WORKING PAPERS

N° 1171

April 2022

"Data, Targeted Advertising, and Quality of Journalism: The Case of Accelerated Mobile Pages (AMP)"

Doh-Shin Jeon and Jun Yan

Toulouse School of Economics

Data, Targeted Advertising, and Quality of Journalism: The Case of Accelerated Mobile Pages (AMP) *

Doh-Shin Jeon † Jun Yan‡

April 21, 2022

Abstract

This paper studies how newspapers' adoption of Google's Accelerated Mobile Pages (AMP), which is a publishing format that enables instant loading of web pages in mobile browsers, changes data allocation and thereby newspapers' incentive to invest in quality journalism when consumer data are used for targeted advertising. The adoption of AMP allows Google to obtain consumer data from AMP articles and to combine it with other sources of consumer data to improve the targeting of the advertisements served by Google on other websites. Even if such data combination increases static efficiency, it can reduce dynamic efficiency when it lowers the ad revenue per newspaper traffic, thereby reducing the quality of journalism. Newspapers face a collective action problem as a newspaper's adoption of AMP generates negative externalities to other newspapers through search ranking and data leakage. By leveraging its market power in search and ad intermediation, Google can induce newspapers to adopt AMP. We provide policy remedies which eliminate the externalities. Our remedies are relevant to current regulatory interventions to make Google pay for displaying newspapers' content.

Keywords: Targeted Advertising, Data Combination, Data Leakage, Quality of Journalism, Search Engine

JEL: D21, L12, L15, L82, M37

*We thank Marc Bourreau, Jacques Crémer, Alexandre de Cornière, Anna D'Annunzio, Antoine Dubus, Luise Eisfeld, Alipio Ferreira, Ulrich Hege, Christian Hellwig, Jan Krämer, Yassine Lefouili, Filip Mrowiec and participants in the 2021 MaCCI conference, 12th Paris Conference on Digital Economics, 4th Doctoral Workshop on The Economics of Digitization, 48th Annual Conference of EARIE and internal workshops at TSE for valuable discussions and comments. Both authors acknowledge funding from ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program).

†Toulouse School of Economics, University of Toulouse Capitole, CEPR, dohshin.jeon@tse-fr.eu.

‡Toulouse School of Economics, University of Toulouse Capitole, jun.yan@tse-fr.eu.

1 Introduction

There is a general tendency by major gatekeeper platforms to use their power to channel consumer interactions with business users into their walled gardens. The development of Super Apps in China by the two major Chinese platforms (Alibaba and Tencent) is an extreme example of such a tendency: within each Super App, a consumer can carry out almost all her activities, including shopping, ride hailing, reading news, gaming, money transfer, and flight bookings, such that she barely needs to leave the Super App. Inspired by the Chinese Super Apps, Facebook and Uber have adopted a similar business strategy. Another example is Google's tendency to "swallow web", about which Shira Ovide, who writes the On Tech newsletter of the New York Times, expresses her concern as follows:

"One longstanding issue is Google's evolution from a website that pointed people to the best links online to one that's swallowing the web. ... Now, Google is more likely to prominently show information or advertisements from its own computer systems or scraped from other companies' websites — and keep you within Google's digital walls. Google isn't a front door to the internet anymore. It's the house. (New York Times, On Tech newsletter, September 24, 2020)"

A main reason for which major platforms expand their walled gardens instead of embracing an open Internet is that they want to collect as much data as possible about consumers' various online activities, which allows them to infer consumers' preferences and to predict their behaviors. This motive is particularly relevant to ad-financed platforms such as Google and Facebook, whose business model consists in harvesting consumer attention and data and monetizing them through targeted advertising (Zuboff, 2019). Platforms' access to business users' data raises a very important question: how does such data access affect the innovation of the platform ecosystem, in particular, the innovation incentives of business users?[1]

We explore this question in the specific context of newspapers' adoption of Google's Accelerated Mobile Pages (henceforth AMP), which is an open-source publishing format developed by Google to enable instant loading of web pages in mobile browsers. Both the CMA Report (2020) of the UK and the report of U.S. House of Representatives (2020) have expressed concerns about Google's anti-competitive practices in its implementation of AMP, especially about the so-called "data leakage" issue[2]. Specifically, Google hosts articles written in AMP format on its servers, thereby collecting consumers' browsing data on these articles. Then, Google can use this data for targeting ads to newspapers' readers on other websites, undermining the value of newspapers' ad inventories.

To provide a perspective on the importance of advertising revenue in the newspaper

[1]For instance, the question arises regarding Amazon's use of business data of third-party sellers who sell on its marketplace, which is under investigation by the European Commission.

[2]There is another issue related to AMP, which we briefly describe in Section 8. For additional information regarding the AMP issues, see also Appendix S of the CMA Report (2020) (p.3 and p.17), Geradin and Katsifis (2019), Scott Morton and Dinielli (2020) and Srinivasan (2020)

industry, we point out two worrying trends. According to Pew Research Center (2020),[3] the industry's advertising revenue fell sharply from 49 billion dollars in 2006 to 8 billion dollars in 2020, and as a consequence, the total number of newsroom employees declined from 74,410 in 2006 to 30,820 in 2020. During the same period, the newspaper industry has made a transition to the online world and has become increasingly dependent on Google, which monopolizes search and ad intermediation.[4] As quality journalism not only matters for consumer surplus but is also a pillar of democratic societies, it is vital to understand how Google's exercise of its market power influences the news industry and its implications for social welfare (Rolnik, Cagé, Gans, Goodman, Knight, Prat, Schiffrin, and Raj, 2019; OECD, 2021).

Against this background, we address the following questions regarding Google's AMP. How does newspapers' adoption of the AMP format change data allocation and thereby newspapers' incentives to invest in quality journalism? What is its impact on static and dynamic welfare? Does Google have any incentive to internalize the impact on the quality of journalism? How does Google leverage its market power in search and ad intermediation to induce newspapers to adopt AMP? What are policy remedies?

To answer these questions, we build a model that captures the online environment in which newspapers operate, involving consumers, a monopolistic search engine, competing providers of ad inventory, ad intermediaries, and advertisers. In particular, our model incorporates some main features of the open display advertising market in which the majority of publishers sell their display advertising inventory to a large number of advertisers through real-time auctions run by ad intermediaries. The ad intermediation market is dominated by Google.[5] We consider competition between Google and another ad tech intermediary T in this market.

In our model, consumers and advertisers interact in two different two-sided markets: the newspaper market and another market (called sector B) which is comprised of non-newspaper content providers. In other words, both consumers and advertisers multihome on these two markets. On the consumer side, newspapers compete among themselves for readership by choosing the quality of journalism. Each consumer chooses a single newspaper whose site she directly visits to read news. In addition, consumers also search for news by using the monopolistic search engine (SE) and read news from multiple newspapers depending on the search results. On the advertiser side, each newspaper faces competition from content providers in sector B as each consumer multihomes on her favorite newspaper and her favorite content provider in sector B. We focus on strategic interactions between the search engine and newspapers.

The ad tech intermediaries collect consumer data for targeted advertising. The data al-

[3]See Newspapers Fact Sheet at https://www.pewresearch.org/journalism/fact-sheet/newspapers/

[4]According to CMA Report (2020), Google accounts for 25% of mobile website traffic for large publishers in the UK, is dominant in different layers of ad intermediation, and in particular has more than 90% share in the publisher ad server market.

[5]See the CMA Report (2020) and Jeon (2021) for detailed analysis of the ad intermediation market.

location between SE and T is determined by the set of content providers each intermediary serves and their respective ability to track consumer activities. In the baseline model introduced in Section 2, we assume that all newspapers use T, which only collects consumer data from direct visit due to its imperfect tracking technology, while the SE serves the sector B and has perfect tracking technology. We assume the ad revenue generated from a consumer by an ad tech intermediary increases with the amount of data it has about the consumer but decreases with the extent of overlap with the data that the rival intermediary has about the consumer.

In this environment, AMP impacts newspapers through three different channels. First, AMP adoption generates the benefit of eliminating the loss of search referral traffic due to the slow loading of pages. Second, a newspaper's adoption decision affects the amount of its search-referral traffic as the SE promotes adopters' rankings and demotes non-adopters' rankings in search results. Third, because the SE hosts AMP articles and thereby collects browsing data on these pages, adopting AMP changes data allocation between the two ad intermediaries and thereby affects newspapers' ad revenues.

Section 3-5 study how AMP impacts news quality and social welfare. In Section 3, we study two benchmarks and show that if AMP does not involve any change in data allocation, it induces newspapers to invest more in the quality of journalism. In Section 4, we explain how AMP changes data allocation and perform a static analysis. We find that when newspapers' quality levels remain unchanged, AMP increases static welfare as the SE's access to AMP data improves ad targeting in sector B and thereby increases total surplus in the advertising market. Then, we endogenize newspapers' quality choices in Section 5 and find that the adoption of AMP leads to two opposite effects for newspapers: (i) the search traffic enhancing effect due to the fast-loading of articles, which is positive; and (ii) the data leakage effect, which reduces the ad revenue per direct traffic and is negative. As a result, when the data leakage effect dominates the search traffic enhancing effect, news quality is reduced, potentially leading to lower dynamic welfare. However, we find that the SE has no incentive to internalize the impact of its data combination on the quality of journalism.

In Section 6, we analyze newspapers' incentives to adopt AMP. Newspapers face a collective action problem as a newspaper's adoption of AMP generates two kinds of negative externalities to other newspapers—search ranking externality and data leakage externality. The first externality arises as the SE promotes adopters' articles in search results. Therefore, a newspaper's adoption will have negative impacts on other newspapers' search referral traffic. Second, a newspaper's adoption leaks to the SE data about other newspapers' direct readers when they are referred to its AMP articles by the SE. This reduces rival newspapers' advertising revenues from direct traffic. As an individual newspaper does not internalize these negative externalities on its competitors, we find that there always exists an equilibrium in which all newspapers adopt AMP. However, another equilibrium in which no newspaper adopts AMP can also exist when the loss in ad revenue from "data leakage" is strong enough.

In Section 7, we further examine the collective action problem by considering an extension where a fraction of newspapers is assumed to use the SE as their ad intermediary. We show that this creates a conflict of interest regarding data leakage between two groups of newspapers depending on whether they use ad tech T or ad tech SE, which Google can exploit to make the adoption equilibrium unique. This result implies that the SE can combine its market power in search and in ad intermediation through a divide-and-conquer strategy to gain control of newspapers' data.

In Section 8, we propose policy remedies which solve the collective action problem by eliminating the two sources of externalities and connect the remedies to the current regulatory interventions in the bargaining between Google and newspapers regarding the compensation Google should make for displaying newspapers' content.

In Section 9 we conclude. Appendix B contains omitted proofs. Readers may refer to Appendix C for an introduction to the open display advertising market, which provides stylized facts that guide our modeling choices.

1.1 Related literature

There is an emerging body of literature studying various data-driven (anti-)competitive strategies and exploring the implication of data combination (also known as data tying or data pooling) on competition(Ghosh, Mahdian, McAfee, and Vassilvitskii, 2015; Condorelli and Padilla, 2020; de Cornière and Taylor, 2020; Bourreau, Hofmann, and Krämer, 2021). Our paper is more related to the papers studying online advertising. Ghosh et al. (2015) is an early paper studying the data leakage issue and explores conditions under which data-sharing enabled by cookie matching can improve one publisher's revenue while harming that of another. Bourreau et al. (2021) is more closely related to our paper, as they also use the AMP issue as one of the motivations and study a data-prominence trade-off faced by publishers. Namely, They consider a game where the dominant platform uses exclusive contract to offer prominent positions in exchange for publisher's data and are interested in its implication on data collection and ranking bias. We have very different focuses and mechanisms as we model the adoption of AMP as a voluntary opt-in game and study its consequence on newspapers' incentives to invest in news quality.

In addition, as we are concerned about a gatekeeper platform's access to business users' data, we contribute to the general discussion about regulation of data collection and usage by gatekeeper platforms. Existing literature explores different types of data-related issues than ours. For instance, Madsen and Vellodi (2021) considers whether vertically integrated marketplaces should be banned from using proprietary sales data of third-party sellers to develop competing product; Johnson, Jungbauer, and Preuss (2021) examines how policies for the ownership and control of consumer browsing data affect market outcomes in the online advertising industry.

As our objective is to show how Google, as a gatekeeper platform, affects the quality of journalism, our paper is related to numerous papers on news aggregators such as Dellarocas, Katona, and Rand (2013); Jeon and Nasr (2016); de Cornière and Sarvary (2022).[6] These papers examine different mechanisms through which a news aggregator (or a large digital platform like Facebook) affects competition among newspapers by influencing news sites' traffic, while taking the advertising revenue per traffic as given. By contrast, we account for Google's influence on newspapers' advertising business by making the ad revenue per traffic endogenous to data allocation, which in turn is influenced by Google's power in search market. In addition, our paper is related to the empirical paper of Calzada and Gil (2020), which examines the impact of Google News' opt-in policy on news publishers' traffic in Germany. A publisher's choice to opt out means shorter excerpts and no image for its articles indexed by Google News. Therefore, this policy generates a collective action problem similar to the one generated by search ranking externality in our model. They find that opting out reduced by 8% the number of visits to the outlets controlled by Axel Springer.

