0-\phi_{0,N})(q^2\phi_{0,N}^2 - 2q^2\phi_{0,N}\phi_1 + q^2\phi_1^2 - \phi_{0,N}^2 + \phi_{0,N})}{(q\phi_{0,N} - q\phi_1 - \phi_{0,N} + 1)^2(q\phi_{0,N} - q\phi_1 - \phi_{0,N})^2},$$

which is positive because $\phi_0 > \phi_{0,N}$, $q^2\phi_1^2 + q^2\phi_{0,N}^2 > 2q^2\phi_{0,N}\phi_1$, and $\phi_{0,N} > \phi_{0,N}^2$. At the same time, higher ϕ_1 increases $I_N(\alpha)$, so a greater advertising load still satisfies the naïfs' PC, leading to once again higher m^*. Thus, the ad-based business model becomes even more likely as ϕ_1 increases, or in other words, the cutoff $\hat{\phi}_0$ decreases in ϕ_1.

Finally, we can see that $q\bar{\pi}_{F_{N1}(\alpha)} + (1-q)\bar{\pi}_{F_{N0}(\alpha)}$ is decreasing in $\phi_{0,N}$:

$$\frac{\partial(q\bar{\pi}_{F_{N1}(\alpha)} + (1-q)\bar{\pi}_{F_{N0}(\alpha)})}{\partial\phi_{0,N}} = (1-e^{-\alpha T})(1-q)q\left(-\frac{\phi_1(\phi_0(1-q)+q\phi_1)}{(\phi_{0,N}+q\phi_1-q\phi_{0,N})^2} - \frac{(1-\phi_1)(1-\phi_0(1-q)-q\phi_1)}{(1-\phi_{0,N}-q(\phi_1-\phi_{0,N}))^2}\right),$$

which is strictly negative. Increasing $\phi_{0,N}$ also reduces the advertising load tolerated by naïves, $\hat{\alpha}^*$, because it also reduces $I_N(\alpha)$ and as a result depresses m^* further. So increasing $\phi_{0,N}$ makes the subscription-based model more attractive, increasing $\hat{\phi}_0$. ∎

Proof of Proposition 3. From Lemma 1, we know that there is no equilibrium transfer from the firm due to advertising, $m^* = 0$. Thus, the subscription model is more attractive to the platform, which yields $T - v$ profit, which is positive by assumption. Because there is no advertising, the firm once again solves $\max_p (p-c)(\beta q - p)$, which occurs when the price is set to $\bar{p}^* = (\beta q + c)/2$. This is the same price of the product as in the base case with no platform. Moreover, the user surplus is determined by $I_S(\alpha) + (1-\alpha)T - (T-v)$ for $\alpha = 0$, which is equal to the agent's outside option v. Thus, welfare is given by $v + qU(z_i^*(q,\bar{p}^*);\, \theta_i = 1) + (1-q)U(z_i^*(q,\bar{p}^*);\, \theta_i = 0) - \bar{p}^* z_i^*(q,\bar{p}^*)$ for both types of users in the base case as well as in the fully-rational benchmark. This establishes that $W_{\text{fully-rational}}(\tau) = W_{\text{base}}(\tau)$, which is independent of the agent's type τ because there is no advertising in equilibrium. ∎

Proof of Proposition 4. For regime (a), we have the platform adopts the same subscription model it does in Proposition 3, leading to identical welfare as in the fully-rational benchmark, which as we showed also has the same welfare as the base case without the platform.

For regime (b), advertising load is at $\hat{\alpha}^*$ and prices are at $\hat{p}^* > \bar{p}^*$. We have that the sophisticates' welfare is given by

$$\begin{aligned}
W(S, x_i = 0) &= v + qU(z_i^*(q,\hat{p}^*);\, \theta_i = 1) + (1-q)U(z_i^*(q,\hat{p}^*);\, \theta_i = 0) - \hat{p}^* z_i^*(q,\hat{p}^*) \\
&< v + qU(z_i^*(q,\bar{p}^*);\, \theta_i = 1) + (1-q)U(z_i^*(q,\bar{p}^*);\, \theta_i = 0) - \bar{p}^* z_i^*(q,\bar{p}^*) \\
&= W_{\text{base}}(S)
\end{aligned}$$

Recall from Proposition 1 that $\Delta(\alpha)$ is the naïve agent's *ex post* consumer surplus lost relative to a sophisticated agent under the same advertising load α. The naïfs' welfare is given by

$$\begin{aligned}
W(N, x_i = 1) =& (1-\hat{\alpha}^*)T + q\mathbb{E}_{\pi\sim F_{N1}}\left[U(z^*(\pi,\hat{p}^*);\, \theta_i = 1) - \hat{p}^* z^*(\pi,\hat{p}^*)\right] \\
&+ (1-q)\mathbb{E}_{\pi\sim F_{N0}}\left[U(z^*(\pi,\hat{p}^*);\, \theta_i = 0) - \hat{p}^* z^*(\pi,\hat{p}^*)\right] \\
=& (1-\hat{\alpha}^*)T + q\mathbb{E}^S_{\pi_i|\theta_i=1}[U(z_i^*(\pi_i,\hat{p}^*);\, \theta_i = 1) - \hat{p}^* z_i^*(\pi_i,\hat{p}^*) \,|\, \hat{\alpha}^*] \\
&+ (1-q)\mathbb{E}^S_{\pi_i|\theta_i=0}[U(z_i^*(\pi_i,\hat{p}^*);\, \theta_i = 0) - \hat{p}^* z_i^*(\pi_i,\hat{p}^*) \,|\, \hat{\alpha}^*] - \Delta(\hat{\alpha}^*) \\
=& (1-\hat{\alpha}^*)T + I_S(\hat{\alpha}^*) - \Delta(\hat{\alpha}^*) + U(z_i^*(q,\hat{p}^*)) - \hat{p}^* z_i^*(q,\hat{p}^*) \\
=& v - \Delta(\hat{\alpha}^*) + (I_S(\hat{\alpha}^*) - I_N(\hat{\alpha}^*)) + \frac{1}{2}(\beta q - \hat{p}^*)^2 \\
<& v + \frac{1}{2}(\beta q - \bar{p}^*)^2 = W_{\text{fully-rational}}(N)
\end{aligned}$$

Finally, we note from Proposition 3 that $W_{\text{fully-rational}}(\tau) = W_{\text{base}}(\tau)$ for both $\tau \in \{S, N\}$. ∎

A.3 Proofs from Section 4

Proof of Proposition 5. We now allow the platform to offer both subscription fees and advertising intensities (but not in the same plan). By Lemma 1, we know that the platform can maximize revenue from sophisticated agents by offering a subscription-based plan and charging $T-v$. This subscription-based plan leaves both naïfs and sophisticates indifferent between participating and not participating in the platform. Consequently, the platform will choose an ad-based plan for naïfs if and only if $I_N(\alpha)+(1-\alpha)T-v \geq 0$. The firm-platform transfer $m^*(\alpha, \phi_0, \lambda, \phi_1, \phi_{0,N})$ conditional on advertising is identical to that of Proposition 2. Consequently, an ad-based plan will extract all the surplus from the naïfs' PC and thus the only two candidates for equilibrium business models are regimes (a) and (b) of Proposition 5.

The revenue generated from business model (a) is given by $T-v$ whereas the revenue generated from business model (b) is given by $\lambda(T-v)+m^*(\alpha, \phi_0, \lambda, \phi_1, \phi_{0,N})$. The difference between the latter and the former is thus $m^*(\alpha, \phi_0, \lambda, \phi_1, \phi_{0,N})-(1-\lambda)(T-v)$. The comparative statics on ϕ_0, ϕ_1, and $\phi_{0,N}$ then follow immediately from Proposition 2, giving the cutoff characterization $\phi_0^*(\lambda, \phi_1, \phi_{0,N})$ and showing that the cutoff is increasing in $\phi_{0,N}$ but decreasing in ϕ_1. To see that it is increasing in λ, we note that $m^*(\alpha, \phi_0, \lambda, \phi_1, \phi_{0,N})-(1-\lambda)(T-v)$ is equal to 0 when $\lambda=1$, and either $-\partial m/\partial \lambda^*\Big|_{\lambda=1} > T-v$ or $-\partial m/\partial \lambda^*\Big|_{\lambda=1} < T-v$. In the former case, we know that $-\partial m/\partial \lambda^* > T-v$ for all λ, which implies that the ad-based plan offered to naïfs generates more revenue than the subscription-based plan for all $\lambda \in (0,1)$. On the other hand, if $-\partial m/\partial \lambda^*\Big|_{\lambda=1} < T-v$, then because m^* is quadratic in λ and $\lambda(T-v)$ is linear with the same intersection at $\lambda=1$, there exists a unique single crossing at $\lambda^* < 1$ where $m^*(\alpha, \phi_0, \lambda^*, \phi_1, \phi_{0,N}) = (1-\lambda^*)T-v$, and the ad-based plan for naïfs is more profitable when $\lambda < \lambda^*$ and the subscription-based plan is more profitable when $\lambda > \lambda^*$. This implies the subscription-based model is more likely as λ increases, which means the corresponding cutoff $\phi_0^*(\lambda, \phi_1, \phi_{0,N})$ is also increasing in λ.

Finally, we note that $m^*(\alpha, \hat{\phi}_0, \lambda, \phi_1, \phi_{0,N}) = T-v$ defines the cutoff for $\hat{\phi}_0$ whereas $m^*(\alpha, \phi_0^*, \lambda, \phi_1, \phi_{0,N}) = (1-\lambda)(T-v)$ defines the cutoff for ϕ_0^*. Because $m^*(\alpha, \cdot, \lambda, \phi_1, \phi_{0,N})$ is increasing and $(1-\lambda)(T-v) < T-v$, it is necessarily the case that $\phi_0^*(\lambda, \phi_1, \phi_{0,N}) < \hat{\phi}_0(\lambda, \phi_1, \phi_{0,N})$. ∎

Proof of Corollary 1. Regime (a) is exactly the same as it was in Proposition 4. For regime (b), the user welfare of the sophisticates is given by

$$\begin{aligned} W(S, x_i=0) &= T-P^* + qU(z_i^*(q,\hat{p}^*);\, \theta_i=1) + (1-q)U(z_i^*(q,\hat{p}^*);\, \theta_i=0) - \hat{p}^* z_i^*(q,\hat{p}^*) \\ &= v + qU(z_i^*(q,\hat{p}^*);\, \theta_i=1) + (1-q)U(z_i^*(q,\hat{p}^*);\, \theta_i=0) - \hat{p}^* z_i^*(q,\hat{p}^*) \\ &< v + qU(z_i^*(q,\bar{p}^*);\, \theta_i=1) + (1-q)U(z_i^*(q,\bar{p}^*);\, \theta_i=0) - \bar{p}^* z_i^*(q,\bar{p}^*) \\ &= W_{\text{base}}(S) \end{aligned}$$

The user welfare of the naïve agents in regime (b) is exactly as in regime (b) of Proposition 4. ∎

Proof of Proposition 6. We know that if $\phi_0 < \phi_0^*(\lambda, \phi_1, \phi_{0,N})$, the platform business model is fully subscription-based and $\hat{W}^*(\tau) = W_{\text{base}}$; in particular, it is constant over this entire range for ϕ_0. By

