WORKING PAPERS

N° 18-912

April 2018

“Economics of News Aggregators"

Doh-Shin Jeon

Toulouse School of Economics

Economics of News Aggregators*

Doh-Shin Jeon†

April 2, 2018

Abstract

The success of news aggregators has generated a heated debate about whether news aggregators steal traffic from newspapers or increase traffic to newspapers. This survey article provides an overview of recent articles on news aggregators. After providing a simple theoretical framework, I first review empirical articles and then theoretical ones. While the empirical articles try to assess the effects on traffic to newspapers, the theoretical ones go beyond and try to analyze the effects on newspapers' incentive to invest in quality journalism. I conclude by raising some questions for future research.

1 Introduction

The traditional ad-based business model of newspapers has been in crisis because of declining revenues from newspaper advertising. According to Pew Research Center (2017), newspapers' revenues from advertising have fallen approximately 62% since 2000: it was $48.67 billion in 2000 but $18.27 billion in 2016. In particular, entry of online classified-ad competitors such as Craigslist substantially reduced newspapers' revenue. Even if the share of digital advertising revenue has increased from 17% in 2011 to 29% in 2016, it is far from reversing the downfall in advertising revenue.[1] As a consequence, newspaper newsroom employment decreased by 37% for the period of 2004-2015 from 65,440 to 41,400.

Newspapers are in stiff competition with new online media. Among online media sources, news aggregators seem to be the most important. According to an Outsell report (2009), 57 percent of news media users go to digital sources, and they are also more likely to turn to an aggregator (31 percent) than to a newspaper site (8 percent) or other news site (18 percent).

*I would like to thank Luis Abreu, Gerard Llobet and Nikrooz Nasr for the very useful comments.

†Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France, dohshin.jeon@gmail.com

[1] According to The Economist ("Taxing Times", 10 Nov. 2012), in France, not a single national newspaper is profitable despite around €1.2 billion in direct and indirect government subsidies.

Indeed, Reuters Institute (2015) shows that aggregators (Yahoo! News, Google News, MSN, Buzzfeed and Huffington Post) attract 80% of the online news traffic in the U.S. In South Korea, the two major news aggregators, Naver and Daum, each had a share of 55.4% and 22.4% in the Internet news traffic in 2016 (Choi, 2017) whereas the traffic to newspaper home pages had only 4% share in 2017 (Korean Press Foundation, 2017).

The success of news aggregators has generated a heated debate about the effects of news aggregators on newspapers' incentive to produce high-quality content. During 2009 to 2010, the FTC hosted three workshops and published a controversial discussion draft (FTC, 2010) that hints at copyright reform and the protection of newspapers from aggregators. In Europe, the German Parliament introduced in 2013 a change in the copyright law that allowed news aggregators to link for free the news stories of news outlets if using excerpts of less than 7 words. Longer excerpts or images would require the payment of a negotiated fee to the news outlets. In 2014, a reform of the Spanish intellectual property law established that firms posting links and excerpts of news stories have to pay a compulsory link fee to the original publishers. In December 2014, Google reacted by shutting down Google News in Spain.

In the debate on news aggregators, content producers argue that news aggregators make money by stealing high-quality content. Since this money is pulled out of content producers' pockets, they have less incentive to produce high-quality content. For instance, according to Rupert Murdoch, chairman of News Corp.:

> "When this work is misappropriated without regard to the investment made, it destroys the economics of producing high-quality content. The truth is that the 'aggregators' need news organizations. Without content to transmit, all our flat-screen TVs, computers, cell phones, iPhones and blackberries, would be blank slates. (Murdoch 2009, 13)."

On the other hand, news aggregators argue that aggregation drives profitable traffic to news sites themselves. In a response to the FTC report (2010), Google (2010) claimed to send more than four billion clicks per month to news publishers via Google Search, Google News, and other products. Google's claim is that each click – each visit – provides publishers with an opportunity to show ads, register users, charge for access to content, and so forth.