The two-sided market feature of our model makes our paper related to the canonical literature on media competition in two-sided markets (Anderson and Coate, 2005; Ambrus, Calvano, and Reisinger, 2016; Athey, Calvano, and Gans, 2018; Anderson, Foros, and Kind, 2018; Anderson and Peitz, 2020) [7] and to its recent development accounting for the role of ad intermediaries (D'Annunzio and Russo, 2020, 2021).[8] We extend the stylized two-sided media market setup, where media firms compete both on the consumer side and on the advertising side, by introducing a gatekeeper platform that can act as an intermediary on both sides. We show that the gatekeeper platform can leverage its market power from the consumer side (i.e. the search market) to the advertising side for its benefit. We explicitly model consumers' generic multi-homing on different services to capture a main consequence of programmatic advertising: the boundary of the advertising market is much larger than that of each product (i.e. content) market.[9] Although the model of Krämer, Schnurr, and Wohlfarth (2019) also captures this situation, our paper is distinguished from theirs for the following reasons. First, they perform a static analysis, whereas we focus on dynamic efficiency in terms of quality choices. Second, they consider a setup of a representative advertiser in which each publisher chooses an ad price, while we consider the

[6]See Jeon (2018) for a survey of the literature on news aggregators.

[7]This literature considers an environment in which publishers directly contract with advertisers to sell ad inventories. Anderson and Coate (2005) is the seminal paper that explicitly accounts for the cross-group externalities between consumers and advertisers in studying advertising market. Relaxing Anderson and Coate (2005)'s assumption that consumers are single-homing, Ambrus et al. (2016), Athey et al. (2018), and Anderson et al. (2018) examine how it affects the advertising market outcomes when a subset of consumers multi-home. Because all of these papers assume that publishers cannot perfectly track consumers, they focus on how repetitive impression affects ad price. By contrast, our paper assumes that all consumers generically multi-home on non-competing content providers and that Google has perfect tracking technology. Our model highlights the role consumer data plays in online targeted advertising.

[8]These two papers examine how the presence of different ad intermediaries affects media competition for advertisers. D'Annunzio and Russo (2020) explores the role of ad networks that track consumers across websites to cap the frequency of impressions. D'Annunzio and Russo (2021) considers ad intermediaries that use consumer browsing data on publishers' websites for targeted advertising and frequency capping. Both papers endogenize publishers' decision of outsourcing ad inventories to ad intermediaries.

[9]A common assumption in these papers is that the boundary of the product (i.e. content) market coincides with that of the advertising market, which means that the same set of media that compete for readership also compete for advertising dollars.

programmatic sale of display advertising. Last, they consider data sharing that improves ad targeting of both parties sharing the data; by contrast, we consider data leakage from newspapers to the SE, which is facilitated by the exercise of its market power in search and ad intermediation.

2 Baseline Model

In this section, we present the baseline model in which a group of consumers and a group of advertisers interact on two different two-sided markets—the newspaper market and sector B to be explained. On the consumer side, a monopolistic search engine (SE) mediates consumers' search for news. On the advertiser side, the SE acts as an ad tech intermediary and competes with another ad tech intermediary T to sell content providers' ad inventory to advertisers.

In our analysis, we focus on the strategic interaction between the SE and n number of ad-funded newspapers that compete on quality, although we also consider quality choice by content providers in sector B. The other market participants—consumers, advertisers, and ad tech intermediary T—are not strategic players. Figure 1 depicts the model's industry structure, which we explain below.

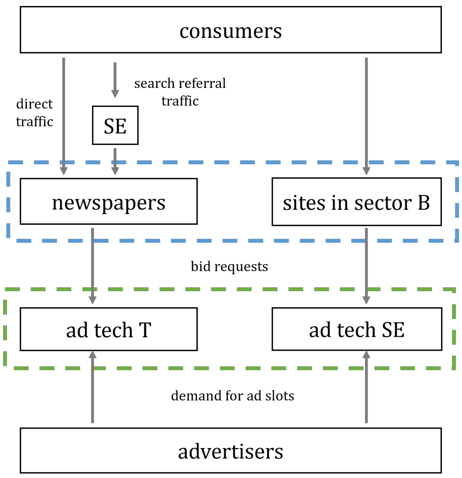

Figure 1: Industry Structure

Notes. The blue rectangle represents the content market in which newspapers compete in quality for both direct traffic and search referral traffic, but they do not compete with sites in sector B for traffic. The green rectangle represents the ad intermediation market where two ad intermediaries (SE and T) sell ad inventory to advertisers.

2.1 Newspaper market and Sector B

There are n online newspapers that are purely financed by advertising. Each newspaper i competes by investing in quality q_i at a cost $c(q_i)$, which is strictly increasing and strictly convex with $c(0) = 0$. Let $\mathbf{q} \equiv (q_1, ..., q_n)$ and $\mathbf{q}_{-i} \equiv (q_1, ..., q_{i-1}, q_{i+1}, ..., q_n)$.

The newspapers are horizontally differentiated, and their demands are determined by the quality vector $\mathbf{q}$. Each newspaper i's demand is composed of two sources—traffic from direct visit (direct traffic) $D^{d,i}(\mathbf{q}) = D^{d,i}(q_i, \mathbf{q}_{-i})$ and traffic referred by the monopolistic search engine (search referral traffic) $D^{s,i}(\mathbf{q}) = D^{s,i}(q_i, \mathbf{q}_{-i})$, $i \in \{1, 2, ..., n\}$. Regarding how these two traffic sources are formed, we have in mind a situation in which consumers have heterogeneous tastes for newspapers and each consumer has a single preferred newspaper whose site she visits directly and regularly. In addition, consumers also search for news by using the search engine (SE) and read news from multiple newspapers depending on the search results.

We make the following assumption about newspaper demand.

Assumption A1 *(i)* $D_i^{d,i} = \frac{\partial D^{d,i}}{\partial q_i} > 0,\ D_j^{d,i} = \frac{\partial D^{d,i}}{\partial q_j} < 0$ *for* $j \neq i$ *and the same for* $D^{s,i}(\mathbf{q})$*;*

(ii) $\sum_{j=1}^{n} D_j^{d,i} \geq 0$ *when* $q = q_1 = ... = q_n$ *and the same for* $D^{s,i}(\mathbf{q})$*;*

(iii) $D_{ii}^{d,i} = \frac{\partial^2 D^{d,i}}{\partial {q_i}^2} \leq 0,\ D_{ij}^{d,i} = \frac{\partial^2 D^{d,i}}{\partial q_j \partial q_i} \leq 0, j \neq i$ *and the same for* $D^{s,i}(\mathbf{q})$*.*

A1 (i) means that newspaper i's demand is increasing in its own quality q_i, while decreasing in any competitor j's quality q_j. A1 (ii) means that at symmetric quality, if all newspapers increase their quality, it at least weakly increases an individual newspaper's demand. This is the market expansion effect. In A1 (iii), $D_{ii}^{d,i} \leq 0$ guarantees that newspapers' profit maximization problems are concave; the property of cross-derivative $D_{ij}^{d,i} \leq 0, j \neq i$ implies that quality choices are strategic substitutes.

Regarding search referral traffic, it is convenient to think that the demand is determined by two stages. In the first stage, consumers visit the SE to search for news and, given the search result, decide which links to click through. The demand determined at this stage is captured by $D^{s,i}(\mathbf{q})$. In the second stage, the SE directs consumers to the news websites whose links are clicked. In this process, consumers may suffer from the slow loading of pages such that they decide not to read the news. We assume that the loss rate is $\delta \in [0, 1]$. As a result, the final search referral traffic is $(1-\delta)D^{s,i}(\mathbf{q})$.

In addition to news consumption activity (considered activity A), we assume that each consumer also uses the Internet to visit other ad-financed content providers' applications or websites. For simplicity, we aggregate all other websites different from newspapers into a single sector called B. We make a reduced-form representation of sector B and assume that the demand of sector B, $D^B(q_B)$, is determined by its quality choice q_B as well, and $\frac{dD^B}{dq_B} > 0,\ \frac{d^2 D^B}{d{q_B}^2} < 0$. The cost of investing in quality is $c_B(q_B)$, which is strictly increasing and convex with $c_B(0) = 0$. The newspaper market and sector B constitute the content market, and we call players in these markets content providers.

2.2 Advertising market and data

Both newspapers and sites (or apps) in sector B are ad-financed and hence sell targeted ads. Every consumer multi-homes on her preferred news site and sector B. Although the sites in sector B do not compete with newspapers on the content side, consumers' multi-homing implies that they do compete with the latter on the advertiser side by offering ad inventory that targets the same consumers.

The sale of inventory for targeted advertising is mediated by ad tech intermediaries that organize real time auctions on behalf of advertisers and publishers. We consider a duopolistic ad tech market—an ad tech system operated by the SE and an alternative system T based on third-party ad intermediaries. We assume that the sites/apps in sector B use the ad tech service of the SE. Actually, we obtain qualitatively the same results if we assume that sector B represents the products of the SE whose advertising are sold by the SE.[10] In the baseline model, all newspapers are assumed to use the service of the ad tech intermediary T. We later extend the baseline model to a situation where an exogenous number of newspapers use the ad tech service of the SE.[11]

To describe targeted advertising revenues generated in the two ad tech systems, we start by describing data allocation. Let Ω^x be the complete data set generated by all of consumer x's online activities. In our model, we have $\Omega^x = \omega^{x,d} \cup \omega^{x,s} \cup \omega^{x,B}$, where $\omega^{x,k}$ represents the set of browsing data generated by consumer x's activity $k \in \{d, s, B\}$, that d stands for direct news reading, s search-referred news reading, and B activities in sector B. Then let $\Omega^{x,h} \subseteq \Omega^x$ be the set of data that ad intermediary $h \in \{SE, T\}$ has about consumer x. An ad tech system can collect and combine data on a consumer's various activities conducted on its customers' websites/apps across time only if it can identify the consumer. We assume that ad intermediary T can perfectly identify consumers in direct traffic but cannot identify consumers in search traffic. By contrast, the SE's tracking technology is superior such that it can combine any data sets it has access to. This assumption of asymmetric tracking technology between ad tech T and SE has a realistic foundation.[12] Together with the assumption on content providers' choices of ad tech intermediaries, this assumption on tracking ability implies that, in the benchmark without AMP, data allocation associated with a certain consumer x is $\Omega^{x,T} = \omega^{x,d}, \Omega^{x,SE} = \omega^{x,B}$.

[10]Google provides more than 53 consumer-facing services and products in the UK and gathers data through them (Appendix F of CMA Report (2020), p. F8)

[11]Studying a full-fledged competition between ad tech intermediaries in which they make offers to content providers is beyond the scope of this paper. To our knowledge, no paper provides yet a formal analysis of the competition in the ad tech market.

[12]Ad techs rely on cookies as identifiers when tracking consumers' activities in mobile browsers while using mobile advertising IDs (MAID) as identifiers for tracking in mobile apps. Due to technical limits, small ad techs may fail to match these two IDs from time to time, leading to the loss of consumer data. By contrast, due to the popularity of its consumer-facing services and Android operating system, Google could use first-party login to facilitate its matching (see detailed explanation in Appendix C). So our model could be interpreted as consumers using apps to read news from their single preferred news outlet while using mobile browsers to search for news from multiple newspapers, and ad tech T is unable to match consumers' cookie IDs with their MAIDs.

Next, we use a reduced-form approach to let data allocation determine targeted advertising revenues, abstracting away from the details of real-time auction. When a content provider uses ad intermediary h, the overall advertising revenue generated by a unit of traffic associated with activity k is

$$\alpha_{k,h}(\Omega^{x,h}, \Omega^{x,h} \cap \Omega^{x,-h})$$

where $k \in \{d, s, B\}, h \in \{SE, T\}$ and $-h$ represents the rival ad tech system. Namely, we characterize the competition between the two ad tech systems by letting the advertising revenue in an ad tech system depend not only on the amount of data it owns but also on the extent of overlap between the two ad tech systems' data sets.

We impose the following assumption on this advertising revenue function:

Assumption B1 *$\forall k \in \{d, s, B\}$ and $\forall h \in \{SE, T\}$, $\alpha_{k,h}(\Omega^{x,h}, \Omega^{x,h} \cap \Omega^{x,-h})$ increases with $\Omega^{x,h}$ given $\Omega^{x,-h}$; and decreases with $\Omega^{x,h} \cap \Omega^{x,-h}$, given $\Omega^{x,h}$.*

We provide a microfoundation for this assumption in Appendix A. The first part of the assumption means that the effect of increasing the amount of data held by ad tech system h on its advertising revenue is positive. Holding $\Omega^{x,h} \cap \Omega^{x,-h}$ constant, this is intuitive, as more data enables better targeting. Furthermore, this effect stays positive even if we take into account the effect through $\Omega^{x,h} \cap \Omega^{x,-h}$ (given $\Omega^{x,-h}$). To understand it, consider a data merger such that a subset of the rival's data, $\Omega^{x,-h} - \Omega^{x,h}$, is added to $\Omega^{x,h}$. This should increase the ad revenue of h even if the effect from better targeting is mitigated because this part of data is also possessed by the rival ad tech $-h$. The second part means that when the rival ad tech $-h$'s data set and hence its overlap with h's data set expand, ad tech h's data becomes less exclusive such that some ads that previously only h was able to target are now accessible to $-h$ as well. As a result, the inventory served by ad tech h becomes less unique, lowering its advertising revenue.

Note that the advertising revenue $\alpha_{k,h}(\Omega^{x,h}, \Omega^{x,h} \cap \Omega^{x,-h})$ also depends on the nature of ad inventories, i.e., which activity k the ad impression is associated with. First, even under symmetric data allocation, α_d and α_B could still be different. This is because the nature of the site on which an advert is displayed affects the willingness to pay of advertisers and thereby the advertising revenue. For instance, reputable advertisers do not want to show their ads besides hate/racism content. Second, $\alpha_{d,T}$ is different from $\alpha_{s,T}$. As we assumed above, ad tech system T is unable to identify consumers in newspaper's search-referral traffic, $\Omega^{x,T} = \varnothing$ for ad inventories associated with activity s. This means that newspapers can only use contextual advertising for search referral traffic, which only uses context data but not consumer behavioral data. Hence, we add a simplifying assumption:

Assumption B2 *All newspapers use contextual advertising for search-referral traffic such that ad revenue for search-referral traffic is constant and given by $\alpha_s \equiv \alpha_{s,T}(\varnothing, \varnothing)$ regardless of data allocation.*

B2 allows us to focus on the substitution between the ad revenue of newspapers' direct traffic and the ad revenue of sector B. It plays a role mainly in the extension in which we analyze a divide-and-conquer strategy by making the analysis tractable. Due to assumption B1, the contextual advertising revenue α_s is lower than the revenue generated from personalized ads that makes use of behavioral data.[13]

Last, we assume that an ad tech h's payoff is a fixed share τ^h of the advertising revenue it generates. τ^h is exogenously given and belongs to $(0,1)$. We also call τ^h h's ad tech take.