Corollary 1, there is a discontinuous jump in welfare down to $\hat{W}^*(\tau) < W_{\text{base}}$ at $\phi_0 = \phi_0^*(\lambda, \phi_1, \phi_{0,N})$. Thus, it just remains to show that $\hat{W}^*(\tau)$ is decreasing in ϕ_0 when $\phi_0 > \phi_0^*(\lambda, \phi_1, \phi_{0,N})$. We know that $\hat{p}^* = \frac{1}{2}\left(\lambda\beta q + (1-\lambda)\beta(q\bar{\pi}_{F_{N1}(\hat{\alpha}^*)} + (1-q)\bar{\pi}_{F_{N0}(\hat{\alpha}^*)}) + c\right)$ in equilibrium, and moreover we observed in Proposition 2 that $q\bar{\pi}_{F_{N1}(\hat{\alpha}^*)} + (1-q)\bar{\pi}_{F_{N0}(\hat{\alpha}^*)}$ is increasing in ϕ_0. Sophisticates' welfare is given by

$$\hat{W}^*(S) = v + qU(z_i^*(q,\hat{p}^*);\, \theta_i = 1) + (1-q)U(z_i^*(q,\hat{p}^*);\, \theta_i = 0) - \hat{p}^* z_i^*(q,\hat{p}^*)$$

which is monotonically decreasing in $\hat{p}^*$ (and thus ϕ_0). On the other hand, from Proposition 2, the welfare of naïve agents is given by

$$\hat{W}^*(N) = v - \Delta(\hat{\alpha}^*) + (I_S(\hat{\alpha}^*) - I_N(\hat{\alpha}^*)) + \frac{1}{2}(\beta q - \hat{p}^*)^2$$

Since $\hat{\alpha}^*$ is constant in ϕ_0, changing ϕ_0 has no bearing on $I_N(\hat{\alpha}^*)$. However, $I_S(\alpha)$ is monotonically decreasing and $\Delta(\alpha)$ is monotonically increasing in ϕ_0 for all α, by Blackwell's theorem (increasing ϕ_0 makes the sophisticates' signal generation process strictly less informative in the Blackwell order). At the same time, $\hat{W}^*(N)$ is monotonically decreasing in $\hat{p}^*$, which is increasing in ϕ_0 (as we saw in Proposition 2 and Proposition 5). Thus, $\hat{W}^*(N)$ is monotone decreasing in ϕ_0.

Because $\phi_0^*(\lambda, \phi_1, \phi_{0,N})$ is decreasing in ϕ_1 by Corollary 1, we know there exists $\phi_1^*(\lambda, \phi_0, \phi_{0,N})$ such that if $\phi_1 < \phi_1^*(\lambda, \phi_0, \phi_{0,N})$ the platform business model is fully subscription-based and $\hat{W}^*(\tau) = W_{\text{base}}$; in particular, it is constant over this entire range of ϕ_1. There is a discontinuous jump in welfare down to $\hat{W}^*(\tau) < W_{\text{base}}$ at $\phi_1 = \phi^*(\lambda, \phi_1, \phi_{0,N})$. Therefore, it just remains to show that $\hat{W}^*(\tau)$ is decreasing in ϕ_1 when $\phi_1 > \phi_1^*(\lambda, \phi_1, \phi_{0,N})$. By Blackwell's theorem, we know that $I_N(\alpha)$ is increasing in ϕ_1, which means that the platform's choice of advertising load $\hat{\alpha}^*$ is increasing in ϕ_1. At the same time $\hat{p}^*$ is increasing in $q\bar{\pi}_{F_{N1}(\hat{\alpha}^*)} + (1-q)\bar{\pi}_{F_{N0}(\hat{\alpha}^*)}$, which is increasing in ϕ_1 and $\hat{\alpha}^*$ (which is, in turn, increasing in ϕ_1). Sophisticates' welfare can be written as

$$\hat{W}^*(S) = v + qU(z_i^*(q,\hat{p}^*);\, \theta_i = 1) + (1-q)U(z_i^*(q,\hat{p}^*);\, \theta_i = 0) - \hat{p}^* z_i^*(q,\hat{p}^*),$$

which is monotonically decreasing in $\hat{p}^*$ (and thus, monotonically decreasing in ϕ_1). The welfare of naïve agents is given by

$$\hat{W}^*(N) = v - \Delta(\hat{\alpha}^*) + (I_S(\hat{\alpha}^*) - I_N(\hat{\alpha}^*)) + \frac{1}{2}(\beta q - \hat{p}^*)^2,$$

where we observe that holding $\hat{\alpha}^*$ constant, $I_S(\hat{\alpha}^*) - \Delta(\hat{\alpha}^*) - I_N(\hat{\alpha}^*)$ is non-increasing in ϕ_1. Also, observe that the first half of the expression for $\Delta(\alpha)$ in Proposition 1 cancels with $I_S(\alpha)$, leaving a difference between the naïve agents' *ex post* utility and their interim utility (according to their subjective model). The resulting final expression is negative and non-increasing in ϕ_1 (the difference is increasing in absolute value) because $\partial(q\bar{\pi}_{F_{N1}(\alpha)} + (1-q)\bar{\pi}_{F_{N0}(\alpha)})/\partial\phi_1 > 0$ (by Proposition 2) and $\partial(q\bar{\pi}_{F_{S1}(\alpha)} + (1-q)\bar{\pi}_{F_{S0}(\alpha)})/\partial\phi_1 = 0$ (by Lemma 1). Thus, $I_S(\hat{\alpha}^*) - \Delta(\hat{\alpha}^*) - I_N(\hat{\alpha}^*)$ is negative and proportional to $1 - e^{-\hat{\alpha}^* T}$, and finally we note that an increase in ϕ_1 raises $\hat{\alpha}^*$, which in turn increases $1 - e^{-\hat{\alpha}^* T}$ and $\hat{p}^*$, simultaneously. Combining these facts, the result is that $\hat{W}^*(N)$ is monotonically decreasing in ϕ_1. ∎

Proof of Proposition 7. The platform will always choose a plan that generates participation from both types of agents because it can always offer the fully subscription-based plan with $P^* = T - v$ to extract some surplus from any non-participating type. Thus, if offering a single plan is a best response for the platform, then the stricter of the two types' participation constraint will bind:

$$\begin{aligned} \max_{\alpha,p,P} \quad & (p-c)(\lambda\beta q + (1-\lambda)\beta q\bar{\pi}_{F_{N1}(\alpha)} + (1-\lambda)\beta(1-q)\bar{\pi}_{F_{N0}(\alpha)} - p) + P \\ \text{subject to} \quad & I_S(\alpha) + (1-\alpha)T - v - P \geq 0, \end{aligned}$$

for which there exists some α^* and corresponding $P^* = I_S(\alpha^*) + (1-\alpha^*)T - v$ that maximizes the above expression. Such (α^*, P^*) is the only candidate for a mixed business model where only one plan is offered. Note also that $\alpha^* \geq \alpha_S^{FB}$ because the platform could increase P^* and $\beta q + (1-\lambda)\beta q\bar{\pi}_{F_{N1}(\alpha)} + (1-\lambda)\beta(1-q)\bar{\pi}_{F_{N0}(\alpha)}$ by increasing α up to α_S^{FB}, which monotonically increases its objective (by Lemmas 2 and 3). Note that $\alpha^* < \hat{\alpha}^*$ because $I_S(\hat{\alpha}^*) + (1-\hat{\alpha}^*)T - v < 0$ by construction in Proposition 2.

In the case of a mixed business model with two plans offered, (α_1^*, P_1^*) and (α_2^*, P_2^*), we maximize the firm's profit subject to the incentive compatibility constraint between the naïve and sophisticated users and their corresponding participation constraints:

$$\begin{aligned} \max_{\alpha_1,\alpha_2,p,P_1,P_2} \quad & (p-c)(\lambda\beta q + (1-\lambda)\beta q\bar{\pi}_{F_{N1}(\alpha_2)} + (1-\lambda)\beta(1-q)\bar{\pi}_{F_{N0}(\alpha_2)} - p) + \lambda P_1 + (1-\lambda)P_2 \\ \text{subject to} \quad & P_2 - P_1 - I_S(\alpha_2) + I_S(\alpha_1) \geq 0 \\ & P_1 - P_2 + I_N(\alpha_2) - I_N(\alpha_1) \geq 0 \\ & I_S(\alpha_1) + (1-\alpha_1)T - v - P_1 \geq 0 \\ & I_N(\alpha_2) + (1-\alpha_2)T - v - P_2 \geq 0, \end{aligned}$$

which yields some profit-maximizing business model (α_1^*, P_1^*) and (α_2^*, P_2^*) together with the profit-maximizing price $p^* = \frac{1}{2}(\lambda\beta q + (1-\lambda)\beta q\bar{\pi}_{F_{N1}(\alpha_2)} + (1-\lambda)\beta q\bar{\pi}_{F_{N0}(\alpha_2)} + c)$. The platform then compares the profits under the single plan, (α^*, P^*), and under the two plans, (α_1^*, P_1^*) and (α_2^*, P_2^*).

We now show that (i) α_2^* is increasing in ϕ_0, (ii) attains at least α_S^{FB}, and (iii) never exceeds $\hat{\alpha}^*$, in the profit-maximizing business model, thus establishing the form of the cutoff $\tilde{\phi}_0(\lambda, \phi_1, \phi_{0,N})$. The first claim follows because $\beta q\bar{\pi}_{F_{N1}(\alpha_2^*)} + \beta q\bar{\pi}_{F_{N0}(\alpha_2^*)}$ is increasing in ϕ_0 while leaving all of the naïfs' constraints (both IC and PC) unaffected, which therefore increases the firm's profit from advertising by Lemma 1. For the second claim, note that if $\alpha_1^* \leq \alpha_2^* < \alpha_S^{FB}$, then the platform can increase the ad load of both α_1^* and α_2^* without needing to reduce P_1^* and P_2^*, which therefore leads to higher profit, yielding a contradiction. The third claim is a direct consequence of the PC constraint of naïve agents and that for any $\alpha > \hat{\alpha}^*$, we have $I_N(\alpha) + (1-\alpha)T - v < 0$. The same comparative statics with respect to $\phi_{0,N}$, ϕ_1, and λ readily follow as in Proposition 5.

Finally, we establish that $\tilde{W}^*_{(a)}(\tau) \leq W^*_{\text{base}}(\tau)$ and that $\tilde{W}^*_{(b)}(\tau) < \tilde{W}^*_{\text{base}}(\tau)$. We leverage the arguments from Proposition 4 and Corollary 1. When a single plan is offered, we know that the platform will set α so that $I_S(\alpha^*) + (1-\alpha^*)T - v - P^* = 0$. At the same time, $\beta q\bar{\pi}_{F_{N1}(\alpha^*)} + \beta q\bar{\pi}_{F_{N0}(\alpha^*)} \geq q$, making the product price $p^* \geq \bar{p}^*$. For the same reasons as in Corollary 1, this will lead to (weakly) lower user welfare relative to the base case for both types of users. When multiple plans are offered, we know that either $I_S(\alpha_1^*) + (1-\alpha_1^*)T - v - P_1^* = 0$ or $I_N(\alpha_2^*) + (1-\alpha_2^*)T - v - P_2^* = 0$, because otherwise both P_1^* and P_2^* could be increased without affecting the incentive compatibility constraints and increasing the platform's profit. If $I_N(\alpha_2^*) + (1-\alpha_2^*)T - v = P_2^*$, then $I_S(\alpha_2^*) + (1-\alpha_2^*)T - v < P_2^*$ by Lemma 3, which means that the sophisticates would choose their outside option over participating on the naïfs' plan, and thus the platform can also set $P_1^* = I_S(\alpha_1^*) + (1-\alpha_1^*)T - v$ without affecting incentive compatibility. Otherwise, $I_S(\alpha_1^*) + (1-\alpha_1^*)T - v = P_1^*$, but then the sophisticates' incentive compatibility constraint implies (by plugging in P_1^*) that $P_2^* - (I_S(\alpha_1^*) + (1-\alpha_1^*)T - v) - I_S(\alpha_2^*) + I_S(\alpha_1^*) \geq 0$ which in turn suggests $I_S(\alpha_2^*) + (1-\alpha_1^*)T - v - P_2^* < 0$, and since $\alpha_1^* \leq \alpha_2^*$, $I_S(\alpha_2^*) + (1-\alpha_2^*)T - v - P_2^* < 0$. In both cases, both sophisticated and naïve agents receive at most the welfare of $v + \frac{1}{2}(\beta q - p^*)^2$, where $p^* \geq \bar{p}^*$, which is no more than base case welfare. It is then straightforward to see that $\tilde{W}^*_{(b)}(\tau) < \tilde{W}^*_{\text{base}}(\tau)$ because $p^* > \bar{p}^*$ when there is a positive level of advertising $\hat{\alpha}^*$ for naïfs. ∎