In this survey, I review empirical and theoretical articles on news aggregators. The empirical articles aim at quantifying whether news aggregators steal traffic from newspapers or help them to receive more traffic. In other words, they study which effect dominates between the two opposite effects, i.e. *the business-stealing effect* and *the readership-expansion effect* which I introduce in Section 2. The theoretical articles aim at identifying different channels through

which news aggregators affect profits of newspapers in order to analyze how news aggregators affect newspapers' incentive to invest in quality.

There exist a variety of news aggregators. Some, like Huffington Post, use editorial staff, while others, like Google News, use an algorithm to find high quality content. After finding high quality articles, each aggregator posts them on its site. This, however, can be done in different ways. Some, like Yahoo! News, post the whole article on their site, with no link to the original content. Usually, this is because the aggregator pays the newspaper for that content and hence has the right to publish it. In 2006, Yahoo! signed an agreement with Newspaper Consortium[2] to use their content. Others, like Google News, show the title and a short summary and provide a link to the original article. These two types of aggregators bring revenue to newspapers in different ways: the first by buying a content license, and the second by sending traffic to newspaper sites. This is why Yahoo! News has kept its service in Spain while Google News has been shutdown in Spain.

The survey is organized as follows. In Section 2, I present a simple theory, which provides a framework to understand the empirical findings reviewed in Section 3. I review theoretical studies in Section 4. Section 5 provides concluding remarks with some questions for future research.

2 A simple theoretical framework

Let me start by providing a simple theory based on Jeon and Nasr (2016) that captures the two opposite effects of news aggregators, the business-stealing effect and the readership-expansion effect. We consider two (major) newspapers and one aggregator and study their competition on the Internet. Suppose that the two newspapers compete on the Hotelling model. The two newspapers are located at the extreme points of a line of length one: newspaper 1 (2) on the left (right) extreme point. The line represents ideological differentiation (Mullainathan and Shleifer 2005 and Gentzkow and Shapiro 2011) and the insights would hold even if the two newspapers' locations are not extreme. A mass one of consumers are uniformly distributed over the line. We assume that consumers single-home, which means that without (with) the aggregator, a consumer consumes only one between the two newspapers (among the two newspapers and the aggregator).

We assume for expositional convenience that there is a continuum of topics which each newspaper covers. Let S be the set of topics. A topic can be about an election, an earthquake, a sport event, the climate change etc. On each given topic, a newspaper can provide either high

[2] http://www.npconsortium.com/
"Is Yahoo a Better Friend to Newspapers Than Google?", New York Times, 8 Apr. 2009

or low quality content. So the strategy of newspaper i, with $i \in \{1,2\}$, is a subset of topics $s_i \in S$ which it covers with high quality. Let $\mu(s)$ represent the measure of any set $s \in S$. Without loss of generality, assume $\mu(S) = 1$. Then, $\mu(s_i)$ represents the average quality of newspaper i. In addition to this vertical dimension of strategy, there is an horizontal dimension of strategy. Namely, given $0 < \mu(s_1), \mu(s_2) \leq 1/2$, for newspaper $i \in \{1,2\}$, if i chooses s_i such that $s_i \cap s_j = \emptyset$, we say that i uses *the maximum differentiation strategy*. If i chooses s_i such that $\mu(s_1 \cap s_2) = \min(\mu(s_1), \mu(s_2))$, then we say that i uses *the minimum differentiation strategy*.

Let $u_0 > 0$ represent a consumer's utility from reading the home page (or the landing page) of a newspaper. u_0 is assumed to be large enough to make all consumers consume a newspaper or the aggregator. The home page provides links to articles with their titles and excerpts. Consumers are assumed to click a link only if the article is of high quality. Let $\mu\Delta u > 0$ represent the utility increase (net of attention cost) a consumer experiences from reading measure μ articles of high quality. If the quality of an article is low, no consumer reads it. Then, the utility that a consumer located at x obtains from consuming newspaper 1 or 2 is given by

$$U^1(x) = u_0 + \mu(s_1)\Delta u - xt; \tag{1}$$

$$U^2(x) = u_0 + \mu(s_2)\Delta u - (1-x)t, \tag{2}$$

where $t > 0$ is the transportation cost parameter and xt (or $(1-x)t$) represents the cost of imperfect match in terms of ideological preferences.