2.3 The search engine and the roles of AMP

As introduced in the previous parts, the SE runs two businesses: it is a monopoly in the search market and faces competition from T in the ad intermediation market.

Due to its market power in the search market, the SE is able to influence newspapers' search referral traffic by altering their rankings in the search result. The search referral demand $D^{s,i}(\mathbf{q})$ represents the demand from an objective and non-distorted ranking of various news articles. When the ranking of a single newspaper, say newspaper i, is purposely promoted by the SE relative to the non-distorted search result, its search-referral traffic becomes $D^{s,i}(q_i, \mathbf{q}_{-i}; i+) \equiv D^{s|i+,i}(q_i, \mathbf{q}_{-i}) > D^{s,i}(q_i, \mathbf{q}_{-\mathbf{i}}), \forall \mathbf{q}$. And when newspaper i's ranking is purposely demoted, its search referral traffic becomes $D^{s,i}(q_i, \mathbf{q}_{-i}; i-) \equiv D^{s|i-,i}(q_i, \mathbf{q}_{-i}) < D^{s,i}(q_i, \mathbf{q}_{-\mathbf{i}}), \forall \mathbf{q}$.

The SE develops AMP technology, which overcomes the slow loading problem of web pages in mobile browsers. In addition, the SE will cache all web pages written in AMP format, allowing it to obtain consumers' browsing data regarding how they interact with these pages. As a result, if all newspapers adopt AMP for their traffic mediated by the search engine, the traffic loss δ will be eliminated but the SE collects the data $\omega^{x,s}$ on every consumer x and hence its data set expands to $\Omega^{x,SE} = \omega^{x,B} \cup \omega^{x,s}$.

2.4 Timing

Newspapers simultaneously decide whether to adopt the AMP and how much to invest in quality $q_i, i \in \{1, 2, ..., n\}$. At the same time, sector B chooses its quality q_B. Our equilibrium concept is Nash equilibrium, and we restrict our attention to the symmetric ones.

3 Two Benchmarks

In this section, we analyze two benchmarks for later use: one is the case without AMP, and the other one is a hypothetical situation where we assume SE cannot, through AMP, gain

[13]For empirical evidence showing that advertisers bid more for impressions that enable identification of consumers than for those with only context information available, see Appendix F p. 29 of CMA Report (2020), Beales (2010) and Srinivasan (2020). For instance, according to CMA Report (2020)'s study of data generated by Google's Randomised Controlled Trial (RCT) of display advertising, UK publishers earned approximately 70 percent less revenue overall when they were unable to sell inventory using personalised advertising (i.e., when cookies were not available) but competed against others who could.

access to data set $\omega^{x,s}$ of consumers.

3.1 Benchmark of no AMP

Recall that when there is no AMP, the data allocation across the two ad techs is $\Omega^T = \omega^d$ and $\Omega^{SE} = \omega^B$. Therefore, in this case, a newspaper's advertising revenue per direct traffic and sector B's advertising revenue per traffic are as follows:

$$\alpha_d^N \equiv \alpha_{d,T}(\Omega^{x,T}, \Omega^{x,T} \cap \Omega^{x,SE}) = \alpha_{d,T}(\omega^{x,d}, \omega^{x,d} \cap \omega^{x,B})$$
$$\alpha_B^N \equiv \alpha_{B,SE}(\Omega^{x,SE}, \Omega^{x,SE} \cap \Omega^{x,T}) = \alpha_{B,SE}(\omega^{x,B}, \omega^{x,B} \cap \omega^{x,d})$$

where the superscript N refers to the regime of no AMP.

Taking other newspapers' quality choices $\mathbf{q_{-i}}$ as given, newspaper i solves the following profit maximization problem, which is concave due to Assumption A1:

$$\max_{q_i} (1 - \tau^T) \left[\alpha_d^N D^{d,i}(q_i, \mathbf{q_{-i}}) + \alpha_s(1-\delta)D^{s,i}(q_i, \mathbf{q_{-i}})\right] - c(q_i)$$

where $\tau^T \in (0,1)$ is the ad tech take of T. Hence, the best response is characterized by the first-order condition:

$$(1 - \tau^T) \left[\alpha_d^N D_i^{d,i}(q_i, \mathbf{q_{-i}}) + \alpha_s(1-\delta)D_i^{s,i}(q_i, \mathbf{q_{-i}})\right] - c'(q_i) = 0.$$

The news quality at symmetric equilibrium is given by the following proposition:

Proposition 1 *In the benchmark without AMP, the news quality at the symmetric equilibrium, denoted by q^N, is characterized by*

$$(1 - \tau^T) \left[\alpha_d^N D_i^{d,i}(q^N, \dots, q^N) + \alpha_s(1-\delta)D_i^{s,i}(q^N, \dots, q^N)\right] = c'(q^N) \textit{ for all } i = 1, ..., n.$$

Sector B solves the problem of

$$\max_{q_B} (1 - \tau^{SE})\alpha_B^N D^B(q_B) - c_B(q_B)$$

where $\tau^{SE} \in (0,1)$ is the ad tech take of SE. Thus, the equilibrium quality q_B^N is determined by:

$$(1 - \tau^{SE})\alpha_B^N D^{B'}(q_B^N) - c'_B(q_B^N) = 0.$$

3.2 Benchmark of AMP without changes in data allocation

In this second benchmark, we consider a hypothetical situation where the SE cannot access to the data sets ω^s through AMP. In other words, newspapers' adoption of AMP does not lead to any change in the allocation of data, and only has the direct effect of eliminating the friction δ in search referral traffic. The advertising revenues are still given by α_d^N and α_B^N. Therefore, it is immediate that:

Proposition 2 *Consider a benchmark in which AMP is adopted by all newspapers, but it does not affect the allocation of consumer data.*

(i) The news quality at the symmetric equilibrium, denoted by q^*, is characterized by

$$(1-\tau^T)\left[\alpha_d^N D_i^{d,i}(q^*,\ldots,q^*)+\alpha_s D_i^{s,i}(q^*,\ldots,q^*)\right]=c'(q^*)\text{ for all } i=1,...,n.$$

(ii) Hence, AMP increases the quality of journalism: $q^* > q^N$.

This second benchmark establishes the result of one of the policy remedies we propose in Section 8.

4 AMP with Changes in Data Allocation: Static Analysis

From now on, we suppose that AMP changes the allocation of consumer data. In this section, we consider a static scenario in which we fix the quality levels in the two sectors at q^N and q_B^N and focus on the equilibrium in which all newspapers adopt the AMP format [14]. We analyze how AMP affects content providers' advertising revenues and the static welfare. The welfare result in this section will be contrasted with that in the next section where quality choices are endogenous.

4.1 The impact of AMP on data allocation and advertising revenue

When all newspapers adopt AMP, the data allocation between the two ad tech systems becomes: $\Omega^T=\omega^d, \Omega^{SE}=\omega^B\cup\omega^s$. Therefore, the ad revenue per unit of newspapers' direct traffic and the ad revenue per traffic in sector B respectively become:

$$\begin{aligned}
\alpha_d^M &\equiv \alpha_{d,T}(\Omega^{x,T},\Omega^{x,T}\cap\Omega^{x,SE})=\alpha_{d,T}(\omega^{x,d},\omega^{x,d}\cap\{\omega^{x,B}\cup\omega^{x,s}\})\\
\alpha_B^M &\equiv \alpha_{B,SE}(\Omega^{x,SE},\Omega^{x,SE}\cap\Omega^{x,T})=\alpha_{B,SE}(\omega^{x,B}\cup\omega^{x,s},\{\omega^{x,B}\cup\omega^{x,s}\}\cap\omega^{x,d}).
\end{aligned}$$

where the superscript M refers to the regime of AMP. Because of Assumption B1, we have

Lemma 1 *Suppose that all newspapers adopt the AMP format. This reduces the ad revenue for direct traffic to newspapers, while increases the ad revenue for traffic related to activity B:*

$$\alpha_d^N > \alpha_d^M, \alpha_B^N < \alpha_B^M.$$

To understand the result of $\alpha_d^N > \alpha_d^M$, consider consumer x who is a direct reader of newspaper 1. In the absence of AMP, $\Omega^{x,T}=\omega^{x,d}$ is the set of data that ad tech T has about this consumer and $\Omega^{x,SE}=\omega^{x,B}$ is the set of data that the SE obtains from her activity B.

[14] The existence of this equilibrium can be proven by following the logic of Proposition 11.

If the two activities are completely uncorrelated (i.e. $\omega^{x,d} \cap \omega^{x,B} = \varnothing$), the most appealing product to consumer x inferred from the data set $\omega^{x,d}$ will be different from the one inferred from $\omega^{x,B}$. If the quality of the data in $\omega^{x,d}$ is much better than that in $\omega^{x,B}$, the ad revenue newspaper 1 generates from x's direct visit would be much higher than what sector B generates. However, in the presence of AMP, if the SE gets access to the consumer x's data $\omega^{x,s}$, we have $\tilde{\Omega}^{x,SE} = \omega^{x,B} \cup \omega^{x,s}$. Then, if $\omega^{x,s}$ is correlated with $\omega^{x,d}$, this may allow the SE to infer sometimes the best match product that would be advertised by newspaper 1 alone in the absence of AMP. In such cases, the SE can engage in advertising arbitrage by targeting consumer x with the best match product in the ad space of sector B, reducing the ad revenue of newspaper 1.

4.2 Static welfare

We study how AMP changes static welfare given quality choices in both sectors. As each newspaper chooses quality q^N, their demands are given by $D^{d,1}(\mathbf{q}^N) = ... = D^{d,n}(\mathbf{q}^N) \equiv D^d(\mathbf{q}^N)$ for direct traffic and $D^{s,1}(\mathbf{q}^N) = ... = D^{s,n}(\mathbf{q}^N) \equiv D^s(\mathbf{q}^N)$ for search referral traffic, where $\mathbf{q}^N = (q^N, ..., q^N)$. And the demand in sector B is $D_B(q_B^N)$. We make the following assumption about advertisers' surplus:

Assumption B3 *(i) Given total advertising inventory (namely, traffic allocation), improving targeting in a subset of inventory increases total advertising surplus; (ii) Advertisers retain a constant share β of the total advertising surplus regardless of the presence of AMP.*

The second part of the assumption implies that we can express the total advertising surplus, which is defined as the sum of advertiser surplus and industry revenue, as $\frac{1}{1-\beta}$ times the joint advertising revenue of content providers and ad tech intermediaries. The advertising industry revenue in the absence of AMP is given as follows:

$$n(1-\tau^T+\tau^T)\underbrace{\left[\alpha_d^N D^d(\mathbf{q}^N)+\alpha_s(1-\delta)D^s(\mathbf{q}^N)\right]}_{\text{a single newspaper's ad revenue}}+(1-\tau^{SE}+\tau^{SE})\underbrace{\alpha_B^N D^B(q_B^N)}_{\substack{\text{total ad revenue}\\ \text{in sector B}}}$$

The welfare change induced by AMP comes from (i) the change in consumer welfare due to the elimination of traffic loss, which is positive; (ii) the change in the total surplus of the advertising sector, which is also positive:

$$\begin{aligned}
&\frac{1}{1-\beta}\left[n\alpha_d^M D^d(\mathbf{q}^N)+n\alpha_s D^s(\mathbf{q}^N)+\alpha_B^M D^B(q_B^N)\right]-\frac{1}{1-\beta}\left[n\alpha_d^N D^d(\mathbf{q}^N)+n\alpha_s(1-\delta)D^s(\mathbf{q}^N)+\alpha_B^N D^B(q_B^N)\right]\\
>&\frac{1}{1-\beta}\left[n\alpha_d^M D^d(\mathbf{q}^N)+n\alpha_s(1-\delta)D^s(\mathbf{q}^N)+\alpha_B^M D^B(q_B^N)\right]-\frac{1}{1-\beta}\left[n\alpha_d^N D^d(q^N)+n\alpha_s(1-\delta)D^s(\mathbf{q}^N)+\alpha_B^N D^B(q_B^N)\right]\\
>&0
\end{aligned}$$

The last inequality follows from Assumption B3 (i), which means that the increase in the advertising revenue from sector B due to better targeting more than compensates for the reduction in the advertising revenue from newspapers' direct traffic.

Proposition 3 *Suppose that all newspapers adopt the AMP format. Under B1-B3, AMP strictly increases static welfare.*

AMP improves static welfare even if AMP does not affect the loss rate of search-referral traffic (i.e., when the loss rate δ is close to zero). This is because AMP enables SE to use the data from search referral traffic to improve targeting efficiency in sector B, which in turn increases total surplus in the advertising industry.

5 AMP with Changes in Data Allocation: Dynamic Analysis

In this section, we consider the main scenario in which content providers' quality choices are endogenously determined and study how AMP affects dynamic welfare. We focus on the equilibrium in which all newspapers adopt AMP. The existence of such equilibrium is verified in the next section.