A.4 Proofs from Section 5

Proof of Proposition 8. As in our baseline game, all decisions for $t = 3, 4, 5$ are generically unique by backward induction, so it suffices to consider just $t \leq 2$. At time $t = 2$, we take the platform advertising rates $\{(\alpha_{1,\rho}^{(j)}, \alpha_{2,\rho}^{(j)})\}_{j=1}^N$ as given. Because product prices are chosen sequentially, Zermelo's theorem guarantees there exists an equilibrium vector of prices $(p_1^*, \ldots, p_N^*)$ attained from backward induction. To show that this is the unique choice of pricing vectors, we just need to prove that, generically, no firm is indifferent between selecting two prices. First, observe that no firm will choose a price below marginal cost in equilibrium – the firm with the lowest price will make negative economic profits and has a profitable deviation to charge any price above marginal cost, which guarantees at least zero profits. Thus, each firm will charge at least $p^* = \min\{p \,|\, p = k\varepsilon > c,\ k \in \mathbb{N}\}$ in equilibrium. The firm j will choose p_j to maximize $\max_{p_j}(p_j - c)\mathbb{E}_{\boldsymbol{\pi}_i \sim \{(\alpha_{1,\rho}^{(j)}, \alpha_{2,\rho}^{(j)})\}_{j=1}^N}[(\pi_i^{(j)}\beta - p_j) \cdot \mathbb{1}_{\beta\pi_i^{(j)} - p_j \in \max_\omega\{\beta\pi_i^{(\omega)} - p_\omega\}}]$, where the prices $\{p_\omega\}_{\omega \neq j}$ are taken as given for all firms who price before firm j and taken as a best-response function of p_j for all firms that price after firm j, with $\mathbb{1}_{\beta\pi_i^{(j)} - p_j \in \max_\omega\{\beta\pi_i - p_\omega\}} = (|\{j : \beta\pi_i^{(j)} - p_j = \max_\omega \beta\pi_i^{(\omega)} - p_\omega\}|)^{-1}$. It is clear that p_j^* is upper bounded by the monopoly price, which we denote by $\tilde{p}^* = \arg\max_{p_j}(p_j - c)\mathbb{E}_{\boldsymbol{\pi}_i \sim \{(\alpha_{1,\rho}^{(j)}, \alpha_{2,\rho}^{(j)})\}_{j=1}^N}[\pi_i^{(j)}\beta - p_j]$, so it is without loss of generality to restrict the firm's equilibrium choice of prices to $p_j^* \in \{p^*, p^* + \varepsilon, \ldots, (k^* + 1)\varepsilon\}$ where $k^* = \max\{k \in \mathbb{N} \,|\, k\varepsilon < \tilde{p}^*\}$.

We proceed by induction on the reverse order of sequential offers. For firm N, all prices $(p_1^*, \ldots, p_{N-1}^*)$ are fixed. Note that unless $\mathbb{P}_{\boldsymbol{\pi}_i \sim \{(\alpha_{1,\rho}^{(j)}, \alpha_{2,\rho}^{(j)})\}_{j=1}^N}[\beta\pi_i^{(N)} - p_N \in \max_\omega\{\beta\pi_i^{(\omega)} - p_\omega\}] = 0$, generically, none of the prices $p_N \in \{p^*, p^* + \varepsilon, \ldots, (k^* + 1)\varepsilon\}$ yield the same profit. Moreover, if $\mathbb{P}_{\boldsymbol{\pi}_i \sim \{(\alpha_{1,\rho}^{(j)}, \alpha_{2,\rho}^{(j)})\}_{j=1}^N}[\beta\pi_i^{(N)} - p_N \in \max_\omega\{\beta\pi_i^{(\omega)} - p_\omega\}] = 0$ for some pricing strategy p_N, then the firm earns zero profits with certainty, and there is a profitable deviation by pricing at p^*, which guarantees strictly positive profits (given that all firms price at p^* or above in a candidate equilibrium). Thus, we can retain a subset, $\mathcal{P}_N$, of $\{p^*, p^* + \varepsilon, \ldots, (k^* + 1)\varepsilon\}$ where $\mathbb{P}_{\boldsymbol{\pi}_i \sim \{(\alpha_{1,\rho}^{(j)}, \alpha_{2,\rho}^{(j)})\}_{j=1}^N}[\beta\pi_i^{(N)} - p_N \in \max_\omega\{\beta\pi_i^{(\omega)} - p_\omega\}] > 0$, which generically yields positive and distinct payoffs for $(p_N - c)\mathbb{E}_{\pi_i \sim \{(\alpha_{1,\rho}^{(j)}, \alpha_{2,\rho}^{(j)})\}_{j=1}^N}[(\pi_i^{(N)}\beta - p_N) \cdot \mathbb{1}_{\beta\pi_i^{(N)} - p_N \in \max_\omega\{\beta\pi_i^{(\omega)} - p_\omega\}}]$ for all $p_N \in \mathcal{P}_N$. Thus, generically, there is a unique p_N^* that maximizes this expression. Finally, note that because the set $\{p^*, p^* + \varepsilon, \ldots, (k^* + 1)\varepsilon\}$ is discrete, firm N's best response choice of p_N^* is insensitive to sufficiently small perturbations in model parameters (such as β).

We leverage this fact to argue the inductive step. For all $p_j \in \{p^*, p^* + \varepsilon, \ldots, (k^* + 1)\varepsilon\}$, we can once again rule out any p_j^* with $\mathbb{P}_{\boldsymbol{\pi}_i \sim \{(\alpha_{1,\rho}^{(j)}, \alpha_{2,\rho}^{(j)})\}_{j=1}^N}[\pi_i^{(j)} - p_j \in \max_\omega\{\pi_i^{(\omega)} - p_\omega\}] = 0$ as an equilibrium strategy, because pricing at p^* results in strictly positive profits. If we have two prices p_j', p_j'' with both $\mathbb{P}_{\boldsymbol{\pi}_i \sim \{(\alpha_{1,\rho}^{(j)}, \alpha_{2,\rho}^{(j)})\}_{j=1}^N}[\pi_i^{(j)} - p_j' \in \max_\omega\{\pi_i^{(\omega)} - p_\omega\}] > 0$ and $\mathbb{P}_{\boldsymbol{\pi}_i \sim \{(\alpha_{1,\rho}^{(j)}, \alpha_{2,\rho}^{(j)})\}_{j=1}^N}[\pi_i^{(j)} - p_j'' \in \max_\omega\{\pi_i^{(\omega)} - p_\omega\}] > 0$ that yield identical profits, then one can introduce a small perturbation to β which has no effect on the best response prices of firms $j+1, \ldots, N$ but breaks the exact equality between $(p_j' - c)\mathbb{E}_{\pi_i \sim \{(\alpha_{1,\rho}^{(j)}, \alpha_{2,\rho}^{(j)})\}_{j=1}^N}[(\pi_i^{(j)}\beta - p_j') \cdot \mathbb{1}_{\beta\pi_i^{(\omega)} - p_j' \in \max_\omega\{\beta\pi_i^{(\omega)} - p_\omega\}}]$ and $(p_j'' - c)\mathbb{E}_{\pi_i \sim \{(\alpha_{1,\rho}^{(j)}, \alpha_{2,\rho}^{(j)})\}_{j=1}^N}[(\pi_i^{(j)}\beta - p_j'') \cdot \mathbb{1}_{\beta\pi_i^{(\omega)} - p_j'' \in \max_\omega\{\beta\pi_i^{(\omega)} - p_\omega\}}]$, making firm j's indifference a knife edge case. There are at most finitely many of these knife-edge cases because pricing is discrete, and these can be disregarded under generic conditions. Therefore, generically we once again get a unique $p_j^* \in \mathcal{P}_j$ that maximizes firm profit, completing the inductive step and establishing the pricing vector $(p_1^*, \ldots, p_N^*)$ as unique. Moreover, this unique pricing vector will be order-independent in the allocation because all firms are ex-ante identical at the beginning of time step $t = 2$.

For $t = 1$, we consider the case of a single platform ($M = 1$) and multiple platforms ($M > 1$) separately. When there is a single platform, there generically exists a unique $(\alpha_*^{(1)}, \ldots, \alpha_*^{(N)})$ (once again, in allocation, up to a renumbering of the firms) which maximizes the transfers $\sum_{j=1}^N m^{(j)}$ the platform can collect from the firms, by extracting the full surplus the firms gain from the advertising. The firms always accept such a contract in equilibrium. When there are multiple platforms, they simultaneously compete to offer the plans most desired by the sophisticated and naïve users. As described in the equilibrium of Proposition 11, these plans have no subscription fees and cater exactly to the advertising levels desired by sophisticates and naïfs out of all feasible plans. These are given by $(\tilde{\alpha}_S^{FB}, \ldots, \tilde{\alpha}_S^{FB})$ and $(\tilde{\alpha}_N^*, \ldots, \tilde{\alpha}_N^*)$ for some $\tilde{\alpha}_S^{FB}$ and $\tilde{\alpha}_N^*$ that maximize $I_S(\alpha^{(1)}, \ldots, \alpha^{(N)}) + (1 - \sum_{j=1}^N \alpha^{(j)})T$ and $I_N(\alpha^{(1)}, \ldots, \alpha^{(N)}) + (1 - \sum_{j=1}^N \alpha^{(j)})T$, respectively (see Appendix C). Platforms simultaneously compete over transfers to firms, driving these transfers to zero in the unique equilibrium, which the firms always accept. The equilibrium is unique because we can fully characterize the unique values of $\tilde{\alpha}_S^{FB}$ and $\tilde{\alpha}_N^*$. ∎

Proof of Proposition 9. Under firm competition, we have that $I_S(\alpha^{(1)}, \alpha^{(2)})$ takes the form of $I_S(\alpha^{(1)}, \alpha^{(2)}) = (1 - e^{-\alpha^{(1)}T})e^{-\alpha^{(2)}T}\chi_S^{(1)} + (1 - e^{-\alpha^{(2)}T})e^{-\alpha^{(1)}T}\chi_S^{(2)} + (1 - e^{-\alpha^{(1)}T})(1 - e^{-\alpha^{(2)}T})\chi_S^{(1,2)}$ for some $\chi_S^{(1)}, \chi_S^{(2)}, \chi_S^{(1,2)}$ that depend on model primitives (e.g., ϕ_0 and ϕ_1). Similarly, $I_N(\alpha^{(1)}, \alpha^{(2)})$ takes the form of $I_N(\alpha^{(1)}, \alpha^{(2)}) = (1 - e^{-\alpha^{(1)}T})e^{-\alpha^{(2)}T}\chi_N^{(1)} + (1 - e^{-\alpha^{(2)}T})e^{-\alpha^{(1)}T}\chi_N^{(2)} + (1 - e^{-\alpha^{(1)}T})(1 - e^{-\alpha^{(2)}T})\chi_N^{(1,2)}$ for some $\chi_N^{(1)} > \chi_S^{(1)}$, $\chi_N^{(2)} > \chi_S^{(2)}$, and $\chi_N^{(1,2)} > \chi_S^{(1,2)}$.