In the absence of the aggregator, given $(\mu(s_1), \mu(s_2))$, the market share of newspaper i is determined by

$$\alpha_i^N = \frac{1}{2} + \frac{\Delta u}{t}(\mu(s_i) - \mu(s_j))$$

where the superscript N means no aggregator and $i, j = 1, 2$ and $i \neq j$.

We consider free newspapers which make revenue from advertising. The advertising revenue of a newspaper is assumed to be proportionate to the attention that consumers spend on the newspaper: but a given unit of attention spent on a home page may generate a larger (or lower) revenue than the same unit of attention spent on individual articles. The advertising revenue generated by a consumer's consumption of a home page is normalized to one. We assume that if a consumer consumes μ measure of high quality articles, it generates an advertising revenue of $\delta\mu$. Therefore, newspaper i's profit without the aggregator is given by

$$\pi_i^N = \alpha_i^N [1 + \mu(s_i)\delta] - c(\mu(s_i)), \tag{3}$$

where $c(\cdot)$ is the cost of producing high quality articles and is increasing and convex.

We model an aggregator along the lines of Google News in that the aggregator provides only a home page without having its own original articles. For each topic, the aggregator chooses one article and publishes its title and its excerpts (called also, snippets) with a link to the original article. We assume that the aggregator chooses the highest quality article for each topic and that if both newspapers produce the same quality articles on a given topic, it chooses one of them with an equal probability.

A consumer who reads the aggregator's home page obtains a utility of $u_0 + u_T$ where $u_T > 0$ is the utility from the aggregation of content from third-parties, i.e., numerous small news sites different from newspaper 1 and 2. In addition, she clicks on the link of each high quality article and spends attention on the newspaper site to which she is directed. The consumer is assumed not to click on the links to low quality articles. Therefore, using the aggregator over her preferred newspaper allows a consumer to access more high quality content, at a higher cost of preference mismatch.

More precisely, consider a consumer with location $x < 1/2$. Then, we have

$$U^{Agg}(x) - U^1(x) - u_T = \underbrace{(\mu(s_1 \cup s_2) - \mu(s_1))\triangle u}_{\text{Benefit from higher quality}} - \underbrace{t(\frac{1}{2} - x)\,(1 + \mu(s_2) - \mu(s_1))}_{\text{Cost from higher preference mismatch}}, \qquad (4)$$

where $U^{Agg}(x)$ represents the utility that a consumer located at x obtains from using the aggregator (see the appendix for the explicit formula). The benefit of using the aggregator instead of newspaper 1 is composed of u_T and the other terms. The term $(\mu(s_1 \cup s_2) - \mu(s_1))\triangle u$ represents surplus increase from consuming more high quality content. This benefit comes with the cost of greater preference mismatch since, for a consumer with location $x < 1/2$, the favorite newspaper is 1; the last term in (4) always has a negative sign for $x < 1/2$.

Jeon and Nasr (2016) assume that producing high-quality articles is costly such that each newspaper i chooses $\mu(s_i) \leq 1/2$. They show that under reasonable assumptions, the maximum differentiation strategy is a dominant strategy for each newspaper as this strategy allows each newspaper to maximize the traffic directed from the aggregator to its individual articles.