5.1 Quality choice

Expecting all other newspapers to adopt AMP and taking their quality choices $\mathbf{q_{-i}}$ as given, the newspaper i that also adopts AMP solves the following profit maximization problem:

$$\max_{q_i} \, (1-\tau^T) \left[\alpha_d^M D^{d,i}(q_i, \mathbf{q_{-i}}) + \alpha_s D^{s,i}(q_i, \mathbf{q_{-i}}) \right] - c(q_i)$$

The best response is determined by the first-order condition:

$$(1-\tau^T) \left[\alpha_d^M D_i^{d,i}(q_i, \mathbf{q_{-i}}) + \alpha_s D_i^{s,i}(q_i, \mathbf{q_{-i}}) \right] - c'(q_i) = 0.$$

Hence, the news quality in the symmetric equilibrium in which all newspapers adopt AMP is given by the following proposition:

Proposition 4 *When all newspapers adopt the AMP format, the news quality at the symmetric equilibrium, denoted by q^M, is characterized by*

$$(1-\tau^T) \left[\alpha_d^M D_i^{d,i}(q^M, ..., q^M) + \alpha_s D_i^{s,i}(q^M, ..., q^M) \right] = c'(q^M) \textit{ for all } i = 1, ..., n.$$

Compared to the condition in Proposition 1 which characterizes the equilibrium news quality in the absence of AMP, newspapers' adoption of AMP has two opposite effects on their incentives to invest in quality. On the one hand, it increases the marginal benefit of investment by eliminating the loss δ in search referral traffic. On the other hand, it discourages investment as the data allocation reduces newspapers' advertising revenue in direct traffic. Therefore, we have the following result:

Proposition 5 *The effect of the AMP on news quality is ambiguous: news quality is reduced when the negative data allocation effect dominates. (e.g., $\delta = 0$, $\alpha_d^M << \alpha_d^N$), while the news quality is increased when the positive search traffic enhancing effect dominates. (e.g., $\delta >> 0, \alpha_d^M \simeq \alpha_d^N$)*

In the presence of AMP, sector B solves the problem of

$$\max_{q_B} (1-\tau^{SE})\alpha_B^M D^B(q_B) - c_B(q_B).$$

So the equilibrium quality q_B^M is determined by:

$$(1-\tau^{SE})\alpha_B^M D^{B'}(q_B^M) - c'_B(q_B^M) = 0.$$

q_B^M is higher than q_B^N as the increased advertising revenue per traffic increases the marginal benefit of investing in quality.

5.2 Welfare analysis

To facilitate welfare analysis, in this subsection, we impose a specific structure on demand functions. Regarding the demand in direct traffic, suppose that each consumer would read k^d number of news articles when she visits her favorite news outlet directly. Assume that q_i stands for the quality of a single article in newspaper i. Then newspaper i's overall quality is $k^d q_i$. As consumers single-home on one newspaper in terms of direct traffic, we apply the discrete-choice Logit model. Let consumer x's utility from direct visit to newspaper i be:

$$U_{x,i}^d = ln\big(v^d(k^d q_i)\big) + \varepsilon_{x,i}^d, \;\; i = 0, 1, 2, ..., n,$$

where "0" stands for the outside option whose quality is q_0. In addition, $\varepsilon_{x,i}^d$ is i.i.d. according to the Type I Extreme Value distribution with scale parameter $\mu^d > 0$. Each consumer chooses the newspaper that delivers the highest utility.

This yields the following direct demand of newspaper i: $D^{d,i}(\mathbf{q}) = \frac{\tilde{v}^d(k^d q_i)}{\Sigma_{j=0}^n \tilde{v}^d(k^d q_j)}$, where $\tilde{v}^d(\cdot) = (v^d(\cdot))^{\frac{1}{\mu^d}}$. When we assume that $v^d(\cdot)$ is increasing and concave and that $\mu^d > 1$, this demand function satisfies the assumptions in A1. From this specification, we obtain the following consumer surplus from direct traffic: $CS^d = \mu^d ln(\sum_{i=0}^n \tilde{v}^d(k^d q_i))$.

In addition to direct demand, each consumer also has demand for k^s number of news articles via search. We assume that each consumer makes k^s independent search queries and that each query is associated with a separate discrete-choice problem. As a result, a consumer multihomes in terms of search-referral traffic in the sense that she reads articles from different newspapers in different searches. Let consumer x 's utility from reading an article of newspaper i discovered through the SE be $U_{x,i}^s = ln\big(v^s(q_i)\big) + \varepsilon_{x,i}^s, \;\; i = 0, 1, 2, ..., n$ in which $\varepsilon_{x,i}^s$ is i.i.d. according to the Type I Extreme Value distribution with scale parameter $\mu^s > 0$. These preference shocks are independent across search queries and uncorrelated with those of direct traffic. Therefore, the search-referral demand of newspaper i without AMP is $D^{s,i}(\mathbf{q}) = k^s(1-\delta)\frac{\tilde{v}^s(q_i)}{\Sigma_{j=0}^n \tilde{v}^s(q_j)}$, where $\tilde{v}^s(\cdot) = (v^s(\cdot))^{\frac{1}{\mu^s}}$.

Last, let consumer x's utility derived from activity B be $U_x^B = ln\big(v^B(q^B)\big) + \varepsilon_x^B$ and the utility derived from activity B's outside option be $U_{x,0}^B = ln\big(v^B(q_0^B)\big) + \varepsilon_{x,0}^B$, in which ε_x^B

and $\varepsilon_{x,0}^B$ are as before i.i.d. according to the Type I Extreme Value distribution with scale parameter $\mu^B > 0$. And let $\tilde{v}^B(\cdot) = (v^B(\cdot))^{\frac{1}{\mu^B}}$ denote the normalized valuation of quality.

Furthermore, we introduce the following assumption:

Assumption A2 *Newspapers are sufficiently differentiated, namely μ^d and μ^s are sufficiently large.*

μ^d and μ^s are sufficiently large in the sense that it guarantees Lemma 3 and Lemma 5 that we introduce later on.

5.2.1 SE's profit

As the SE's profit is a constant share τ^{SE} of the advertising revenue generated in sector B for supplying ad tech service, it is immediate from the last subsection that SE always benefits from newspapers' adoptions of AMP—the data allocation effect of AMP not only increases the advertising revenue per traffic but also expands ad inventory in sector B:

$$\Pi^{SE,N} = \tau^{SE}\alpha_B^N D^B(q_B^N) < \Pi^{SE,M} = \tau^{SE}\alpha_B^M D^B(q_B^M)$$

Formally, we have the following result:

Proposition 6 *The SE always gains from newspapers' adoption of the AMP format. Therefore, even if the adoption of the AMP format reduces the quality of journalism, the SE has no incentive to internalize it.*

We assumed that the SE has no consumer-facing services related to activity B. In reality, Google owns many consumer-facing services, which will even strengthen our results, as Google will retain the whole benefit from having access to search-referral data instead of just having a fraction τ^{SE} of it.

5.2.2 Comparison of newspaper profit

When there is no AMP, the equilibrium newspaper profit is:

$$\pi^N(\alpha_d^N, \delta) = (1-\tau^T)[\alpha_d^N D^d(\mathbf{q}^N) + \alpha_s(1-\delta)D^s(\mathbf{q}^N)] - c(q^N).$$

We have the two following lemmas.

Lemma 2 *When there is no AMP, the news quality at the symmetric equilibrium q^N is decreasing in traffic friction δ.*

Lemma 3 *Under Assumption A2, a newspaper's profit π^N in the symmetric equilibrium without AMP satisfies $\frac{\partial \pi^N(\alpha_d^N,\delta)}{\partial \alpha_d^N} > 0$ and $\frac{\partial \pi^N(\alpha_d^N,\delta)}{\partial \delta} < 0$.*

The lemma means that the direct effect of a positive exogenous shock to the industry, such as increased ad price or reduced loss rate, dominates the negative effect of intensified competition on industry profit at symmetric equilibrium.

When there is AMP, the equilibrium newspaper profit becomes:

$$\pi^M(\alpha_d^M) = (1-\tau^T)[\alpha_d^M D^d(\mathbf{q}^M) + \alpha_s D^s(\mathbf{q}^M)] - c(q^M).$$

We obtain the following result about the comparison of newspapers' profit between the two regimes:

Proposition 7 *Under assumptions A1, A2, B1, and B2 we have:*

(1). the presence of AMP reduces the newspaper industry profit when δ is close to zero;

(2). When a_d^M is close enough to a_d^N, there exists a threshold $0 < \tilde{\delta} < 1$ determined by $\pi^N(\alpha_d^N, \tilde{\delta}) = \pi^M(\alpha_d^M)$, such that AMP increases the newspaper industry profit if δ belongs to $(\tilde{\delta}, 1)$.

5.2.3 Comparison of consumer surplus in content market

When there is no AMP, the equilibrium consumer surplus is:

$$CS^N(\delta) = \mu^d ln\big(\tilde{v}^d(k^d q_0) + n\tilde{v}^d(k^d q^N)\big) + k^s(1-\delta)\mu^s ln\big(\tilde{v}^s(q_0) + n\tilde{v}^s(q^N)\big) + \mu^B ln\big(\tilde{v}^B(q_0^B) + \tilde{v}^B(q_B^N)\big)$$

Note that only the first two terms in CS^N depend on δ.

Lemma 4 *When there is no AMP, consumer surplus CS^N is decreasing in δ.*

From Lemma 2, when there is no AMP, the equilibrium news quality is decreasing in the loss rate δ. And because consumer surplus is increasing in the news quality, it is immediate that CS^N is decreasing in δ.

When there is AMP, total consumer surplus is:

$$CS^M = \mu^d ln\big(\tilde{v}^d(k^d q_0) + n\tilde{v}^d(k^d q^M)\big) + k^s\mu^s ln\big(\tilde{v}^s(q_0) + n\tilde{v}^s(q^M)\big) + \mu^B ln\big(\tilde{v}^B(q_0^B) + \tilde{v}^B(q_B^M)\big)$$

Note that CS^M does not depend on δ.

Proposition 8 *Under assumptions A1, A2, B1, and B2, we have:*

(1) If $CS^N|_{\delta=0} \leq CS^M$, consumer surplus is always higher with AMP;

(2) If $CS^N|_{\delta=0} > CS^M$, there exists a threshold $\bar{\delta} > 0$ determined by $CS^N(\bar{\delta}) = CS^M$ such that when $0 \leq \delta \leq \min\{\bar{\delta}, 1\}$, consumer surplus is lower with AMP; when $\min\{\bar{\delta}, 1\} < \delta \leq 1$, consumer surplus is higher with AMP.

This proposition suggests that when AMP lowers the quality of journalism, AMP's impact on consumer surplus depends on the trade-off between the gain from higher content quality in sector B and the loss from lower quality of journalism. For instance, consider the case of $\delta \approx 0$, where AMP does not bring much efficiency but reduces the equilibrium quality of newspapers through data leakage. If AMP lowers consumer surplus in this case, then AMP reduces consumer surplus for any δ below a certain threshold.

5.2.4 Comparison of social welfare

Because newspapers and sites in sector B in our model are ad-financed, they create values by providing content to consumers as well as by selling advertising inventories to advertisers through ad intermediaries. Under Assumption B3, the aggregate social welfare without AMP is:

$$W^N = \underbrace{CS^N - nc(q^N) - c^B(q_B^N)}_{\text{social surplus in content industry}} + \underbrace{\frac{1}{1-\beta}\left[n\alpha_d^N D^d(\mathbf{q}^N) + n\alpha_s(1-\delta)D^s(\mathbf{q}^N) + \alpha_B^N D^B(q_B^N)\right]}_{\text{social surplus in advertising industry}}$$

And the aggregate social welfare with AMP is:

$$W^M = CS^M - nc(q^M) - c^B(q_B^M) + \frac{1}{1-\beta}\{n\alpha_d^M D^d(\mathbf{q}^M) + n\alpha_s D^s(\mathbf{q}^M) + \alpha_B^M D^B(q_B^M)\}$$

Lemma 5 *Under Assumption A2, newspapers always underinvest in quality relative to the quality level chosen by a social planner to maximize social welfare.*

Our model does not consider positive externalities to the society from high quality journalism such as improving voting outcomes by informing voters or making politicians accountable and so on. The lemma should hold a fortiori if we take into account such positive externalities.

In the following proposition, we provide a sufficient condition for the adoption of AMP to be socially efficient:

Proposition 9 *If $q^M > q^N$, then aggregate social welfare is higher with AMP (i.e., $W^M > W^N$).*

In this case, the effect of AMP on social welfare can be decomposed into three parts: (1) the adoption of AMP directly creates more surplus in both content market and advertising market by eliminating the loss of traffic due to slow loading; (2) a higher content quality in each sector improves consumer surplus and creates more advertising opportunities by expanding traffic to newspapers and sites in sector B; (3) data leakage to site B increases ad targeting, which further increases surplus in advertising market.

However, if the equilibrium quality of newspapers is lower with AMP, the effect of AMP on welfare is ambiguous, as it depends on:

(i) whether consumers are affected more by the decreased quality in journalism or by the increased quality in sector B;

(ii) whether the direct effect of eliminating the loss of search-referral traffic δ brought by AMP is large enough, which includes gains in both consumer surplus and advertising surplus;

(iii) whether the advertising surplus generated by increased ad inventory in sector B is high enough relative to the loss from the reduced inventory of newspapers;

(iv) to what degree data leakage improves matching efficiency of ad inventories in sector B.

Therefore, AMP is highly likely to reduce welfare when δ is close to zero, consumers value the quality of newspapers much more than that of sector B, and advertisers value the ad inventory of newspapers much more than that of sector B.

The next proposition provides a sufficient condition for AMP adoption to reduce welfare. In the proposition, we shut down both the effect on the traffic loss rate and the one on the quality in sector B to focus on the main trade-off, but by continuity the result carries over when the two effects are small.