For regime (i), notice that when $\phi_0 = \phi_{0,N} = \phi_1 = 0$, beliefs of the agents do not change as a result of advertising, and therefore the firms are not willing to transfer any amount to the platform from an ad contract. At the same time, we observe that there exists $\bar{\phi}_1$ such that for all $\phi_1 < \bar{\phi}_1$, the maximal subscription fee the platform can generate for both sophisticates and naïfs occurs where $\alpha^{(1)} = \alpha^{(2)} = 0$. This is a direct consequence of noting that $\partial I_S(\alpha^{(1)}, \alpha^{(2)})/\partial\alpha^{(1)} = Te^{-\alpha^{(1)}T}e^{-\alpha^{(2)}T}\chi_S^{(1)} + Te^{-\alpha^{(1)}T}(1 - e^{-\alpha^{(2)}T})(\chi_S^{(1,2)} - \chi_S^{(2)}) < T$ for all $\alpha^{(1)}$ and $\alpha^{(2)}$, for ϕ_1 sufficiently close to 0, because $\chi_S^{(1)}, \chi_S^{(2)}$

and $\chi_S^{(1,2)}$ are sufficiently close to 0 (and the same for $\partial I_S(\alpha^{(1)}, \alpha^{(2)})/\partial\alpha^{(2)}$, $\partial I_N(\alpha^{(1)}, \alpha^{(2)})/\partial\alpha^{(1)}$, and $\partial I_N(\alpha^{(1)}, \alpha^{(2)})/\partial\alpha^{(2)}$). So, in particular, $\alpha^{(1)} = \alpha^{(2)} = 0$ maximizes both $I_S(\alpha^{(1)}, \alpha^{(2)}) + (1 - \alpha^{(1)} - \alpha^{(2)})T$ and $I_N(\alpha^{(1)}, \alpha^{(2)}) + (1 - \alpha^{(1)} - \alpha^{(2)})T$ for all $\phi_1 < \bar{\phi}_1$. Thus, at $\phi_1 = 0$, the profit-maximizing business model of the platform is to set $\alpha^{(1)} = \alpha^{(2)} = 0$ and extract full surplus from a subscription fee $P^* = T - v$. We also know that user i will purchase all of the product which has the largest $\pi_i - p_j^*$, which is equivalent to largest $q - p_j^*$ when there is no advertising, reducing to standard Bertrand competition on price without horizontal differentiation. This leads to $p_1^* = p_2^* = c$ as the unique equilibrium under regime (i). Finally, by Blackwell's theorem, we know that the transfers from the firms will be monotonically increasing in ϕ_1 for every pair $(\alpha^{(1)}, \alpha^{(2)})$, therefore, we can choose $\phi_1^{\mathcal{F}} \leq \bar{\phi}_1$ to be maximal such that for any $\phi_1 > \phi_1^{\mathcal{F}}$, there is an ad-based plan offered in equilibrium.

In regime (ii), let us consider $\phi_1 > \phi_1^{\mathcal{F}}$ and $\phi_0 = \phi_1$. We note that $\chi_S^{(1)} = \chi_S^{(2)} = \chi_S^{(1,2)} = 0$, and therefore the platform can extract the maximal amount of revenue from the sophisticates by charging the subscription fee $T - v$ and setting $\alpha^{(1)} = \alpha^{(2)} = 0$, as in regime (i). At the same time, under the fully-rational benchmark, for the same reasons as in the previous paragraph, the equilibrium involves the platform offering one subscription-based plan and the firms competing over price alone to offer $\bar{p}_1^* = \bar{p}_2^* = c$. However, by construction of $\phi_1^{\mathcal{F}}$, we know that $\alpha_2^{(1)*} + \alpha_2^{(2)*} > 0$ and the naïfs opt into the ad-based plan in the equilibrium. Without loss of generality, let $\alpha_2^{(1)*} > 0$. Then with probability $\phi_0(1 - e^{-\alpha_2^{(1)*}})e^{-\alpha_2^{(2)*}} > 0$, the user sees a positive ad from firm 1 and no ad from firm 2, leading to belief $\pi_N^{(1)}(G) > q$ about product 1 but belief $\pi_N^{(2)} = q$ about product 2. We know that pricing at or below marginal cost is dominated by a strategy that sets $p_1^* = c + \beta(\pi_N(G) - q)$, which attains strictly positive profits; thus, $\hat{p}_1^* > \bar{p}_1^* = c$ in equilibrium. At the same time, with probability $(1 - \phi_0)(1 - e^{-\alpha_2^{(1)*}})e^{-\alpha_2^{(2)*}} > 0$, the user sees a negative ad from firm 1 and no ad from firm 2, leading to a belief $\pi_N^{(1)}(B) < q$ about product 1 but belief $\pi_N^{(2)} = q$ about product 2. Similarly,we know that for firm 2 pricing at or below marginal cost is a is dominated by a strategy that sets $p_2^* = c + \beta(q - \pi_N(B))$, which attains strictly positive profits. Thus, firm 2 will set $\hat{p}_2^* > \bar{p}_2^* = c$ in equilibrium. The claim in (ii) follows by choosing the minimal $\phi_0^{\mathcal{F}}(\phi_1)$ where this property holds. ∎

Proof of Proposition 10. In the fully-rational benchmark of regimes (i) and (ii) we have that the platform offers a subscription-based plan with $P^* = T - v$ and firms price at marginal cost $\bar{p}_1^* = \bar{p}_2^* = c$. The equilibrium played in regime (i) corresponds exactly to the fully-rational benchmark, so we obtain that $\hat{W}_{2,1}(\tau)^* = \bar{W}_{2,1}^*(\tau)$ for both user types $\tau \in \{S, N\}$. For regime (ii), advertising load is at $\alpha_2^* = \alpha_2^{(1)*} + \alpha_2^{(2)*} > 0$ for naïfs and prices are at $\hat{p}_1^* > c$ and $\hat{p}_2^* > c$. Sophisticates subscribe to the same plan in equilibrium as they do under the fully-rational benchmark, but face strictly higher prices in the product market, so $\hat{W}_{2,1}^*(S) < \bar{W}_{2,1}^*(S)$. Naïve agents, on the other hand, receive welfare of at most $\hat{W}_{2,1}^*(S) + I_S(\alpha^{(1)*}, \alpha^{(2)*}) - (\alpha^{(1)*} + \alpha^{(2)*})T - \Delta(\alpha^{(1)*}, \alpha^{(2)*}) < \hat{W}_{2,1}^*(S) < \bar{W}_{2,1}^*(S) = \bar{W}_{2,1}^*(N)$, by construction of $\phi_1^{\mathcal{F}}$ and $\phi_0^{\mathcal{F}}(\phi_1)$ that $(\alpha^{(1)}, \alpha^{(2)}) = (0, 0)$ maximizes $I_S(\alpha^{(1)}, \alpha^{(2)}) + (1 - \alpha^{(1)} - \alpha^{(2)})T$. ∎

Proof of Proposition 11. We claim that both platforms compete to offer two plans with advertising intensities $\alpha_N^{1,2}$ and $\alpha_S^{1,2}$ at prices $P_N^{1,2} = 0$ and $P_S^{1,2} = 0$ to naïfs and sophisticates, respectively, where $\alpha_N^{1,2} \in \arg\max_\alpha I_N(\alpha) + (1 - \alpha)T$ and $\alpha_S^{1,2} \in \arg\max_\alpha I_S(\alpha) + (1 - \alpha)T$. Note that if these two plans are offered then naïfs and sophisticates will choose their respective plan regardless of whether there is another plan $(\tilde{\alpha}, \tilde{P})$ offered. By way of contradiction, suppose that sophisticates participate on platform 1's plan $(\tilde{\alpha}_S^{1,2}, P_S)$ where $\tilde{\alpha}_S^{1,2} \neq \alpha_S^{1,2}$ or $P_S \neq 0$. If $P_S > 0$, then platform 2 has a profitable deviation to offer $(\tilde{\alpha}^{1,2}, P_S - \epsilon)$ for $\epsilon > 0$, whereas if $\tilde{\alpha}_S^{1,2} \neq \alpha_S^{1,2}$, then platform 2 has a profitable deviation to offer $((\tilde{\alpha}^{1,2} + \alpha_S^{1,2})/2, P_S + \epsilon)$ for $\epsilon > 0$ sufficiently small. The same is true for naïve agents.

By Blackwell's theorem, $I_S(\alpha)$ and $I_N(\alpha)$ are both increasing in ϕ_1 and by Lemma A.1, $I_S(\alpha) = I_N(\alpha) = 0$ when $\phi_1 = 0$. Recall that by the same reasoning as in Lemma 3, we know that $I_N(\alpha) = (1 - e^{-\alpha T})\chi_N(\phi_1)$ and $I_S(\alpha) = (1 - e^{-\alpha T})\chi_S(\phi_1)$ with $\chi_N(0) = \chi_S(0) = 0$. We observe that $I'_N(\alpha) = Te^{-\alpha T}\chi_N(\phi_1) < I'_N(0) = T\chi_N(\phi_1)$. Moreover, there exists some $\phi_1^{\mathcal{P}} \in (0, 1)$ such that for all $\phi_1 < \phi_1^{\mathcal{P}}$, $\chi_N(\phi_1) < 1$, implying that $I'_N(\alpha) < T$ for all $\alpha > 0$, thus $\alpha_N^{1,2} = 0$. This in turn suggests that $I'_S(\alpha) < T$ for all $\alpha > 0$, and $\alpha_S^{1,2} = 0$. Therefore, the equilibrium under regime (i) involves $\alpha_N^{1,2} = \alpha_S^{1,2} = 0$ and $P_N^{1,2} = P_S^{1,2} = 0$. In the fully-rational benchmark, all agents are sophisticated, so the firm solves $\max_p (p - c)(\beta q - p)$ by Lemma 1 regardless of the level of advertising in equilibrium. When there is no advertising, the firm solves exactly the same problem, so $p^* = \bar{p}^*$ in regime (i).

We choose $\phi_1^{\mathcal{P}}$ to be maximal so that for any $\phi_1 > \phi_1^{\mathcal{P}}$, $\chi_N(\phi_1) > 1$ and $I'_N(0) > T$, which implies the advertising level to naïve agents in equilibrium will be such that $\alpha_N^{1,2} > 0$. Note that $\alpha_N^{1,2}$ solves $I'_N(\alpha_N^{1,2}) = T$, which necessarily occurs before $\hat{\alpha}^*$ where $I_N(\hat{\alpha}^*) + (1 - \hat{\alpha}^*)T - v = 0$. Similarly, because $I'_S(\alpha) < I'_N(\alpha)$, we know that $\alpha_N^{1,2} > \alpha_S^{1,2} = \alpha_S^{FB}$. In the fully-rational benchmark, we have that $\bar{p}^* = (\beta q + c)/2$ whereas the firm will charge $p^* = \frac{1}{2}\left(\lambda\beta q + (1 - \lambda)\beta(q\bar{\pi}_{F_{N1}(\alpha_N^{1,2})} + (1 - q)\bar{\pi}_{F_{N0}(\alpha_N^{1,2})}) + c\right) > (\beta q + c)/2$ for $\alpha_N^{1,2} > 0$ when $\phi_1 > \phi_1^{\mathcal{P}}$. ∎

Proof of Proposition 12. In the fully-rational benchmark, all agents receive advertising at the rate α_S^{FB}, zero subscription fees, and $p^* = \bar{p}^*$. Under regime (i), we have $\alpha_S^{FB} = 0$ so the equilibrium played is exactly that of the fully-rational benchmark; trivially, we have $\hat{W}_{1,2}^*(\tau) = \bar{W}_{1,2}^*(\tau)$ for both $\tau \in \{S, N\}$.