Under the maximum differentiation strategy, given (s_1, s_2), newspaper i's profit is given by:

$$\pi_i^A(s_i|\max) = \alpha_i^A\,[1 + \mu(s_i)\delta] + \delta\mu(s_i)(1 - \alpha_i^A - \alpha_j^A) - c\mu(s_i)^2; \qquad (5)$$

where $j \in \{1, 2\}$, $j \neq i$ and the superscript A means that the aggregator is present. When compared with the profit without the aggregator (3), the term in the middle of the R.H.S. of (5) is new and represents the advertising revenue from the consumers directed by the aggregator to i's articles as $(1 - \alpha_i^A - \alpha_j^A)$ represents the aggregator's share in home page traffic.

Both the business-stealing effect and the readership-expansion effect are defined in terms of traffic. *The business-stealing effect* captures the reduction in the traffic to the home pages of the newspapers which results as some consumers read the home page of the aggregator and is given by $\alpha_i^A - \alpha_i^N (= -\Delta\alpha_i) < 0$. *The readership-expansion effect* captures the traffic increase to high quality articles which result as high quality articles of a newspaper can reach not only its loyal readers, but also those using the aggregator. The latter includes consumers who would read the rival newspaper if there were no news aggregator. If traffic is measured in terms of the number of pageviews, $\Delta\alpha_j \mu(s_i) > 0$ represents the readership-expansion effect to newspaper i where $\Delta\alpha_j$ represents the consumers who switch from newspaper $j\,(\neq i)$ to the aggregator in terms of home page consumption. Note that in Jeon and Nasr (2016), the total number of consumers is fixed and hence the readership-expansion effect means that consumers on average read more articles. However, one can also consider another kind of readership-expansion effect, which means that the aggregator increases the number of consumers who read news (see Dellarocas, Katona and Rand (2013) in Section 4).

The empirical papers reviewed in Section 3 try to study which of the two effects dominates. In addition, they try to see how the effects interact with the characteristics of newspapers. For instance, within the simple framework presented in this section, the small (unknown) newspapers whose content is aggregated by the aggregator and is captured by u^T for sure gain from the presence of the aggregator as they attract no traffic in its absence. Although this extreme result is obvious and is driven by assumption, I will provide some empirical evidence for a more generalized version of the result in the next section.

In the end, what matters for each newspaper is how its profit is affected. Given (s_1, s_2), the effect of the aggregator on newspaper i's profit is given by

$$\begin{aligned} \pi_i^A(s_i|\max) - \pi_i^N(s_i) \quad = \quad & \underbrace{-\Delta\alpha_i}_{\text{Business-stealing effect (-)}} \\ + \quad & \delta * \underbrace{\Delta\alpha_j \mu(s_i)}_{\text{Readership-expansion effect (+)}} \end{aligned}$$

where $\delta > 0$ captures the monetary value of a unit traffic to articles relative to that of a unit traffic to home page in terms of advertising revenue and is typically smaller than one. Hence, even if the total effect on the traffic of newspaper i is positive, the total effect on its profit can be negative.

The theoretical papers reviewed in Section 4 investigate how the aggregator affects each newspaper's incentive to invest in quality, which can be studied only after one understands how the aggregator affects each newspaper's profit for given quality choices. However, as it is hard to

find data on profits, the empirical papers seldom study the effect on profits. Therefore, there is a gap between theoretical papers and empirical papers. This is why I review first the empirical papers before reviewing theoretical papers.

Finally, each newspaper can employ another strategy, which consists in opting out from the news aggregator. Jeon and Nasr (2016) show that if an increase in the third-party content indexed by the aggregator generates more traffic to each newspaper, then each newspaper has no incentive to opt out. Opting out implies losing traffic from the aggregator. This adverse effect of the opting out should increase with the market share of the aggregator, which in turn increases with the amount of the third-party content indexed by the aggregator (represented by u_T). I will review below an empirical study of opting in/out decisions in Germany by Calzada and Gil (2017).

In summary, the simple theoretical framework generates the following questions to be answered. On the empirical side, we have:

- Which effect dominates between the business-stealing effect and the readership-expansion effect?
- How do the effects vary depending on the characteristics of newspapers?
- Does a newspaper have an incentive to opt-out?