Proposition 10 *Suppose $\delta = 0$, $q_B^M = q_B^N \equiv q_B$ and $q^M \ll q^N$. Then, if the positive effect of data leakage on advertising surplus is dominated by the negative effect of lower quality of journalism on welfare, aggregate social welfare is lower with AMP (i.e., $W^M < W^N$).*

6 Newspapers' Incentive to Adopt AMP

In this section, we study the incentives of newspapers to adopt AMP. As before, we focus on symmetric equilibria.

Newspapers face a collective action problem as one newspaper's adoption of AMP generates two kinds of negative externalities to the other newspapers: search ranking externality and data leakage externality. First, the search ranking externality arises as the SE promotes adopters' articles in search results. Therefore, a newspaper's adoption will have a negative impact on the search referral traffic of other newspapers. Second, a newspaper's adoption leaks data of other newspapers' direct readers to the SE, which reduces their advertising revenues from direct traffic.

To facilitate the exposition in this section, we refine the notation of the data set $\omega^{x,s}$ to be $\omega^{x,s}(l)$, where l is the number of newspapers adopting AMP. So in the previous sections we mean $\omega^{x,s}(n)$ by $\omega^{x,s}$. Then, when there are l newspapers adopting AMP, we denote the ad revenue per direct traffic of newspapers and the ad revenue per traffic of sector B as:

$$\alpha_d(l) \equiv \alpha_{d,T}(\omega^{x,d}, \omega^{x,d} \cap \{\omega^{x,B} \cup \omega^{x,s}(l)\})$$
$$\alpha_B(l) \equiv \alpha_{B,SE}(\{\omega^{x,B} \cup \omega^{x,s}(l)\}, \{\omega^{x,B} \cup \omega^{x,s}(l)\} \cap \omega^{x,d}),$$

where for simplicity, we assume that, given l, $\alpha_d(l)$ is the same for all newspapers, regardless of whether one has adopted AMP or not. Therefore, $\alpha_d^M \equiv \alpha_d(n)$, $\alpha_B^M \equiv \alpha_B(n)$ and $\alpha_d^N \equiv \alpha_d(0)$, $\alpha_B^N \equiv \alpha_B(0)$.

From Assumption B1, $\alpha_d(l)$ decreases and $\alpha_B(l)$ increases in l. This is because when there are more newspapers adopting AMP, the scale of data leakage is larger. This effect is in place in the absence of search distortion. In addition, the SE's promotion of AMP articles in search results further increases data leakage.

We first verify the existence of an equilibrium in which all newspapers adopt the AMP format. As explained in Section 2, the SE could leverage its market power in the search market to implement AMP by promoting adopters' rankings in the search result. When newspaper i is the only non-adopter, its articles are demoted in the search result, and hence, its search referral traffic becomes $D^{s,i}(q_i, \mathbf{q}_{-i}; i-) \equiv D^{s|i-,i}(q_i, \mathbf{q}_{-i}) < D^{s,i}(q_i, \mathbf{q}_{-\mathbf{i}})$. When n is sufficiently large, the demotion of newspaper i in search ranking implies that its articles almost never appear in the first pages of search results and hence $D^{s|i-,i}(q_i, \mathbf{q}_{-i})$ is close to zero. Moreover, this implies $\alpha_d(n-1) \simeq \alpha_d(n)$. This is because when n is large, consumers are always directed to AMP articles such that the SE can perfectly track their search-referred news reading activities. Then, additional adoption of newspaper i has no impact on SE's data set.

Formally, we assume:

Assumption A3 *The number of competing newspapers n is sufficiently large such that the demotion of newspaper i in search rankings makes $D^{s|i-,i}(q_i, \mathbf{q}^M_{-i})$ close to zero for any $(q_i, \mathbf{q}^M_{-i})$ and $\mathbf{q}^M_{-i} = (q^M, \ldots, q^M)$, and that $\alpha_d(n-1) \simeq \alpha_d(n)$.*

Note that for the above assumption to hold, n does not need to be very large. For instance, if 5 articles per mobile page are shown in a search result and most consumers stop scrolling down from the fifth page, then $n > 21$ is enough.

Then we have:

Proposition 11 *Under assumptions A1, A3, B1, and B2, there exists an equilibrium in which all newspapers adopt the AMP format.*

We emphasize that when $\delta \approx 0$, the SE's leverage of its search monopoly power through demotion of non-adopters' positions is crucial in sustaining the all-adoption equilibrium. As we can see in the above inequality, without punishment in terms of ranking, non-adoption will not affect the search-referral traffic and remove the (small) negative impact of data leakage on the advertising revenue in direct traffic (i.e., $\alpha_d(n-1) > \alpha_d(n)$). This induces newspaper i not to adopt the AMP format.

Next, we examine when there exists another symmetric equilibrium in which no newspaper adopts the AMP format. Suppose newspaper i is the only adopter of AMP such that the SE promotes its AMP articles in search results. Hence its search referral traffic becomes $D^{s,i}(q_i, \mathbf{q}_{-i}; i+) \equiv D^{s|i+,i}(q_i, \mathbf{q}_{-i}) > D^{s,i}(q_i, \mathbf{q}_{-\mathbf{i}})$. Newspapers' ad revenue per direct traffic becomes $\alpha^M_d(1)$, satisfying $\alpha^M_d \equiv \alpha_d(n) < \alpha_d(1) < \alpha_d(0) \equiv \alpha^N_d$.

We provide a sufficient condition for the existence of the non-adoption equilibrium:

Proposition 12 *Under assumptions A1, B1 and B2, there exists an equilibrium in which no news-*

paper adopts the AMP format if the following condition is satisfied:

$$\underbrace{(\alpha_d^N - \alpha_d(1))D^{d,i}(q^+, \mathbf{q}_{-i}^N)}_{\substack{\textit{reduction in ad revenue}\\ \textit{due to data leakage}}} > \underbrace{\alpha_s \left[D^{s|i+,i}(q^+, \mathbf{q}_{-i}^N) - (1-\delta)D^{s,i}(q^+, \mathbf{q}_{-i}^N)\right]}_{\substack{\textit{increase in search referral traffic}\\ \textit{due to promotion and higher speed}}}$$

where q^+ is the quality choice of newspaper i after its deviation:

$$q^+ = \arg\max_{q_i}(1-\tau^T)\left[\alpha_d(1)D^{d,i}(q_i, \mathbf{q}_{-i}^N) + \alpha_s D^{s|i+,i}(q_i, \mathbf{q}_{-i}^N)\right] - c(q_i).$$

In short, the all-adoption equilibrium always exists and the non-adoption equilibrium can also exist if the loss from data leakage is large enough relative to the expansion of search referral demand.

7 Extension: Divide-and-Conquer

This section relaxes the assumption in the baseline model that all newspapers use the ad tech system T by supposing that $m(< n)$ number of newspapers use the ad tech SE instead. We assume that search ranking of the SE is neutral with respect to whether a newspaper uses the ad tech T or SE.

In this case, there are two ways through which the SE collects consumers' browsing data on news pages directed by search. First, when the news page a consumer visits via search belongs to a newspaper using the ad tech SE, the SE collects her browsing data. Second, when the news page belongs to a newspaper that uses the ad tech T but has adopted AMP, the SE obtains her data by hosting the relevant page. We denote the amount of data the SE captures regarding a consumer x in search referral traffic as $\omega^{x,s}(m,l,g)$, where m is the number of newspapers using the ad tech SE, l is the number of newspapers adopting AMP among those who use the ad tech T, and g is the number of newspapers adopting AMP among those who use the ad tech SE . We impose the following assumption on $\omega^{x,s}(m,l,g)$:

Assumption B4 *The amount of data the SE obtains from search referral traffic on a consumer x, $\omega^{x,s}(m,l,g)$, has the following properties:*[15]
(1) Given l and g, it is increasing with the number of newspapers using its ad tech service: $\frac{\partial \omega^{x,s}(m,l,g)}{\partial m} \geq 0$.
(2) Given m, it is increasing with the number of newspapers adopting AMP: $\frac{\partial \omega^{x,s}(m,l,g)}{\partial l} \geq 0$ *and* $\frac{\partial \omega^{s}(m,l,g)}{\partial g} \geq 0$.

The properties of $\frac{\partial \omega^{x,s}(m,l,g)}{\partial m} \geq 0$ and $\frac{\partial \omega^{x,s}(m,l,g)}{\partial l} \geq 0$ are straightforward as the SE gains more access to data. The property of $\frac{\partial \omega^{x,s}(m,l,g)}{\partial g} \geq 0$ is because the SE promotes AMP articles

[15] m, l, g are integers. In this assumption, we abuse the notation of partial derivative to simplify the expression.

in search results, which increases data leakage to the SE. For instance, more data is leaked to the SE when a consumer is diverted to an AMP article away from a non-AMP article of a newspaper using the ad tech T.

Now, the amount of data a newspaper has about its direct readers and hence the ad revenue per direct traffic will depend on which ad tech service it uses. For any newspaper using the ad tech T, the data allocation regarding its direct reader x is $\Omega^{x,T} = \omega^{x,d}$, $\Omega^{x,SE} = \omega^{x,s}(m,l,g) \cup \omega^{x,B}$. We denote the ad revenues generated from consumer x's direct visit to the newspaper and her visit to sector B as:

$$\alpha_d^T(m,l,g) = \alpha_{d,T}(\omega^{x,d}, \omega^{x,d} \cap \left\{\omega^{x,s}(m,l,g) \cup \omega^{x,B}\right\})$$
$$\alpha_B^T(m,l,g) = \alpha_{B,SE}(\left\{\omega^{x,s}(m,l,g) \cup \omega^{x,B}\right\}, \left\{\omega^{x,s}(m,l,g) \cup \omega^{x,B}\right\} \cap \omega^{x,d})$$

With some abuse of notation, the superscript T in $\alpha_B^T(m,l,g)$ denotes the advertising revenue of sector B from a consumer who is a direct reader of a newspaper that uses the ad tech T.

For any newspaper using the ad tech SE, given a direct consumer x', the SE tracks various online activities of the consumer and, therefore, collects the data set $\Omega^{x',SE} = \omega^{x',d} \cup \omega^{x',s}(m,l,g) \cup \omega^{x',B}$. Therefore, the ad revenues generated from her direct visit to the newspaper and from her visit to sector B are:

$$\begin{aligned}
\alpha_d^{SE}(m,l,g) &\equiv \alpha_{d,SE}(\Omega^{x',SE}, \Omega^{x',SE} \cap \Omega^{x',SE}) \\
&= \alpha_{d,SE}(\omega^{x',d} \cup \omega^{x',s}(m,l,g) \cup \omega^{x',B}, \omega^{x',d} \cup \omega^{x',s}(m,l,g) \cup \omega^{x',B}) \\
\alpha_B^{SE}(m,l,g) &\equiv \alpha_{B,SE}(\Omega^{x',SE}, \Omega^{x',SE} \cap \Omega^{x',SE}) \\
&= \alpha_{B,SE}(\omega^{x',d} \cup \omega^{x',s}(m,l,g) \cup \omega^{x',B}, \omega^{x',d} \cup \omega^{x',s}(m,l,g) \cup \omega^{x',B}).
\end{aligned}$$

In this case, both ad inventories associated with consumer x' are served by the same ad tech intermediary. We add the following assumption:

Assumption B5 *When both the ad inventory from direct traffic and the one from Sector B are served by the same ad tech intermediary, we assume $\alpha_{k,h}(\Omega^{x,h}, \Omega^{x,h})$ is increasing with $\Omega^{x,h}, \forall k \in \{d,s,B\}, h \in \{T, SE\}$.*

This assumption simply says that when an ad tech gains more data, it can improve ad targeting and thus increase each client's advertising revenue.

We point out the conflict of interest between two groups of newspapers. Whereas the ad revenue per direct traffic of a newspaper using the ad tech SE increases with the amount of search-referral data collected by the SE, the ad revenue of a newspaper using T decreases with the amount of search-referral data collected by the SE.

We first show that it is a dominant strategy for a newspaper using the ad tech service of the SE to adopt the AMP format. Consider newspaper i that uses the ad tech SE and takes as given the quality and the adoption choices of other newspapers. If newspaper i adopts the

AMP, it improves its ranking and eliminates traffic loss δ, while the SE keeps having access to the browsing data from $m + l$ newspapers in search-referred news reading activities. So the only change in terms of data leakage is that i's adoption expands $\omega^{x,s}(m, l, g)$ to $\omega^{x,s}(m, l, g + 1)$ as the SE promotes its AMP articles, which in turn increases the ad revenue per direct traffic from $\alpha_d^{SE}(m, l, g)$ to $\alpha_d^{SE}(m, l, g + 1)$ according to B5. Therefore, we have:

Lemma 6 *Under A1, B1, B2, B4 and B5, it is a dominant strategy for a newspaper using the ad tech service of the SE to adopt the AMP format.*

From an argument analogous to the one used for Proposition 11, it is straightforward to see the existence of the adoption equilibrium in which all newspapers using the ad tech T also adopt AMP.