For regime (b), advertising load is at $\alpha_1^* = \alpha_S^{FB}$ and $\alpha_2^* \in (\alpha_S^{FB}, \hat{\alpha}^*)$ and prices are at $\hat{p}_{\mathcal{P}}^*$. We have the user welfare for the sophisticated agents is given by

$$\begin{aligned}
\hat{W}_{1,2}^*(S) =&(1 - \alpha_S^{FB})T + q\mathbb{E}_{\pi \sim F_{S1}}\left[U(z^*(\pi, \hat{p}_{\mathcal{P}}^*);\, \theta_i = 1) - \hat{p}_{\mathcal{P}}^* z^*(\pi, \hat{p}_{\mathcal{P}}^*)\right] \\
&+ (1 - q)\mathbb{E}_{\pi \sim F_{S0}}\left[U(z^*(\pi, \hat{p}_{\mathcal{P}}^*);\, \theta_i = 0) - \hat{p}_{\mathcal{P}}^* z^*(\pi, \hat{p}_{\mathcal{P}}^*)\right] \\
=&(1 - \alpha_S^{FB})T + I_S(\alpha_S^{FB}) + U(z_i^*(q, \hat{p}_{\mathcal{P}}^*)) - \hat{p}_{\mathcal{P}}^* z_i^*(q, \hat{p}_{\mathcal{P}}^*) = \bar{W}_{1,2}^*(S) + \frac{1}{2}(\beta q - \hat{p}_{\mathcal{P}}^*)^2 - \frac{1}{2}(\beta q - \bar{p}^*)^2 < \bar{W}_{1,2}^*(S)
\end{aligned}$$

Consumer welfare for the naïve agents is given by

$$\begin{aligned}
\hat{W}_{1,2}^*(N) =&(1 - \alpha_2^*)T + q\mathbb{E}_{\pi \sim F_{N1}}\left[U(z^*(\pi, \hat{p}_{\mathcal{P}}^*);\, \theta_i = 1) - \hat{p}_{\mathcal{P}}^* z^*(\pi, \hat{p}_{\mathcal{P}}^*)\right] \\
&+ (1 - q)\mathbb{E}_{\pi \sim F_{N0}}\left[U(z^*(\pi, \hat{p}_{\mathcal{P}}^*);\, \theta_i = 0) - \hat{p}_{\mathcal{P}}^* z^*(\pi, \hat{p}_{\mathcal{P}}^*)\right] \\
=&(1 - \alpha_2^*)T + q\mathbb{E}^S_{\pi_i|\theta_i=1}[U(C_i(\pi_i, \hat{p}_{\mathcal{P}}^*);\, \theta_i = 1) - \hat{p}_{\mathcal{P}}^* z_i^*(\pi_i, \hat{p}_{\mathcal{P}}^*) \,|\, \alpha_2^*] \\
&+ (1 - q)\mathbb{E}^S_{\pi_i|\theta_i=0}[U(z_i^*(\pi_i, \hat{p}_{\mathcal{P}}^*);\, \theta_i = 0) - \hat{p}_{\mathcal{P}}^* z_i^*(\pi_i, \hat{p}_{\mathcal{P}}^*) \,|\, \alpha_2^*] - \Delta(\alpha_2^*) \\
=&(1 - \alpha_2^*)T + I_S(\alpha_2^*) - \Delta(\alpha_2^*) + U(z_i^*(q, \hat{p}_{\mathcal{P}}^*)) - \hat{p}_{\mathcal{P}}^* z_i^*(q, \hat{p}_{\mathcal{P}}^*) \\
<&(1 - \alpha_S^{FB})T + I_S(\alpha_S^{FB}) + (I_S(\alpha_2^*) - I_N(\alpha_2^*)) + \frac{1}{2}(\beta q - \hat{p}_{\mathcal{P}}^*)^2 < \bar{W}_{1,2}^*(N) + \frac{1}{2}(\beta q - \hat{p}_{\mathcal{P}}^*)^2 - \frac{1}{2}(\beta q - \bar{p}^*)^2,
\end{aligned}$$

which is strictly less than $\bar{W}_{1,2}^*(N)$, establishing the welfare claim. ∎

A.5 Proofs from Section 6

Proof of Proposition 13. First, we claim that if the menu offered is (α_1, P_1, p_1) and (α_2, P_2, p_2) with $\alpha_2 > \alpha_1$, and the sophisticates self-select into the plan with advertising, α_2, then so do the naïfs. By assumption, we know that $I_S(\alpha_1) + (1-\alpha_1)T - P_1 + \frac{1}{2}(\beta q - p_1)^2 < I_S(\alpha_2) + (1-\alpha_2)T - P_2 + \frac{1}{2}(\beta q - p_2)^2$, which implies that $(e^{-\alpha_1 T} - e^{-\alpha_2 T})\chi_S > (\alpha_2 - \alpha_1)T - P_1 + P_2 + \frac{1}{2}(\beta q - p_1)^2 - \frac{1}{2}(\beta q - p_2)^2$, where χ_S is defined as in Lemma 3. Therefore, $(e^{-\alpha_1 T} - e^{-\alpha_2 T})\chi_N > (\alpha_2 - \alpha_1)T - P_1 + P_2 + \frac{1}{2}(\beta q - p_1)^2 - \frac{1}{2}(\beta q - p_2)^2$ given $\chi_N > \chi_S$, and rearranging gives us $I_N(\alpha_1) + (1-\alpha_1)T - P_1 + \frac{1}{2}(\beta q - p_1)^2 < I_N(\alpha_2) + (1-\alpha_2)T - P_2 + \frac{1}{2}(\beta q - p_2)^2$.

Second, we claim that if the menu offered is again (α_1, P_1, p_1) and (α_2, P_2, p_2), but the sophisticates opt for the plan with advertising α_1, then welfare is larger if naïve agents also choose α_1 instead of α_2. Recall that naïfs' welfare can be written as $I_S(\alpha) + (1-\alpha)T - \Delta(\alpha) - P + \frac{1}{2}(\beta q - p)^2$, where $\Delta(\alpha)$ is an increasing function in α. However, $I_S(\alpha_1) + (1-\alpha_1)T - P_1 + \frac{1}{2}(\beta q - p_1)^2 > I_S(\alpha_2) + (1-\alpha_2)T - P_2 + \frac{1}{2}(\beta q - p_2)^2$ and $\Delta(\alpha_1) < \Delta(\alpha_2)$; thus, $I_S(\alpha_1) + (1-\alpha_1)T - \Delta(\alpha_1) - P_1 + \frac{1}{2}(\beta q - p_1)^2 > I_S(\alpha_2) + (1-\alpha_2)T - \Delta(\alpha_2) - P_2 + \frac{1}{2}(\beta q - p_2)^2$, implying that welfare for naïfs is higher under plan with advertising load α_1.

We chose α^{SB} to maximize welfare given by $\lambda W(S;\alpha^{SB}) + (1-\lambda)W(N;\alpha^{SB})$. The lower bound on α^{SB} trivially holds when $\alpha_N^{FB} = 0$, so suppose $\alpha_N^{FB} > 0$. Then we know that $\lambda W'(S;\alpha) + (1-\lambda)W'(N;\alpha) > 0$ for all $\alpha \leq \alpha_N^{FB}$ because $I_S'(\alpha) > 0$ and $I_S'(\alpha) - \Delta'(\alpha) \geq 0$, so choosing load α leads to lower welfare than choosing some $\alpha_N^{FB} + \epsilon$ for some small $\epsilon > 0$. Likewise, if $\alpha_S^{FB} = 0$, then it straightforward to see $\alpha^{SB} = 0$, so let us take $\alpha_S^{FB} > 0$. Then we know for all $\alpha \geq \alpha_S^{FB}$ that $\lambda W'(S;\alpha_S^{FB}) + (1-\lambda)W'(N;\alpha) < 0$ because $I_S'(\alpha) = 0$ and $I_S'(\alpha) - \Delta'(\alpha) < 0$, so choosing α leads to lower user welfare than choosing $\alpha_S^{FB} - \epsilon$ for small $\epsilon > 0$. This implies that α^{SB} lies somewhere in the interval $[\alpha_N^{FB}, \alpha_S^{FB}]$.

To see that $W_{FB}(\tau) > W_{SB}(\tau)$ for both $\tau \in \{S, N\}$, we consider the construction of α^{SB}, α_S^{FB}, and α_N^{FB}. We know that when $\alpha_N^{FB} > 0$, then $\alpha^{SB} \in (\alpha_N^{FB}, \alpha_S^{FB})$, so for sophisticated agents we have $I_S(\alpha^{SB}) + (1-\alpha^{SB})T < I_S(\alpha_S^{FB}) + (1-\alpha_S^{FB})T$. For naïve agents we have that $\alpha^{SB} > \alpha_N^{FB}$, so for naïfs we have $I_S(\alpha^{SB}) + (1-\alpha^{SB})T - \Delta(\alpha^{SB}) < I_S(\alpha_N^{FB}) + (1-\alpha_N^{FB})T - \Delta(\alpha_N^{FB})$, again by our construction of α_N^{FB} in Proposition 1.

Finally, we note that $\lambda W_{\text{base}}(S) + (1-\lambda)W_{\text{base}}(N) = v$ and the firm charges a price strictly higher than marginal cost. On the other hand, we know that $I_S(\alpha^{SB}) + (1-\alpha^{SB})T - (1-\lambda)\Delta(\alpha^{SB})$ is maximized for our choice of α^{SB}, so in particular $I_S(\alpha^{SB}) + (1-\alpha^{SB})T - (1-\lambda)\Delta(\alpha^{SB}) \geq I_S(\tilde{\alpha}) + (1-\tilde{\alpha})T - (1-\lambda)\Delta(\tilde{\alpha}) = v$, where $\tilde{\alpha}$ exists because $I_S(1) < v$ and $I_S(0) + T > v$. The shadow price of the good is cheaper under the second best (it is equal to marginal cost); thus, $\lambda W_{SB}(S) + (1-\lambda)W_{SB}(N) > \lambda W_{\text{base}}(S) + (1-\lambda)W_{\text{base}}(N)$. ∎

Proof of Proposition 14. Consider $m^*(\alpha)$ to be the transfer to the platform as a function of ad load α, while holding all other parameters constant. We know that $\sup_{\alpha \in [0,1]} \partial m^*(\alpha)/\partial\alpha < L$ for some constant L because total demand $(\lambda\beta q + (1-\lambda)\beta(q\bar{\pi}_{F_{N1}(\alpha)} + (1-q)\bar{\pi}_{F_{N0}(\alpha)}) - p)$ has a bounded derivative in α. Second, we see that the platform's potential gain in subscription revenue from the agents of type τ for all $\alpha > \alpha^{SB}$ is given by $(\lambda\mathbb{1}_{\tau=S} + (1-\lambda)\mathbb{1}_{\tau=N})\int_{\alpha^{SB}}^{\alpha}(I_\tau'(x) - T)\,dx = (\lambda\mathbb{1}_{\tau=S} + (1-\lambda)\mathbb{1}_{\tau=N})(I_\tau(\alpha) - I_\tau(\alpha^{SB}) - T(\alpha - \alpha^{SB}))$, which has a derivative bounded above by $I_\tau'(\alpha) < I_N'(0) < \infty$ by Lemma 3.