In what follows, we show that for m large enough, the adoption equilibrium is the unique equilibrium. We take it as given that the m newspapers using the ad tech SE adopt AMP. Consider an equilibrium candidate in which no newspaper using the ad tech T adopts the AMP format. Let $D^{s|-,i}(q_i, \mathbf{q}_{-i}; m)$ (respectively, $D^{s|+,i}(q_i, \mathbf{q}_{-i}; m)$) represent the search-referral traffic when newspaper i is a non-adopter (respectively, an adopter) when there are m number of adopters. Denote the quality vector in this equilibrium candidate $\mathbf{q} = (q^T, \ldots, q^T, q^{SE}, \ldots, q^{SE})$, where $q_i = q^T$ satisfies

$$(1 - \tau^T) \left[\alpha_d^T(m, 0, m) D_i^{d,i}(\mathbf{q}) + \alpha_s(1 - \delta) D_i^{s|-,i}(\mathbf{q}; m) \right] = c'(q^T);$$

and $q_i = q^{SE}$ satisfies

$$(1 - \tau^{SE}) \left[\alpha_d^{SE}(m, 0, m) D_i^{d,i}(\mathbf{q}) + \alpha_s D_i^{s|+,i}(\mathbf{q}; m) \right] = c'(q^{SE}).$$

A newspaper using the service of T has an incentive to deviate by adopting the AMP format if the following inequality holds:

$$\begin{aligned} & (1 - \tau^T) \left[\alpha_d^T(m, 0, m) D^{d,i}(\mathbf{q}) + \alpha_s(1 - \delta) D^{s|-,i}(\mathbf{q}; m) \right] - c(q^T) \\ \leq \; & (1 - \tau^T) \left[\alpha_d^T(m, 1, m) D^{d,i}(\widehat{q}, \mathbf{q}_{-i}) + \alpha_s D^{s|+,i}(\widehat{q}, \mathbf{q}_{-1}; m + 1) \right] - c(\widehat{q}) \end{aligned}$$

where $\widehat{q}$ is determined by

$$\widehat{q} = \arg\max_{q_i} (1 - \tau^T) \left[\alpha_d^T(m, 1, m) D^{d,i}(q_i, \mathbf{q}_{-i}) + \alpha_s D^{s|+,i}(q_i, \mathbf{q}_{-1}; m + 1) \right] - c(q_i).$$

A sufficient condition is that the inequality holds at $\widehat{q} = q^T$, which is equivalent to

$$\left[\alpha_d^T(m, 1, m) - \alpha_d^T(m, 0, m) \right] D^{d,i}(\mathbf{q}) + \alpha_s \left[D^{s|+,i}(\mathbf{q}; m + 1) - (1 - \delta) D^{s|-,i}(\mathbf{q}; m) \right] \geq 0.$$

When m is large enough (this is possible as n is large),[16] because of the SE's promotion of AMP articles, almost all the search traffic is directed to the AMP articles, implying that

[16]For the same reason given right after A3, m does not need to be very large

$D^{s|-,i}(\mathbf{q}; m)$ is close to zero. This in turn implies $\alpha_d^T(m,1,m) \simeq \alpha_d^T(m,0,m)$. Therefore, $\alpha_s D^{s|+,i}(\mathbf{q}; m+1)$ dominates all the other terms in the above inequality and, hence , the condition is satisfied. The same logic applies to any equilibrium candidate in which $l(< n-m)$ number of newspapers using the ad tech T adopt the AMP format in addition to the m number of newspapers using the ad tech SE. Therefore, we have the following result:

Proposition 13 *Under A1, A3, B1, B2, B4 and B5, there exists a threshold $\widehat{m}$ such that for $m > \widehat{m}$, all newspapers adopt the AMP format in the unique equilibrium.*

To conclude, we have shown in this section that the SE can combine its market power in search and ad intermediation to deploy a divide-and-conquer strategy to achieve the unique outcome in which all newspapers adopt AMP.

Last, we verify that the SE is indeed better off in an equilibrium where all newspapers adopt AMP relative to the benchmark without AMP. This is straightforward because all kinds of ad revenues generated by the ad tech SE are greater under AMP due to its larger data collection, i.e., $\omega^s(m, n-m, m) \geq \omega^s(m, n-m, 0) > \omega^s(m, 0, 0)$:

$$\begin{aligned}
\alpha_B^T(m, n-m, m) &> \alpha_B^T(m, 0, 0), \\
\alpha_d^{SE}(m, n-m, m) &> \alpha_d^{SE}(m, 0, 0), \\
\alpha_B^{SE}(m, n-m, m) &> \alpha_d^{SE}(m, 0, 0).
\end{aligned}$$

8 Policy Remedies and Discussion

8.1 Remedies

We propose two policy remedies to eliminate the two types of externalities, search ranking externality and data leakage externality, that generate the collective action problem among newspapers. First, in order to eliminate the search ranking externality, we propose that the SE use an objective criterion of loading speed and treat all articles meeting such criterion in a non-discriminatory way, regardless of whether the technology is developed by Google or not. This remedy can be interpreted as neutrality obligations. In particular, it allows the SE to demote articles that do not meet the speed criterion. Such exercise of search market power is socially desirable as long as newspapers' adoption of speed-enhancing technology improves welfare. If the SE does not discriminate articles at all, newspapers may not adopt the technology, for instance, when the adoption requires each newspaper to pay a high fixed cost. Second, in order to eliminate the data leakage externality, we propose to prohibit the SE from hosting articles on its server. If all newspapers adopt technologies meeting the speed criterion, the combination of the two policy remedies leads to the equilibrium we described in the benchmark of AMP without change in data allocation (characterized in Proposition 2), where adoption of AMP improves quality of journalism and social welfare.

The potential anti-competitive effect of gatekeeper platforms' collection and use of data has raised policy makers' concerns. Our remedy of prohibiting the dominant platform from unfairly gaining access to business user's data is echoed in the European Commission's Digital Markets Act (2020): "a gatekeeper shall...refrain from using, in competition with business users, any data not publicly available, which is generated through activities by those business users, including by the end users of these business users, of its core platform services or provided by those business users of its core platform services or by the end users of these business users".

One may argue that requiring gatekeeper platforms to share with news publishers data regarding how end users interact with the latter's content on the platform is a more efficient remedy for addressing the negative impact of AMP, as it preserves targeting efficiency in the advertising market. However, it is easy to show that this does not necessarily eliminate the negative impact of AMP on newspapers' ad revenues in direct traffic, particularly when the shared data is much more valuable to the platform than to the publishers. Then, it is more effective to forbid the platform from gaining access to the data in the first place.

8.2 Discussions

There are two additional channels through which Google leverages its market power in search and/or ad intermediation to impose anti-competitive rules on news publishers, which deserve discussion due to their close relationship with this paper.

The first channel is directly related to AMP and concerns the header bidding issue. Browser-side header bidding is a programming technique that enables a publisher to directly call multiple ad exchanges from the user's browser to collect bids. However, Google restricts the use of browser-side header bidding in AMP, which allows Google to leverage its monopoly power in the publisher ad server market to favor its own ad exchange over competitors.[17] We exclude this channel from our analysis, as it is about Google's leverage of its power in one layer of the ad intermediation market to favor its own business in another, and analyzing it requires us to model the complex chain of ad intermediation, which is beyond the scope of this paper. [18]

The second channel is the controversial news aggregator problem, which has been widely debated in the past decade. [19] This debate is about Google's practice of displaying snippets and images of news articles in its Search and Google News products without compensating news publishers for the use. Therefore, news publishers in different countries and continents, including those in the EU, the U.S. and Australia, have called for the reform of

[17]This is because to work around the AMP constraint on the use of JavaScript, the header bidding auctions have to take place in the publisher ad server, which is monopolized by Google. Google's publisher ad server disadvantages competing ad exchanges by imposing tighter latency restrictions on them. This will not only make it hard for bids collected by competing ad exchanges to arrive in time, but will also result in more cookie syncing failure and hence consumer data loss for these competitors.

[18]Interested readers could refer to Srinivasan (2020) and the complaint of the state of Texas, et al. against Google (2020) for more details on this specific aspect of AMP and refer to Jeon (2021) and the ACCC Final Report (2021) for the broader problem of Google's leveraging its power in different layers of ad tech value chain to promote its own services.

[19]See Jeon (2018) for a survey of the literature on news aggregators.

copyright laws to make news aggregators such as Google and Facebook pay for the use of their content.

This issue is closely related to our paper, as publishers' negotiations with Google over compensation features a similar coordination problem to that we have identified in Section 6. Specifically, Google's power to allocate newspapers' search-referral traffic by manipulating rankings in search results generates search ranking externalities among newspapers. For instance, Google can make only newspapers that have licensed their content eligible for the "Top Stories" carousel. To avoid being demoted in search results, newspapers would license their content to Google in exchange for a small amount of or zero compensation.

Policy markers in various jurisdictions are wary of this collective action issue. To address it, the Australian news media bargaining code[20]clearly requires that a digital platform, in its activities of crawling, indexing, making available and distributing news businesses' content, not discriminate by whether a news business is paid or not paid by the platform for making available the news business' covered news content.

In a similar vein, in the decision 21-D-17 of 12 July 2021[21], the French Authorité de la Concurrence fined Google for violating the neutrality obligation not to link the negotiation over related rights to the indexing, classification and presentation of protected content taken up by Google on its services. What Google did was to link the negotiation on the compensation to participation in the new Google News Showcase program. Because Showcase is initially embedded in the Google News and will be integrated into general Google search results, it is reasonable to expect that whether or not a news publisher participates in the Showcase program will greatly impact its search-referral traffic. As a consequence, "the strategy implemented by Google has thus strongly encouraged publishers to accept the contractual conditions of the Showcase service and to renounce negotiations relating specifically to the current uses of protected content...under the risk of seeing their exposure and their remuneration degraded compared to their competitors who would have accepted the proposed terms", says the decision.[22] Our policy remedy can be interpreted as neutrality obligations in search results as long as publishers' articles meet an objective criterion of loading speed.

The neutrality obligations can have even broader implications beyond the search engine market for preventing digital platforms from using their ranking algorithm to induce business users into unfair contractual terms. For instance, in Italy, Amazon was fined for harming competitors by favoring third-party sellers that use the company's logistics services in the Amazon website's search results. As a remedy, the Italian regulator ordered Amazon to

[20]https://www.accc.gov.au/focus-areas/digital-platforms/news-media-bargaining-code27

[21]https://www.autoritedelaconcurrence.fr/en/press-release/remuneration-related-rights-press-publishers-and-agencies-autorite-fines-google-500

[22]The German Bundeskartellamt also opened an investigation into the Google News Showcase issue, particularly regarding the announced integration of the Google News Showcase service into Google's general search function. See https://www.bundeskartellamt.de/SharedDocs/Meldung/EN/Pressemitteilungen/2021/04_06_2021_Google_Showcase.html

offer ”fair and nondiscriminatory standards” for listings from third-party sellers, which it would monitor through an appointed trustee[23].

9 Conclusion

AMP allows Google to collect data from articles written in this format and to combine them with data from other sources in order to improve targeting of the ads served by Google on various websites. Even if such data combination improves static welfare, we found that it can reduce dynamic welfare by reducing newspapers' incentive to invest in the quality of journalism. In particular, we showed that Google has no incentive to internalize the impact of its conduct on news quality.

In this paper, we considered only Google's ad provision to third-party sites/apps. However, in reality, Google owns many consumer-facing products and serves ads in these products as well. Considering Google-owned ad inventory will strengthen the conflict between newspapers' investment in journalism and Google's data combination for ad targeting that we identified. In particular, given that a small fraction of valuable consumers explains most ad revenue of publishers,[24] Google has an incentive to engage in cream-skimming by showing ads to valuable consumers in its own products, which further exacerbates the conflict.

Our paper compared two particular data allocations and combinations that arise depending on whether AMP exists or not. In this simple environment, neutrality obligations in search results together with no hosting of news content by Google are desirable policy remedies. However, we consider our paper a call for future research to address a more fundamental and general question: what is the optimal scope of data combination that takes into account both static and dynamic efficiency? If the optimal scope of data combination turns out to be narrower than Google's current practice, which combines a vast majority of data from third-party publishers with its own first-party data, then a policy intervention would be required to implement the optimal scope because a collective action problem would prevent publishers from maintaining a proper control of their data.

References

ACCC Final Report (2021), "Digital advertising services inquiry-final report." Report, ACCC, Austrilia.

Ambrus, Attila, Emilio Calvano, and Markus Reisinger (2016), "Either or both competition: A" two-sided" theory of advertising with overlapping viewerships." *American Economic Journal: Microeconomics*, 8, 189–222.

[23] Eric Sylvers and Sam Schechner. "Amazon Fined $1.3 Billion in Italian Antitrust Case." The Wall Street Journal, Dec. 9, 2021, https://www.wsj.com/articles/amazon-fined-1-3-billion-in-italian-antitrust-case-11639043714

[24] See the complaint of the state of Texas, et al v. Google, LLC (2020) at https://www.texasattorneygeneral.gov/sites/default/files/images/admin/2020/Press/20201216%20COMPLAINT_REDACTED.pdf

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, Øystein Foros, and Hans Jarle Kind (2018), "Competition for advertisers and for viewers in media markets." *The Economic Journal*, 128, 34–54.

Anderson, Simon P and Martin Peitz (2020), "Media see-saws: Winners and losers in platform markets." *Journal of Economic Theory*, 186, 104990.

Athey, Susan, Emilio Calvano, and Joshua S Gans (2018), "The impact of consumer multihoming on advertising markets and media competition." *Management Science*, 64, 1574–1590.

Beales, Howard (2010), "The value of behavioral targeting." *Network Advertising Initiative*, 1, 2010.

Bourreau, Marc, Janina Hofmann, and Jan Krämer (2021), "Prominence-for-data schemes in digital platform ecosystems: Implications for platform bias and consumer data collection." Working paper.

Calzada, Joan and Ricard Gil (2020), "What do news aggregators do? evidence from Google news in Spain and Germany." *Marketing Science*, 39, 134–167.

CMA Report (2020), "Online platforms and digital advertising market study final report." Report, Competition and Markets Authority, UK.

Condorelli, Daniele and Jorge Padilla (2020), "Data-driven envelopment with privacy-policy tying." working paper.

D'Annunzio, Anna and Antonio Russo (2020), "Ad networks and consumer tracking." *Management Science*, 66, 5040–5058.

D'Annunzio, Anna and Antonio Russo (2021), "Intermediaries in the online advertising market."

de Cornière, Alexandre and Miklos Sarvary (2022), "Social media and news: Content bundling and news quality." *Management Science*.

de Cornière, Alexandre and Greg Taylor (2020), "Data and competition: a general framework with applications to mergers, market structure, and privacy policy." Working paper.

Dellarocas, Chrysanthos, Zsolt Katona, and William Rand (2013), "Media, aggregators, and the link economy: Strategic hyperlink formation in content networks." *Management Science*, 59, 2360–2379.

Digital Markets Act (2020), "Proposal for a regulation on digital markets act." Report, European Commission, EU.

Geradin, Damien and Dimitrios Katsifis (2019), "An EU competition law analysis of online display advertising in the programmatic age." *European Competition Journal*, 15, 55–96.

Ghosh, Arpita, Mohammad Mahdian, R Preston McAfee, and Sergei Vassilvitskii (2015), "To match or not to match: Economics of cookie matching in online advertising." *ACM Transactions on Economics and Computation (TEAC)*, 3, 1–18.

Jeon, Doh-Shin (2018), "Economics of news aggregators." In *Economic Analysis of the Digital Revolution* (Juan-José Ganuza and Gerard Llobet, eds.), 343–366, Funcas, Spain.

Jeon, Doh-Shin (2021), "Market power and transparency in open display advertising–a case study." Final report, Expert Group for the Observatory on the Online Platform Economy.

Jeon, Doh-Shin and Nikrooz Nasr (2016), "News aggregators and competition among newspapers on the internet." *American Economic Journal: Microeconomics*, 8, 91–114.

Johnson, Justin, Thomas Jungbauer, and Marcel Preuss (2021), "Online advertising, data sharing, and consumer control." Working paper.

Krämer, Jan, Daniel Schnurr, and Michael Wohlfarth (2019), "Winners, losers, and Facebook: The role of social logins in the online advertising ecosystem." *Management Science*, 65, 1678–1699.

Madsen, Erik and Nikhil Vellodi (2021), "Insider imitation." *Available at SSRN 3832712*.

OECD (2021), "Competition issues concerning news media and digital platforms." OECD competition committee discussion paper. https://www.oecd.org/daf/competition/competition-issues-in-news-media-and-digitalplatforms.htm.

Rolnik, Guy, Julia Cagé, Joshua Gans, Ellen Goodman, Brian Knight, Andrea Prat, Anya Schiffrin, and Prateek Raj (2019), "Protecting journalism in the age of digital platforms." *Committee for the Study of Digital Platforms Media Subcommittee. Chicago: Stigler Center for the Study of the Economy and the State. University of Chicago Booth School of Business*.

Scott Morton, Fiona M and David C Dinielli (2020), "Roadmap for a digital advertising monopolization case against google." *Omidyar Network, May*.

Srinivasan, Dina (2020), "Why Google dominates advertising markets." *Stan. Tech. L. Rev.*, 24, 55.

U.S. House of Representatives (2020), "Investigation of competition in digital markets, majority staff report and recommendations." Report, Subcommittee on Antitrust, Commercial and Administrative Law of the Committee on the Judiciary, U.S.

Zuboff, Shoshana (2019), *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power: Barack Obama's Books of 2019*. Profile Books.

Appendix

A Micro-foundation of B1

Suppose that there are $N > 0$ advertisers who are interested in showing ads to consumer x.

Suppose that a given set of data about consumer x, Ω^x, is available to the advertisers. They use the data to estimate their willingness to pay. Their estimations generate a vector of the willingness to pay

$$\mathbf{v}(\Omega^x) = (\widetilde{v}_1(\Omega^x), \widetilde{v}_2(\Omega^x), ..., \widetilde{v}_N(\Omega^x))$$

where $\widetilde{v}_k(\Omega^x)$ is the kth-highest willingness to pay and is a random variable. In one extreme of $\Omega^x = \varnothing$, we assume that $\widetilde{v}_1(\varnothing) = \widetilde{v}_2(\varnothing) = ..., = \widetilde{v}_N(\varnothing) = v^e$ where v^e is a positive constant. In the other extreme of perfect information $\Omega^x = \mathbf{\Omega}^x$, $\widetilde{v}_i(\mathbf{\Omega}^x) = v_i$ for $i = 1, ..., N$ with

$$v_1 > v_2 > ...(> v^e >)... > v_{N-1} > v_N.$$

As Ω^x increases from $\varnothing$ to $\mathbf{\Omega}^x$, the expected values of $\widetilde{v}_1(\Omega^x)$ and $\widetilde{v}_2(\Omega^x)$ increase to v_1 and $v_2(>> v^e)$ whereas the expected values of $\widetilde{v}_{N-1}(\Omega^x)$ and $\widetilde{v}_N(\Omega^x)$ decrease to $v_{N-1}(<< v^e)$ and v_N.

We assume that the expected values of the three highest valuations $\widetilde{v}_1(\Omega^x), \widetilde{v}_2(\Omega^x), \widetilde{v}_3(\Omega^x)$ are increasing in Ω^x.

Consider two sets $\Omega^{x,A}$ and $\Omega^{x,B}$ such that $\Omega^{x,A} \cap \Omega^{x,B} = \varnothing$. Consider two independent second-price auctions, each selling one spot: auction A uses data $\Omega^{x,A}$ and auction B uses data $\Omega^{x,B}$. Then, we assume that the probability that the highest bidder of one auction will be also the highest bidder or the second-highest bidder of the other auction is zero. This in turn implies that the outcomes of the two auctions do not depend on the sequential order of the auctions.

Consider now expanding $\Omega^{x,B}$ to $\Omega^{x,B\prime}$ such that $\Omega^{x,A} \cap \Omega^{x,B\prime} \neq \varnothing$. If auction A runs before auction B, the change in $\Omega^{x,B}$ does not affect the outcome of auction A: we here make a simplifying assumption that advertisers are myopic and hence the advertiser with valuation $\widetilde{v}_1(\Omega^{x,A})$ prefers participating in the first auction instead of giving up the first auction in order to participate in the second auction. If auction B runs before auction A, there is a probability $p(\Omega^{x,A} \cap \Omega^{x,B\prime})$, which increases with $\Omega^{x,A} \cap \Omega^{x,B\prime}$, that the winner of auction B has either $\widetilde{v}_1(\Omega^{x,A})$ or $\widetilde{v}_2(\Omega^{x,A})$. In this case, the ad revenue of auction A will be $\widetilde{v}_3(\Omega^{x,A})$ instead of $\widetilde{v}_2(\Omega^{x,A})$.

Finally, assume that consumer x is reader of newspaper i. She visits everyday the site of newspaper i and another site for activity B. But the order of her visit is random: with equal probability, she visits each site first and then visits the other site. Each site sells one ad spot per day. Then, the expected ad revenue of the newspaper from direct visit is

$$\begin{aligned}
& \alpha_A(\Omega^{x,A}, \Omega^{x,A} \cap \Omega^{x,B\prime}) \\
= \quad & (1-\frac{1}{2}p(\Omega^{x,A} \cap \Omega^{x,B\prime}))\widetilde{v}_2^e(\Omega^{x,A}) + \frac{1}{2}p(\Omega^{x,A} \cap \Omega^{x,B\prime})\widetilde{v}_3^e(\Omega^{x,A}) \\
= \quad & \widetilde{v}_2^e(\Omega^{x,A}) - \frac{1}{2}p(\Omega^{x,A} \cap \Omega^{x,B\prime}) \left[\widetilde{v}_2^e(\Omega^{x,A}) - \widetilde{v}_3^e(\Omega^{x,A})\right]
\end{aligned}$$

where the superscript e represents expectation. $\widetilde{v}_2^e(\Omega^{x,A})$ increases with $\Omega^{x,A}$. $\frac{1}{2}p(\Omega^{x,A} \cap \Omega^{x,B\prime})$ increases with $\Omega^{x,A} \cap \Omega^{x,B\prime}$ for given $\Omega^{x,A}$, which satisfies the second part of B1. In order to satisfy the first part of B1, either the second component $\frac{p(\Omega^{x,A} \cap \Omega^{x,B\prime})}{2}\left[\widetilde{v}_2^e(\Omega^{x,A}) - \widetilde{v}_3^e(\Omega^{x,A})\right]$ is weakly decreasing in $\Omega^{x,A}$ or the effect from the first component $\widetilde{v}_2^e(\Omega^{x,A})$ should dominate the effect from the second component.

B Omitted Proofs

Proof of Lemma 2. Differentiating the equilibrium condition in Proposition 1 with respect to δ on both sides, we get:

$$(1-\tau^T)\left[\alpha_d^N \Sigma_{j=1}^n D_{i,j}^{d,i}(\mathbf{q}^N) + \alpha_s(1-\delta)\Sigma_{j=1}^n D_{ij}^{s,i}(\mathbf{q}^N)\right]\frac{\partial q^N}{\partial \delta} - (1-\tau^T)\alpha_s D_i^{s,i}(\mathbf{q}^N) = c''(q^N)\frac{\partial q^N}{\partial \delta}$$

Solving for $\frac{\partial q^N}{\partial \delta}$, we obtain:

$$\frac{\partial q^N}{\partial \delta} = \frac{(1-\tau^T)\alpha_s D_i^{s,i}(\mathbf{q}^N)}{(1-\tau^T)\left[\alpha_d^N \Sigma_{j=1}^n D_{i,j}^{d,i}(\mathbf{q}^N) + \alpha_s(1-\delta)\Sigma_{j=1}^n D_{ij}^{s,i}(\mathbf{q}^N)\right] - c''(q^N)}$$

By Assumption A1, we have $D_i^{s,i} > 0$, $D_{ij}^{d,i} < 0$ and $D_{ij}^{s,i} < 0$ for $i, j = 1, 2, .., n$, and the cost function $c(q)$ is convex. Therefore, we have $\frac{\partial q^N}{\partial \delta} < 0$. □

Proof of Lemma 3. We only prove $\frac{\partial \pi^N(\alpha_d^N, \delta)}{\partial \delta} < 0$, as $\frac{\partial \pi^N(\alpha_d^N, \delta)}{\partial \alpha_d^N} > 0$ can be proved analogously.

$$\begin{aligned}
-\frac{\partial \pi^N}{\partial \delta} &= -\Big[(1-\tau^T)[\alpha_d^N \Sigma_{j=1}^n D_j^{d,i}(\mathbf{q}^N) + \alpha_s(1-\delta)\Sigma_{j=1}^n D_j^{s,i}(\mathbf{q}^N)] - c'(q^N)\Big]\frac{\partial q^N}{\partial \delta} + (1-\tau^T)\alpha_s D^{s,i}(\mathbf{q}^N) \\
&= \underbrace{-(1-\tau^T)[\alpha_d^N \Sigma_{j\neq i} D_j^{d,i}(\mathbf{q}^N) + \alpha_s(1-\delta)\Sigma_{j\neq i} D_j^{s,i}(\mathbf{q}^N)]\frac{\partial q^N}{\partial \delta}}_{\text{the effect of intensified competition (-)}} + \underbrace{(1-\tau^T)\alpha_s D^{s,i}(\mathbf{q}^N)}_{\text{the direct effect of positive shock (+)}} \\
&> 0
\end{aligned}$$

The second equality follows from the optimality of q^N for newspaper i's profit maximization problem. When newspapers are sufficiently differentiated, the effect of intensified competition will be small such that the last inequality holds. □

Proof of Proposition 7. (1) As $\frac{\partial \pi^N(\alpha_d^N, \delta)}{\partial \alpha_d^N} > 0$ from Lemma 3, when $\delta = 0$, we have $\pi^N(\alpha_d^N, 0) > \pi^N(\alpha_d^M, 0) = \pi^M(\alpha_d^M)$. This implies that the presence of AMP reduces the newspaper industry profit. (2) Taking the values of δ and α_d^N as given, we have $\lim_{\alpha_d^M \uparrow \alpha_d^N} \pi^M(\alpha_d^M) = \pi^N(\alpha_d^N, 0)$. This property together with $\frac{\partial \pi^N}{\partial \delta} < 0$ (from Lemma 3) guarantees the existence of the threshold $0 < \tilde{\delta} < 1$. $\square$

Proof of Proposition 8. (1) This is straightforward as consumer surplus CS^N is decreasing in δ from Lemma 4. (2) Recall from Proposition 5 that AMP raises news quality when the positive search traffic enhancing effect dominates the negative data allocation effect, and this condition holds particularly when $\delta >> 0$ and α_d^M is close to α_d^N. In this case, both the quality of newspapers and that of sector B are higher under AMP, implying $CS^M > CS^N(\delta)$. This result, together with the condition $CS^N|_{\delta=0} > CS^M$ and Lemma 4, implies that there is a threshold $\bar{\delta}$, below which $CS^N(\delta) > CS^M$ and above which $CS^N(\delta) < CS^M$. $\square$

Proof of Lemma 5. For instance, in the case without AMP, the newspapers' private incentive to invest in quality is given by the first-order condition in Proposition 1. By contrast, the first-order condition of the social planner's problem is:

$$\begin{aligned}\frac{\partial W^N}{\partial q^N} =& \frac{\partial CS^N}{\partial q^N} - nc'(q^N) + \frac{1}{1-\beta}\Big[n\alpha_d^N\left[D_i^{d,i}(\mathbf{q}^N) + \Sigma_{j\neq i}^n D_j^{d,i}(\mathbf{q}^N)\right] \\ &+ n\alpha_s(1-\delta)\left[D_i^{s,i}(\mathbf{q}^N) + \Sigma_{j\neq i}^n D_j^{s,i}(\mathbf{q}^N)\right]\Big] \\ =& 0\end{aligned}$$