Thus, taking $\bar{\mu} = I'_N(0) + L$, we see that the platform's marginal revenue from setting $\alpha > \alpha^{SB}$ is upper bounded by $R(\alpha) = m^*(\alpha) - m^*(\alpha^{SB}) + \lambda(I_S(\alpha) - I_S(\alpha^{SB}) - T(\alpha - \alpha^{SB})) + (1-\lambda)(I_N(\alpha) - I_N(\alpha^{SB}) - T(\alpha - \alpha^{SB}))$ which is strictly less than the tax $\mu(\alpha - \alpha^{SB})$ for all $\alpha > \alpha^{SB}$ if $\mu > \bar{\mu}$. This implies that the platform will set an ad load no greater than α^{SB} in any of its plans.

To see that it will never set an ad load strictly less than α^{SB}, note that because $\alpha^{SB} \leq \alpha_S^{FB}$, if $\alpha^* < \alpha^{SB}$ in one of the offered plans, the platform could instead offer $((\alpha^* + \alpha^{SB})/2, P^* + \epsilon)$ for sufficiently small ϵ and both the sophisticates and naïfs would prefer $((\alpha^* + \alpha^{SB})/2, P^* + \epsilon)$ to (α^*, P^*). Moreover, because we know that $m^*(\alpha)$ is monotonically increasing in α, both its advertising and subscription revenue would increase if the platform instead offered $(\alpha^* + \alpha^{SB})/2$ instead of α^* (and it would not be subject to the tax). This represents a profitable deviation; thus, the platform implements exactly one plan with ad load α^{SB}.

Getting the product price down to marginal cost c in equilibrium can be accomplished with a per-unit subsidy δ as follows. Let $Z(\alpha^{SB})$ denote the demand for the good as a function of the advertising load to naïfs (recall by Lemma 1 that aggregate demand of sophisticates is unaffected by their advertising load). Then the firm solves $\max_p (p - c)(Z(\alpha^{SB}) - p)$. Instead let us provide a subsidy to the good in the amount $\delta = Z(\alpha^{SB}) - c$. Then p solves $\arg\max_p (p - c + \delta)(Z(\alpha^{SB}) - p) = \frac{1}{2}(Z(\alpha^{SB}) + c - \delta) = c$, as desired. Here, $Z(\alpha^{SB}) = \lambda\beta q + (1-\lambda)\beta(q\bar{\pi}_{F_{N1}(\alpha^{SB})} + (1-q)\bar{\pi}_{F_{N0}(\alpha^{SB})})$.

Finally, we show that we can implement zero subscription fees with the appropriate per-user subsidy to the platform. Consider if the platform offers one plan at advertising load α^{SB}, and suppose it charges subscription fee $P^* = I_N(\alpha^{SB}) + (1 - \alpha^{SB})T - v$ to maximally extract consumer surplus from the naïve agents (with sophisticates refraining from participation). Then setting $\bar{\eta} = P^*$ means that if $\eta > \bar{\eta}$, the platform can generate the most revenue from each user conditional on advertising at α^{SB} by charging no subscription fee, which implements the second-best plan with $P^{SB} = 0$. ∎

Proof of Proposition 15. The proof consists of three parts. The first part shows that if there is an ad-based plan in the decentralized equilibrium, then we necessarily have an ad-based plan that naïfs opt into with ad load $\alpha^* > \alpha_N^* = \arg\max_\alpha I_N(\alpha) + (1-\alpha)T$. The second part shows that if there is an ad-based plan with $\alpha^* > \alpha_N^*$, a flat digital ad tax with a sufficiently high tax rate can implement a lower ad plan for naïfs at load $(\alpha^* + \alpha_N^*)/2$. The third part argues that the digital ad tax has no impact on the plan offered to sophisticates but improves user welfare for naïfs. Welfare of both agents improve because of the price spillovers to sophisticates from lower advertising to naïfs.

Let us first show that if an ad-based plan is offered in equilibrium to naïfs, then it has ad load $\alpha^* \geq \alpha_N^* = \arg\max_\alpha I_N(\alpha) + (1-\alpha)T$. The claim holds trivially when $\alpha_N^* = 0$, so suppose $\alpha_N^* > 0$ and the platform offers (α_2, P_2) which is selected by the naïfs, with $\alpha_S^{FB} < \alpha_2 < \alpha_N^*$ (it cannot be that $\alpha_2 < \alpha_S^{FB}$, because the platform can extract more subscription fee and ad revenue from both sophisticates and naïfs by offering a plan at least at α_S^{FB}). For the same reasons as in the proof of Proposition 14, because $I'_N(\alpha_2) > T$, the platform could instead offer the plan $((\alpha_2 + \alpha_N^*)/2, P_2 + \epsilon)$ for sufficiently small $\epsilon > 0$ and naïfs would prefer $((\alpha_2 + \alpha_N^*)/2, P_2 + \epsilon)$ to (α_2, P_2), which leads to strictly higher subscription revenue and advertising revenue $m^*(\alpha)$, which is thus a profitable deviation (as sophisticates would remain in their same plan). To show that it is a strict inequality, note that $\partial m^*(\alpha)/\partial\alpha > 0$ for all α, but the loss in subscription revenue from naïfs is equal to $(1-\lambda)(T(\alpha - \alpha_N) - (I_N(\alpha) - I_N(\alpha_N^*))$, with

$\frac{\partial}{\partial\alpha}(1-\lambda)(T(\alpha-\alpha_N)-(I_N(\alpha)-I_N(\alpha_N^*)) = T - I_N'(\alpha)$, which is equal to zero when evaluated at $\alpha = \alpha_N^*$ > 0 (which is true by assumption that an ad-based plan is offered in equilibrium and there is a plan with ad load at least α_N^*). Thus, there exists some small $\nu > 0$ such that $m^*(\alpha_N^* + \nu) - (1-\lambda)(T\nu - (I_N(\alpha_N^* + \nu) - I_N(\alpha_N^*)) > m^*(\alpha_N^*)$, thus the platform can generate greater revenues by setting its ad load to at least $\alpha_N^* + \nu$, establishing the strict inequality. In other words, setting ad load at $\alpha^* = \alpha_N^* + \nu$ leads to a higher sum of subscription fees and ad revenue than setting it at exactly α_N^*, showing the first part.

For the second part, we show that a flat digital ad tax can implement advertising load $(\alpha_N^* + \alpha^*)/2$ in the naïfs' advertising plan. With a flat digital ad tax, the platform will solve $\max_{\alpha \geq \alpha_N^*}(1-\gamma)m^*(\alpha) - (1-\lambda)(T(\alpha-\alpha_N) - (I_N(\alpha) - I_N(\alpha_N^*)))$. Note that $\partial(\beta q \bar{\pi}_{F_{N1}(\alpha)} + \beta(1-q)\bar{\pi}_{F_{N0}(\alpha)})/\partial\alpha$ is bounded for all α, and hence there exists some $L > 0$ such that $\partial m^*/\partial\alpha < L$. But also notice that

$$\frac{\partial}{\partial\alpha}(1-\gamma)m^*(\alpha) - (1-\lambda)(T(\alpha-\alpha_N) - (I_N(\alpha) - I_N(\alpha_N^*))) < (1-\gamma)L - (1-\lambda)(T - I_N'((\alpha^* + \alpha_N^*)/2)$$

which is less than 0 for $\gamma > 1 - \frac{(1-\lambda)(T - I_N'((\alpha^*+\alpha_N^*)/2))}{L} \in (0,1)$. Given this restriction on γ, the platform will always opt to select an advertising load less than or equal to $(\alpha^* + \alpha_N^*)/2$.

For the final part, note the platform can still maximize subscription revenue from the sophisticates by offering their first-best plan α_S^{FB}, which yields the highest subscription fee the platform can extract from them (and recall that, from Lemma 1, there are no digital ad revenues from sophisticates). The lower ad load of $(\alpha^* + \alpha_N^*)/2$ for naïfs leads to better user welfare than the equilibrium without policy, because $\alpha^* > \alpha_N^* > \alpha_S^{FB}$ and the product price p^* is increasing in the ad load of the naïfs' plan. To ensure sophisticates and naïfs participate in their respective plans, the platform then just sets the subscription prices (P_1, P_2) to maximize $\lambda P_1 + (1-\lambda)P_2$ subject to the incentive compatibility constraints from Proposition 7, leaving the same platform surplus for naïfs as before the policy. ∎

A.6 Proofs from Section 7

Proof of Proposition 16. Let $\Pi(\phi_{0,N})$ denote the platform's maximal profit by setting the misspecification parameter to $\phi_{0,N}$. We first show that $\Pi(\phi_{0,N})$ is decreasing in $\phi_{0,N}$. Fix any allocation (α_N, α_S) that the platform wishes to implement, with naïfs choosing planning (α_N, P_N) and sophisticates choosing plan (α_S, P_S) (not necessarily distinct). The set of (P_S, P_N) that implements this allocation is characterized by the four standard constraints from Proposition 7:

$$\begin{aligned}
I_N(\alpha_N;\phi_{0,N}) - P_N &\geq 0, \\
I_S(\alpha_S;\phi_{0,N}) - P_S &\geq 0, \\
I_N(\alpha_N;\phi_{0,N}) - P_N &\geq I_N(\alpha_S;\phi_{0,N}) - P_S, \\
I_S(\alpha_S;\phi_{0,N}) - P_S &\geq I_S(\alpha_N;\phi_{0,N}) - P_N.
\end{aligned}$$

When $\phi_{0,N}$ falls the LHS of the third inequality (the naïf's IC constraint) rises relative to the RHS while the fourth inequality (the sophisticate's IC constraint) remains unchanged. Hence, for the same (α_N, α_S), the feasible subscription price set expands as $\phi_{0,N}$ falls: every (P_S, P_N) that was incentive compatible at $\phi_{0,N}$ remains feasible at any $\phi'_{0,N} < \phi_{0,N}$, and additional (higher-revenue) pairs become feasible because the naïf's IC constraint slackens.

There are two immediate consequences. First, holding (α_N, α_S) fixed, the maximum total subscription revenue $\lambda P_S + (1-\lambda)P_N$ implementable is larger at any $\phi'_{0,N} < \phi_{0,N}$. Intuitively, naïfs can be charged more for the ad plan (their participation constraint can be driven closer to zero while keeping their IC satisfied) and/or sophisticates can be charged more for their plan because the rent they must receive to deter them from mimicking naïfs does not increase. Second, because the first inequality (the naïf's participation constraint) binds in any optimal implementation of the naïf plan, a drop in $\phi_{0,N}$ relaxes that constraint in α_N. Thus, the platform can raise α_N while keeping the participation constraint binding; since ad revenue $m^*(\cdot)$ is increasing in α_N by Lemma 2, this raises ad revenue from naïfs and leads to higher overall platform profits. Thus, we have for any allocation (α_N, α_S) the platform's profit (holding the allocation fixed) is larger at lower $\phi_{0,N}$. Because $\Pi(\phi_{0,N})$ is the pointwise supremum over allocations of these decreasing (in $\phi_{0,N}$) profit functions, $\Pi(\phi_{0,N})$ itself is decreasing in $\phi_{0,N}$. Hence any profit-maximizing platform chooses the smallest feasible misspecification, $\phi^*_{0,N} = \underline{\phi}$.

Finally, note that $\hat{W}_{\phi_0}(\tau)$ is the same as the welfare under the fully-rational benchmark, so the result follows from parts (a) and (b) of Proposition 7. ∎

Proof of Proposition 17. For part (i), have the ordinary differential equation

$$\frac{d\lambda(t)}{dt} = \mu(\tilde{\lambda} - \lambda(t)) + \alpha_N \Lambda(1-q)\phi_0(1-\lambda(t)),$$

so $\frac{d\lambda(t)}{dt} = a - b\lambda$ where $a \equiv \mu\tilde{\lambda} + \alpha_N\Lambda(1-q)\phi_0 > 0$ and $b \equiv \mu + \alpha_N\Lambda(1-q)\phi_0 > 0$. The unique fixed point solves $0 = a - b\lambda^*$, i.e., Equation (3). Uniqueness is immediate because the law of motion is linear in λ, and we even have the closed-form solution $\lambda(t) = \lambda^* + (\lambda(0) - \lambda^*)e^{-bt}$ showing global convergence.

For part (ii), letting $K \equiv \Lambda(1-q)$, note that differentiation yields

$$\frac{\partial \lambda^*}{\partial \alpha_N} = \frac{\mu(1-\tilde{\lambda})\,K\,\phi_0}{(\mu + \alpha_N K\phi_0)^2} \geq 0, \qquad \frac{\partial \lambda^*}{\partial \phi_0} = \frac{\mu(1-\tilde{\lambda})\,\alpha_N K}{(\mu + \alpha_N K\phi_0)^2} \geq 0, \tag{4}$$

and, on any compact parameter set, $\max\{|\partial\lambda^*/\partial\alpha_N|, |\partial\lambda^*/\partial\phi_0|\} = O(1/\mu)$. In particular,

$$\lambda^*(\alpha_N, \phi_0) - \tilde{\lambda} = \frac{\alpha_N K\phi_0}{\mu + \alpha_N K\phi_0}\,(1-\tilde{\lambda}) > 0 \quad \text{whenever } \alpha_N\phi_0 > 0. \tag{5}$$

Let $\Pi_r(\phi_0, \lambda)$ denote the platform's *optimized* profit in mixed regime $r \in \{a, b\}$ of Proposition 7, and define the baseline difference $\Delta_0(\phi_0; \lambda) \equiv \Pi_b(\phi_0, \lambda) - \Pi_a(\phi_0, \lambda)$. Proposition 7 implies: $\partial_{\phi_0}\Delta_0(\phi_0; \lambda) > 0$ (single crossing in ϕ_0), $\partial_\lambda \Delta_0(\phi_0; \lambda) < 0$, and there is a unique cutoff $\tilde{\phi}_0(\lambda)$ with $\Delta_0(\tilde{\phi}_0(\lambda); \lambda) = 0$ and $\tilde{\phi}'_0(\lambda) > 0$.

Endogenize λ by setting $\lambda = \lambda^*(\alpha^r_N(\phi_0), \phi_0)$, where $\alpha^r_N(\phi_0)$ solves regime-r's inner problem. Define the dynamic difference

$$\Delta_{\text{dyn}}(\phi_0) \equiv \Pi_b\big(\phi_0, \lambda^*(\alpha^b_N(\phi_0), \phi_0)\big) - \Pi_a\big(\phi_0, \lambda^*(\alpha^a_N(\phi_0), \phi_0)\big).$$

By the envelope theorem, along each regime's optimizer,

$$\frac{d}{d\phi_0}\Pi_r\big(\phi_0, \lambda^*(\alpha_N^r(\phi_0), \phi_0)\big) = \partial_{\phi_0}\Pi_r(\phi_0, \lambda^*) + \partial_\lambda \Pi_r(\phi_0, \lambda^*)\,\frac{d\lambda^*}{d\phi_0},$$

with $\frac{d\lambda^*}{d\phi_0} = \frac{\partial\lambda^*}{\partial\phi_0} + \frac{\partial\lambda^*}{\partial\alpha_N}\frac{d\alpha_N^r}{d\phi_0}$. The terms $\partial_{\phi_0}\Pi_r$, $\partial_\lambda\Pi_r$ and $d\alpha_N^r/d\phi_0$ are bounded on compacts, while Equation (4) implies $d\lambda^*/d\phi_0 = O(1/\mu)$. Hence there exists $\bar{\mu}$ such that for all $\mu > \bar{\mu}$,

$$\frac{d}{d\phi_0}\Delta_{\text{dyn}}(\phi_0) = \partial_{\phi_0}\Delta_0(\phi_0; \lambda^*) + O(1/\mu) \ > \ 0.$$

Therefore Δ_{dyn} is strictly increasing and admits a unique zero, the dynamic cutoff $\tilde{\phi}_0^{\text{dynamic}}(\phi_1, \phi_{0,N})$, and all regime and comparative static conclusions of Proposition 7 continue to hold when evaluated at $\lambda = \lambda^*(\alpha_N, \phi_0)$.

To locate the dynamic cutoff relative to the baseline, evaluate Δ_{dyn} at $\phi_0 = \tilde{\phi}_0(\tilde{\lambda})$, where $\Delta_0(\tilde{\phi}_0(\tilde{\lambda}); \tilde{\lambda}) = 0$. Using Equation (5) and $\partial_\lambda\Delta_0 < 0$,

$$\Delta_{\text{dyn}}\big(\tilde{\phi}_0(\tilde{\lambda})\big) = \Delta_0\big(\tilde{\phi}_0(\tilde{\lambda}); \lambda^*(\alpha_N^b, \tilde{\phi}_0)\big) - \Delta_0\big(\tilde{\phi}_0(\tilde{\lambda}); \lambda^*(\alpha_N^a, \tilde{\phi}_0)\big) \leq 0,$$

with strict inequality under nondegenerate primitives (e.g., $\alpha_N^r > 0$). Since Δ_{dyn} is strictly increasing in ϕ_0, its zero must lie at a higher value than $\tilde{\phi}_0(\tilde{\lambda})$; thus, $\tilde{\phi}_0^{\text{dynamic}}(\phi_1, \phi_{0,N}) \ \geq \ \tilde{\phi}_0(\tilde{\lambda}, \phi_1, \phi_{0,N})$. This proves that, for $\mu > \bar{\mu}$, Proposition 7 holds identically under the replacement $\lambda \mapsto \lambda^*(\alpha_N, \phi_0)$, and that the cutoff shifts up to $\tilde{\phi}_0^{\text{dynamic}}$. ∎

Proof of Proposition 18. For part (i), write the demand intercepts for sophisticates and naïfs as $A_S \equiv \beta q$ and $A_N(\alpha_N^*) = \beta(q\bar{\pi}_{F_{N1}(\alpha_N^*)} + (1-q)\bar{\pi}_{F_{N0}(\alpha_N^*)})$. By Lemma 2, we know that $A_N(\alpha_N^*) \geq A_S$ for all α_N^* with strict inequality when $\alpha_N^* > 0$. The pooled monopoly price with naïf ad load α_N is $p^* = \frac{c+\lambda A_S + (1-\lambda)A_N(\alpha_N^*)}{2}$ which follows from maximizing $(p-c)(\lambda A_S + (1-\lambda)A_N(\alpha_N^*) - p)$. If the firm conditions on type, each segmented market is linear with intercept A_τ, so the type-specific monopoly prices are $p_S^* = \frac{c+A_S}{2}$ and $p_N^* = \frac{c+A_N(\alpha_N)}{2}$. Since $A_N \geq A_S$, it follows that

$$p_N^* = \frac{c + A_N}{2} \geq p^* = \frac{c + \lambda A_S + (1-\lambda)A_N}{2} \geq p_S^* = \frac{c + A_S}{2}.$$

For part (ii), let the platform offer the mixed menu $M = \{(\alpha_1, P_1), (\alpha_2, P_2)\}$ with $\alpha_1 \leq \alpha_2$. The downstream price part of the platform's objective under a given α_2 is the firm's monopoly profit evaluated at the effective demand intercepts. With type unobserved ($\kappa = 0$), this

$$\pi^*(\bar{A}(\alpha_2)) = \max_p (p-c)(\bar{A}(\alpha_2) - p) = \frac{1}{4}(\bar{A}(\alpha_2) - c)^2$$

where $\bar{A}(\alpha_2) \equiv \lambda A_S + (1-\lambda)A_N(\alpha_2)$. When the firm observe type with probability κ, expected downstream profit at the same α_2 becomes

$$\pi^\kappa(\alpha_2) = (1-\kappa)\pi^*(\bar{A}(\alpha_2)) + \kappa(\lambda\pi^*(A_S) + (1-\lambda)\pi^*(A_N(\alpha_2)))$$

Convexity of $x \mapsto x^2$ implies the uplift from price discrimination at a fixed α_2 is

$$\pi^\kappa(\alpha_2) - \pi^*(\bar{A}(\alpha_2)) = \frac{\kappa}{4}\lambda(1-\lambda)(A_N(\alpha_2) - A_S)^2 \tag{6}$$

which is strictly positive when $A_N(\alpha_2) \neq A_S$. Economically, segmentation raises the margins on naïfs and the platform extracts these via the transfer m^*.

Crucially, the IC and PC constraints that select $(\alpha_1, \alpha_2, P_1, P_2)$ in Proposition 7 do not contain κ: they depend on (α_ℓ, P_ℓ) and I_S, I_N only. Thus, allowing $\kappa > 0$ simply adds the uplift to Equation (6), without changing the feasible set. Now consider the two mixed regimes of Proposition 7: in regime (a), the optimal naïf load satisfies $\alpha_2^{*(a)} < \hat{\alpha}^*$ and in regime (b), the naïf plan sets $\alpha_2^{*(b)} = \hat{\alpha}^*$. Because Equation (6) is strictly increasing in α_2, the gain from introducing $\kappa > 0$ is larger in regime (b) than regime (a), so $(\pi^\kappa - \pi^*)\Big|_{\alpha_2^{*(b)}} > (\pi^\kappa - \pi^*)\Big|_{\alpha_2^{*(a)}}$. Let $\Pi_a(\phi_0; 0)$ and $\Pi_b(\phi_0; 0)$ be the platform profits (including subscription fees and ad rents) at $\kappa = 0$ under the two mixed regimes, evaluated at their respective optimal menus. By Proposition 7, there is a unique cutoff $\tilde{\phi}_0$ at which $\Pi_a(\tilde{\phi}_0; 0) = \Pi_b(\tilde{\phi}_0; 0)$.

Turning on $\kappa > 0$ adds Equation (6) to both objectives. By our previous observation, the increment to regime (b)'s objective strictly exceeds that to regime (a)'s at the same ϕ_0, so $\Pi_b(\tilde{\phi}_0; \kappa) - \Pi_a(\tilde{\phi}_0; \kappa) > 0$. Since the mixed-model profit difference is weakly increasing in ϕ_0 (i.e., higher ϕ_0 raises naïfs' post-ad demand shift and thus ad rents), the equality point between the two mixed profits must occur at a lower false-positive rate: there exists $\tilde{\phi}_0^\kappa < \tilde{\phi}_0$ such that $\Pi_b(\tilde{\phi}_0^\kappa; \kappa) = \Pi_a(\tilde{\phi}_0^\kappa; \kappa)$.