Therefore, the difference in their incentives to invest in quality can be expressed as::

$$\begin{aligned}\frac{1}{n}\frac{\partial W^N}{\partial q^N} - \frac{\partial \pi^N(\mathbf{q}^N)}{\partial q_i} =& \underbrace{\frac{1}{n}\frac{\partial CS^N}{\partial q^N}}_{(+)} + \underbrace{\left(\frac{1}{1-\beta} - 1 + \tau^T\right)\left(\alpha_d^N D_i^{d,i} + \alpha_s(1-\delta)D_i^{s,i}\right)}_{(+)} \\ &+ \underbrace{\frac{1}{1-\beta}\left(\alpha_d^N \Sigma_{j\neq i}^n D_j^{d,i} + \alpha_s(1-\delta)\Sigma_{j\neq i}^n D_j^{s,i}\right)}_{(-)}\end{aligned}$$

When newspapers are sufficiently differentiated (as we assume in A2), the third term, which is the business stealing effect, is weak and dominated by the first two positive terms. As a result, the social planner has a greater incentive to improve newspapers' quality. $\square$

Proof of Proposition 9. Recall that we assume newspapers are sufficiently differentiated such that q^M is lower than the quality chosen by the social planner. Similarly, sites in sector B also have a smaller incentive to invest in quality than the social planner. As the model's assumptions on demand and cost functions guarantee that the social surplus function is concave, we have $W^M(q^M, q_B^M) > W^M(q^N, q_B^N)$ when $q^M > q^N$. Under Assumption B3 (i), we have $n\alpha_d^M D^d(\mathbf{q}^N) + n\alpha_s D^s(\mathbf{q}^N) + \alpha_B^M D^B(q_B^N) > n\alpha_d^N D^d(\mathbf{q}^N) + n\alpha_s D^s(\mathbf{q}^N) + \alpha_B^N D^B(q_B^N)$. Therefore, $W^M(q^M, q_B^M) > W^M(q^N, q_B^N) > CS^M(\mathbf{q}^N, q_B^N) - nc(q^N) - c^B(q_B^N) + \frac{1}{1-\beta}\left[n\alpha_d^N D^d(\mathbf{q}^N) + n\alpha_s D^s(\mathbf{q}^N) + \alpha_B^N D^B(q_B^N)\right] > W^N$. $\square$

Proof of Proposition 10. Under the condition $\delta = 0$ and $q_B^M = q_B^N$,

$$W^M(q^M) - W^N(q^N, \delta = 0) = \underbrace{\left[W^M(q^M) - W^N(q^M, \delta = 0)\right]}_{\text{efficiency effect of data leakage}} - \underbrace{\left[W^N(q^N, \delta = 0) - W^N(q^M, \delta = 0)\right]}_{\text{direct effect of lower quality}}$$

The main trade-off is captured by the above equation. By taking the quality of newspapers q^M as given, the first bracket represents the positive effect of data leakage, which leads to improved matching in targeted advertising, as captured in the static welfare analysis. The second captures the negative effect of lower quality on welfare. As assumptions on demand functions and cost functions guarantee that social surplus is concave in symmetric quality and q^N is assumed to be lower than the social optimum, any quality lower than q^N induces lower welfare. $\square$

Proof of Proposition 11. Let q^- be the best response of newspaper i when it unilaterally deviates by not adopting AMP but expects all other newspapers to adopt AMP and to choose the equilibrium quality q^M:

$$q^- = \arg\max_{q_i}(1 - \tau^T)\left[\alpha_d(n-1)D^{d,i}(q_i, \mathbf{q}_{-i}^M) + \alpha_s(1-\delta)D^{s|i-,i}(q_i, \mathbf{q}_{-i}^M)\right] - c(q_i).$$

Newspaper i has no incentive to deviate by not adopting the AMP format because:

$$\begin{aligned}
& (1 - \tau^T)\left[\alpha_d^M D^{d,i}(q^M, \mathbf{q}_{-i}^M) + \alpha_s D^{s,i}(q^M, \mathbf{q}_{-i}^M)\right] - c(q^M) \\
\geq\ & (1 - \tau^T)\left[\alpha_d^M D^{d,i}(q^-, \mathbf{q}_{-i}^M) + \alpha_s D^{s,i}(q^-, \mathbf{q}_{-i}^M)\right] - c(q^-) \\
>\ & (1 - \tau^T)\left[\alpha_d(n-1)D^{d,i}(q^-, \mathbf{q}_{-i}^M) + \alpha_s(1-\delta)D^{s|i-,i}(q^-, \mathbf{q}_{-i}^M)\right] - c(q^-)
\end{aligned}$$

where the first inequality is from the optimality of q^M in newspaper i's profit maximization problem, and the second inequality holds under A3. $\square$

Proof of Proposition 12.

$$\begin{aligned}
& (1 - \tau^T)\left[\alpha_d^N D^{d,i}(q^N, \mathbf{q}_{-i}^N) + \alpha_s(1-\delta)D^{s,i}(q^N, \mathbf{q}_{-i}^N)\right] - c(q^N) \\
\geq & (1 - \tau^T)\left[\alpha_d^N D^{d,i}(q^+, \mathbf{q}_{-i}^N) + \alpha_s(1-\delta)D^{s,i}(q^+, \mathbf{q}_{-i}^N)\right] - c(q^+) \\
\geq & (1 - \tau^T)\left[\alpha_d(1)D^{d,i}(q^+, \mathbf{q}_{-i}^N) + \alpha_s D^{s|i+,i}(q^+, \mathbf{q}_{-i}^N)\right] - c(q^+)
\end{aligned}$$

The first inequality follows from the optimality of q^N for newspaper i's profit maximization problem when no newspaper adopts the AMP. The second inequality follows from the condition in the proposition. $\square$

C Industry Background on Open Display Market

In this section, we briefly introduce how the open programmatic display advertising market works to help readers get familiar with the context of our model and the motivation behind our modeling choices. For further references, see CMA Report (2020) and its online appendices for a detailed and comprehensive survey of digital advertising market. Also see Geradin and Katsifis (2019) and Srinivasan (2020) for their analysis of online display advertising issues from the angle of competition law.

The online display advertising market is composed of two segments, depending on whether or not ad inventories are sold through intermediaries. The first involves own-and-operated platforms such as Google Search and Facebook App, which sell a large amount of ad inventories from their consumer-facing services through their proprietary ad interfaces. The other segment, which is the focus of this paper, is the open display advertising market, in which a large number of publishers (such as newspapers, blogs, app owners and any other content/service providers) sell their ad inventory to a large number of advertisers through a complex chain of third-party ad intermediaries. These intermediaries, which are also called ad techs, organize and/or participate in real-time bidding auctions on behalf of publishers and advertisers. Examples of open display ads are the banner or video ads we frequently see on websites and apps.

How is consumer data used in personalized targeting? As display ads are usually targeted, consumer data plays an important role in determining what ads are relevant to a consumer and how much advertisers bid for ad impressions. To learn about consumers' purchasing intents, ad techs track consumers' online activities across websites and devices to infer what products might appeal to them. For instance, an ad intermediary can predict that a consumer may be interested in seeing the ad of the latest iPhone if she spends a lot of time reading tech news and reviews on smart phones. Based on the collected data, the essential work of ad techs is to build consumer profiles, each of which is a group of segments. Using the last example, the consumer profile could be {Female, France, Phone,...}, with each entry representing a segment. Accordingly, advertisers will create their audience by defining targeting criteria in terms of segments. For instance, a smart phone retailer can set her targeting audience as $\{$location=France, monthly income $>$ 1K, Phone$\}$. Then, when a consumer visits a publisher's website, its ad server will send a bid request together with some data about the consumer (including user identifiers). This information is passed to ad intermediaries such as demand-side platforms (DSPs) which help advertisers to evaluate their willingness to pay and to make bids by matching the received information with consumer profiles that they built. An advertiser will bid on a consumer if she belongs to predefined audience. The winner finally displays an ad of her product on the page the consumer is browsing.

Tracking.[25] In such an environment, an ad tech's success largely depends on how much consumer data it has, which in turn depends on the ability to track consumers across websites and devices. To compile a certain consumer's browsing activities conducted on different sites, the tracker needs to (i) learn that the consumer is visiting a web page when the event takes place; and (ii) identify the consumer in order to associate his different browsing events together.

[25] See a detailed explanation in Online Appendix G of CMA Report (2020)

The first point is done by embedding third-party codes on first-party websites. When a consumer visits a web page in which the code writes that it needs content input from third-party websites, her browser will send requests to both the first party (the website she is visiting) and the servers of the third-party websites. Information such as the referrer's URL, device info, IP addresses, etc can be passed along with the request. In this way, a third-party tracker learns that a consumer is visiting a website that contains its codes. The content fetched from a third-party tracker's server could be a banner/video ad if the tracker provides ad serving service to the first-party website or simply a 1x1 pixel transparent GIF that is invisible to visitors if the tracker provides analytic services to the first-party website. On the web, these third-party codes are called tags and pixels. Their counterparts in mobile apps are Third Party Libraries (TPLs) or Software Development Kits (SDKs). For a tracker to recognize that it is the same individual who visits a series of websites that embed its codes, the request sent to the tracker needs to contain a unique user identifier attached to the consumer or to her browser. This is mostly done via the best-known use of cookies. Cookies are small text files that a website's server drops in the browser when the server responds to the browser's request. Most importantly, it contains a randomly generated string of letters and numbers to serve as an identifier. For example, suppose that both the WSJ and the NYT use the ad service of Google's DoubleClick (which would be a third-party tracker in the example). When a consumer visits the WSJ for the first time, the browser will make requests to both WSJ and DoubleClick's servers, as the page needs both the news content and ads to fill spaces. When sending back those required contents, the WSJ and DoubleClick respectively set a cookie in the visitor's browser. The WSJ's cookie is called first-party cookie as it belongs to the domain the consumer is visiting, while the DoubleClick's is called third-party cookie.

Cookies are private to domains such that only the domain that sets the cookies can read them. However, cookies can be sent back whenever the browser requests content from their owners, as long as the user did not delete it. Continuing with the above example, suppose now that the consumer visits the NYT. As the NYT also requests ad input from DoubleClick, the cookie set earlier by DoubleClick will be sent back along with this request. By reading the cookie identifier, DoubleClick knows it is the same consumer who previously visited WSJ now being on NYT. As a result, DoubleClick can compile consumer activities on these two websites together.

One issue with cookies is that, as they are randomly generated, the identifiers in cookies set in a browser from different domains are different. As a consequence, when a publisher uses ad tech A to serve ads, whereas an advertiser uses ad tech B, the advertiser cannot identify the consumer of which the impression is on sale with the cookie ID set by ad tech A. Then, ad tech B engages in cookie matching (also called cookie syncing) during a real-time bidding process to identify the same consumer in its own database and to evaluate the advertiser's willingness to pay for the impression. This process of cookie matching is prone to failure, resulting in approximately 30 percent failed matching.

In addition to cookie IDs, trackers can also use email addresses, IP addresses, user account IDs, device info, or a combination of them to identify consumers. In particular, trackers in mobile apps use mobile advertising IDs (MAIDs) as user identifiers, which are unique to mobile devices and shared with all apps. Therefore, all tracking parties in mobile apps share a common identifier associated with each device, and they save the trouble of cookie matching, as in the web tracking case.

Finally, to build more complete user profiles, trackers need to perform cross-device tracking to link MAIDs with cookie IDs. This can be greatly facilitated by IP addresses, email addresses or first-party login details/internal IDs.

C.1 Stylized Facts on the Competition in the Ad Intermediation Market

The Ad intermediation market consists of several layers along its complex value chain from publishers to advertisers. On the supply side, there are publisher ad servers and supply side platforms (SSPs) and on the demand side, there are demand side platforms (DSPs) and advertiser ad servers.

Because of various acquisitions and the leverage of data, advertising inventories, and speed advantage, Google is currently the dominant player at each vertical layer of ad intermediation. Below, we report Google's market shares in the UK provided by CMA Report (2020). The publisher ad server market is monopolized by Google, as Google Ad Manager accounts for more than 90 percent of the display ads served in the UK. Google has a 50-60 percent share in the SSP market in the UK. Google's DSP DV360 has a 30-40 percent market share. Google operates a DSP through Google Ads, which has a 10-20 percent market share. Hence, the combined market share in DSP becomes 40-60 percent in the UK. The advertiser ad server market is highly concentrated, and Google accounts for approximately 80-90 percent of the ads served to UK users. We describe in detail how Google gains data advantage that can be leveraged in competition.

Sources of Google's Data Advantage:

- Google offers a wide range of leading consumer-facing services. For instance, Google provides more than 53 consumer-facing services and products in the UK, including Google Search, YouTube, and Gmail. (Appendix F of CMA Report (2020), 2020, p. F8). This allows it to collect a vast amount of first-party consumer data and to derive valuable insights about users. For instance, search data are very useful to advertisers as a source of learning purchase intent.

- Google can leverage the first-party data it has to attract publishers and advertisers to use its own ad intermediary by restricting access to those valuable first-party data to its proprietary platforms. To provide services, Google places its trackers on customers' websites and apps. According to CMA Report (2020), Google was found to be present as a third-party on approximately 85% of websites.

- Because consumers, especially Android users, log in to their Google account on each of their devices, Google has an advantage in cross-device tracking.
- Mainstream browsers are starting to ban the use of third-party cookies to protect consumer privacy. For instance, Apple's Safari and Mozilla's Firebox have blocked third party cookies by default and Google also plans to do so in Chrome in the coming years. This will hurt rival ad techs more than Google, as the former rely more heavily on the use of third-party cookies to collect information.

Implications for Competition Outcome. The lack of competition in ad intermediation translates into high ad tech fees, which is commonly referred to as "ad tech take". Ad tech take represents the difference between what advertisers pay and what publishers earn from digital advertising. The CMA report estimates that "on average publishers receive around 65% of initial advertising revenue that is paid by advertisers (i.e., the overall 'ad tech take' is around 35%)". Another estimate on the ad tech tax from the Wall Street Journal could be as high as 60%. [26].

[26] See https://www.wsj.com/articles/behavioral-ad-targeting-not-paying-off-for-publishers-study-suggests-11559167195