For part (iii), let E^0 be the equilibrium when $\kappa = 0$ with naïf plan $M^0 = (\alpha_N^0, P_N^0)$ and pooled price $p^*(\alpha_N^0)$. Let E^κ be the equilibrium when $\kappa > 0$ with naïf plan $M^\kappa = (\alpha_N^\kappa, P_N^\kappa)$; in that equilibrium a naï faces price $p_N^*(\alpha_N^\kappa)$ with probability κ and $p^*(\alpha_N^\kappa)$ with probability $1-\kappa$. Naïf ex-post welfare at plan (α, P) and product price p is given by $\hat{W}_N(\alpha, P, p) = I_S(\alpha) - \Delta(\alpha) + (1-\alpha)T - P + \frac{1}{2}(\beta q - p)^2$, where $\Delta(\cdot)$ increasing and I_S independent of prices. Observe that (i) for any α, $p_N^*(\alpha) \geq p^*(\alpha)$ (strict if ads are offered to naïfs), and $p^*(\alpha)$ increases in α; (ii) in the mixed model the platform's best response with $\kappa > 0$ sets $\alpha_N^\kappa \geq \alpha_N^0$ (the extractable margin on naïfs rises with κ and α); and (iii) for fixed (α, P), $\hat{W}_N$ is decreasing in the productive price p and for fixed prices, $\hat{W}_N$ is decreasing in α because $I_S(\alpha) - \Delta(\alpha) + (1-\alpha)T$ falls in α for $\alpha \geq \alpha_S^{FB}$. This yields the chain of inequalities $\hat{W}_N' \leq \hat{W}_N(M^\kappa; 0) \leq \hat{W}_N(M^0; 0) = \hat{W}_N$, where the first inequality uses only the price ranking $p_N^*(\alpha_N^\kappa) \geq p^*(\alpha_N^\kappa)$ and that $\frac{1}{2}(\beta q - p)^2$ is decreasing in p, and the second inequality follows from $\alpha_N^\kappa \geq \alpha_N^0$ together with monotonicity of the non-price part in α and the fact that even the pooled price $p^*(\alpha)$ rises with α. Thus, $\hat{W}_N' \leq \hat{W}_N$, with a strict inequality whenever either $p_N^*(\alpha_N^\kappa) > p^*(\alpha_N^\kappa)$ (nondegenerate price discrimination) or $\alpha_N^\kappa > \alpha_N^0$ (the platform indeed pushes more ads on naïfs). ∎

Proof of Proposition 19. Consider an arbitrary plan (α, P). The (interim) utility for type $(\tau, \nu) \in \{S, N\} \times \{\nu_L, \nu_H\}$ is $\tilde{U}_{\tau,\nu}(\alpha, P) = I_\tau(\alpha) + (1-\alpha)\nu T - P$. For any two intended types a, b with assigned plans (α_a, P_a) and (α_b, P_b), their IC constraints are $\tilde{U}_a(\alpha_a, P_a) \geq \tilde{U}_a(\alpha_b, P_b)$ and $\tilde{U}_b(\alpha_b, P_b) \geq \tilde{U}_b(\alpha_a, P_a)$. Subtracting these two inequalities cancels subscription prices and yields the key monotonicity condition that $\Delta_{a,b}(\alpha_a) - \Delta_{a,b}(\alpha_b) \geq 0$ where $\Delta_{a,b}(\alpha) \equiv \tilde{U}_a(\alpha, 0) - \tilde{U}_b(\alpha, 0)$. Hence if $\Delta_{a,b}(\cdot)$ is (weakly) increasing, IC forces $\alpha_a \geq \alpha_b$; if decreasing, IC forces $\alpha_a \leq \alpha_b$.

Apply this to part (i) by fixing τ and taking $a = (\tau, \nu_H)$ and $b = (\tau, \nu_L)$. Here $\Delta_{a,b}(\alpha) = (\nu_H - \nu_L)(1-$

$\alpha)T$ is decreasing in α, so IC implies that $\alpha_{\tau,H} \leq \alpha_{\tau,L}$. The price ordering then following by writing the two IC inequalities explicitly subtracting the I_τ-terms:

$$P_{\tau,H} - P_{\tau,L} \geq (\nu_H - \nu_L)T(\alpha_{\tau,L} - \alpha_{\tau,H}) \geq 0,$$

so $P_{\tau,H} \geq P_{\tau,L}$.

For part (ii), fix ν and take $a = (N, \nu)$ and $b = (S, \nu)$. Now $\Delta_{a,b}(\alpha) = I_N(\alpha) - I_S(\alpha)$ which is increasing in α; the same monotonicity condition therefore gives $\alpha_{N,\nu} \geq \alpha_{S,\nu}$: naïfs receive weakly more ads than sophisticates at any given ν. ∎

B Online Appendix: Kinked Linear Demand

In the main text we assumed that β was sufficiently large that $z_i < 0$, so that we could drop the constraint $z_i \geq 0$. Here, we relax this and suppose β can take any value, and users are constrained to only non-negative consumption. In particular, agents now solve $\max_{z_i \geq 0} \mathbb{E}^{\tau_i} [U(z_i; \theta_i) - pz_i]$, which yields a kinked demand curve $z_i^* = (\beta\pi_i - p)_+$, instead of our previously linear demand curve $z_i^* = \beta\pi_i - p$. Our results are impacted as follows.

First, it is possible to extract some surplus from sophisticated agents by advertising to them when demand is kinked. This follows from Jensen's inequality, since $\mathbb{E}^S[(\beta\pi_i - p)_+] > (\beta\mathbb{E}^S[\pi_i] - p)_+ = (\beta q - p)_+$. Thus, in the baseline model, the platform is no longer completely indifferent between sophisticated agents' participation and not, as advertising to them leads to profits $\Pi_S(\alpha)$ which are increasing in α for the same reasons as Lemma 2. By Lemma 3, naïfs will always participate on the ad-based plan whenever sophisticates do, but not necessarily vice-versa. When an ad-based platform decides on the advertising load, it trades off a lower ad load $\hat{\alpha}' < \hat{\alpha}^*$ that keeps sophisticates on the platform (capturing informational surplus from a greater fraction of the population) with the higher ad load $\hat{\alpha}^*$ that sacrifices Π_S but can extract more surplus from naïfs at load $\hat{\alpha}^*$. This implies that our Proposition 2 now turns on both λ and ϕ_0. Fixing ϕ_0 (and other model primitives), there exists λ^* such that if $\lambda < \lambda^*$, the platform chooses ad load $\hat{\alpha}^*$ with $I_N(\hat{\alpha}^*) + (1 - \hat{\alpha}^*) - v = 0$ just as in the baseline model under Lemma 4. However, when $\lambda > \lambda^*$, there are sufficiently many sophisticates and the platform prefers to retain their participation, instead choosing $\hat{\alpha}'$ where $I_S(\hat{\alpha}') + (1 - \hat{\alpha}')T - v = 0$.

Within each of these regimes, the dependence on ϕ_0 is also slightly different. For sufficiently small values of λ, we recover exactly the cutoff structure of Proposition 2, where a higher rate of false positives ϕ_0 results in the platform adopting an ad-based model with ad load $\hat{\alpha}^*$. On the other hand, when λ is sufficiently close to 1, we end up with a flipped cutoff structure: There exists $\hat{\hat{\phi}}_0$ such that the ad-based business model is adopted if and only if $\phi_0 < \hat{\hat{\phi}}_0$. The reason here is that Blackwell informativeness is one-to-one with Π_S, so the platform can extract more informational rent if the ads themselves are more informative, which happens when the gap between ϕ_0 and ϕ_1 is larger. Regardless of which regime we are in, we still have our main result from Proposition 4: The platform will leave the sophisticated agents with no excess surplus from platform usage, and product prices will be higher under an ad-based business model. The naïfs' welfare is always below that of the sophisticates, so in the presence of advertising, the welfare of both types will fall below base case levels.

Most of these insights generalize immediately to mixed platform business models, competition, and the policy analysis. When the platform can offer a menu that segments sophisticates and naïfs, there will be a pair of cutoffs (ϕ_0^*, ϕ_0^{**}) where $\phi_0 > \phi_0^*$ leads to an advertising plan for naïve agents (as in Proposition 5) and where $\phi_0 < \phi_0^*$ leads to an advertising plan for sophisticated agents (as opposed to always offering a subscription-based plan to sophisticates). Because competitive forces push the firms and platforms to cater more to users (rather than extract full surplus), both Proposition 9 and 11 apply identically. Finally, our policy implications remain fully intact: The first and second-best are allocations, so are unaffected by the surplus the platform can extract from sophisticates, and the flat digital ad tax always helps improve welfare in the decentralized equilibrium.

C Online Appendix: Digital Ad Taxation with Multiple Firms/Platforms

We consider how our digital ad taxation policy changes in the case of multiple firms (N) and platforms (M). When there are multiple firms, there will generally be a menu of advertising vectors $(\alpha^{(1)}, \ldots, \alpha^{(N)})$ instead of just a menu of advertising intensities α. The informational value from advertising will generally take the form of $I_S(\alpha^{(1)}, \ldots, \alpha^{(N)}) = \sum_{j \in \{0,1\}^N} \prod_{k=1}^{N} (e^{-\alpha^{(j)}T}(1-j_k) + (1-e^{-\alpha^{(j)}T})j_k)\chi_S^{(j)}$ (for some $\chi_S^{(j)}$ that depend on model primitives) and analogously for $I_N(\alpha^{(1)}, \ldots, \alpha^{(N)})$. To solve for the first-best level of advertising, we maximize $I_S(\alpha^{(1)}, \ldots, \alpha^{(N)}) + (1 - \sum_{j=1}^{N} \alpha^{(j)}T)$ for sophisticates and maximize $I_S(\alpha^{(1)}, \ldots, \alpha^{(N)}) + (1 - \sum_{j=1}^{N} \alpha^{(j)}T) - \Delta(\alpha^{(1)}, \ldots, \alpha^{(N)})$ for naïfs, for an appropriately defined Δ under the multi-advertiser case, just as in the proof of Proposition 1. It is easy to see from there that there exists a unique α_S^{FB} such that $(\alpha_S^{FB}, \ldots, \alpha_S^{FB})$ is first-best for sophisticates, and a corresponding unique $\alpha_N^{FB} \leq \alpha_S^{FB}$ such that $(\alpha_N^{FB}, \ldots, \alpha_N^{FB})$ is first-best for naïve agents. Once again, IC constraints from the users will prevent us from implementing the first-best level of advertising, and there will be a tension that makes it impossible to separate sophisticates and naïfs in the second-best, leading to the offering of a single plan which takes the form of $(\alpha^{SB}, \ldots, \alpha^{SB})$ with $\alpha_N^{FB} \leq \alpha^{SB} \leq \alpha_S^{FB}$. Note that because users participate in its most one plan, our first-best and second-best allocations are exactly the same in the presence of multiple firms and platforms.

To implement the second best, we consider the case of a single platform with multiple firms and multiple platforms and potentially multiple firms separately. For a single platform, the planner can similarly levy a sufficiently high advertising tax on ad quantities above α^{SB} for any individual advertiser. Note that in general it will not be sufficient to regulate total advertising, because the platform may not play a symmetric advertising strategy, so taxing the sum of advertising above $N\alpha^{SB}$ may not implement the second best. However, a policy of the form of Proposition 14 with the tax applying to each individual advertiser will have the same desired effect as in the single advertiser case. With multiple platforms, our analysis from Proposition 11 extends here to show that the platforms will compete to offer the plans that are viewed as most desirable to sophisticated and naïve agents. In other words, the platforms will offer one plan for sophisticates with no subscription fee that has advertising at $(\alpha_S^{FB}, \ldots, \alpha_S^{FB})$ and will offer a second plan to naïfs that has advertising load $(\alpha_N^*, \ldots, \alpha_N^*)$ with $\alpha_N^* > \alpha^{SB}$ and no such option fee. This maximizes naïfs' utility under their subjective model. In such a setting, a similar policy as in the single platform case will implement the second best, but the tax rate may need to be even higher. This is because competition drives platforms not to offer the most profitable business models, but the ones that are seen as utility-maximizing by the agents. Thus, the tax must be high enough that platforms who advertise above $(\alpha^{SB}, \ldots, \alpha^{SB})$ in fact operate at loss, as opposed to just making less profit than when $(\alpha^{SB}, \ldots, \alpha^{SB})$ is played.