On the theoretical side, we have:

- How does the aggregator affect each newspaper's profit?
- How does the aggregator affect each newspaper's incentive to invest in quality?

3 Empirical studies of news aggregators

In this section, I review empirical studies of news aggregators. I start by reviewing papers that study Google News: Google News shutdown in Spain, Google News opt-in policy in Germany and other events related to Google News. And then, I review a paper that studies Facebook as a news aggregator and an experimental paper studying attention allocation between a news aggregator and original articles. Finally, I review a paper studying news slants of aggregators.

Before reviewing the empirical results, let me point out the fact that all empirical papers find that the business-stealing effect is dominated by the readership-expansion effect.

3.1 Empirical studies of Google News

3.1.1 Events regarding Google News in Spain and in Germany[3]

Let me first describe the events regarding Google News in Spain and in Germany. On January 1, 2014, because of the lobbying of the publishers' association AEDE, the Spanish Parliament passed a reform of the law of intellectual property right. The new law established that online outlets posting links and excerpts of news articles originated elsewhere must pay a link fee to the original publishers. A unique feature of the Spanish regulation is that link fees are mandatory: publishers cannot refuse to receive a fee from news aggregators as the link fee must be collected by a private entity called CEDRO which will redistribute the revenues to the news outlets. (Calzada and Gil, 2017).

Although the implementation of the law was subject to a lot of uncertainty, on December 16, 2014, Google shut down the Spanish edition of Google News. The shutdown had an important and immediate impact on the Spanish news market such that the publishers in AEDE urged the government to negotiate a solution with Google. Some large publishers in AEDE even announced that they would renounce any compensation payment for sharing content with news aggregators.

The German Parliament passed an addendum to the copyright law on March 1, 2013. It granted publishers the right to charge search engines and other online aggregators for reproducing their content beyond headlines and short excerpts but also allowed free use of text in links and brief excerpts. The main differences of the German regulation with respect to the Spanish one are: link fees have to be negotiated between the parties and brief excerpts are not affected by the regulation.

In June 2014, VG Media, a consortium of more than 200 publishers, sued Google and other news aggregators for displaying excerpts and preview images along with the links to their news articles. On October 2, 2014, the German edition of Google News announced the change from an opt-out to an opt-in system: those publishers who want to be indexed by Google must explicitly grant permission and renounce any type of compensation. Publishers associated with VG Media decided not to opt in. A leading publisher in the group of VG Media was Axel Springer, which asked VG Media not to issue free licenses for its websites. On October 23, 2014, Google News and other German news aggregators stopped showing large excerpts, video and images from the publishers that did not opt in. The change significantly reduced traffic to VG Media news sites that on November 5, 2014, Axel Springer and other VG Media publishers decided to opt in.

[3]The description of the events is mainly based on Calzada and Gil (2017).

3.1.2 Google News shutdown in Spain

Athey, Mobius and Pal (2017) study Google News shutdown in Spain by using browser log data of desktop users. Control users are chosen to have identical news consumption patterns as treatment users after the shutdown. Before the shutdown, treatment users used Google News whereas control users did not. They estimate the effect of the shutdown by comparing the news consumption of treatment and control users before the shutdown.

They find that treatment users have 19.7 percent higher consumption in terms of pageviews in the pre-shutdown period compared to control users, including their consumption of the Google News home page. This volume change comes from two sources: Google News users consume 28.8 percent more articles but 8.5 percent fewer landing pages (omitting the Google News landing page). Hence, the readership-expansion effect dominates the business-stealing effect: in other words, Google News is a complement to overall news reading.

Athey, Mobius and Pal (2017) also break out the volume effect by distinguishing top 20 outlets from below top 20 outlets. They find that the effect of Google News on the top 20 outlets is not statistically different from zero as the positive effect on articles cancels out the negative effect on landing pages. By contrast, smaller outlets gain as much as 26.3 percent from the presence of Google News: the landing page traffic is unaffected but article pageviews increase by 44.6 percent. They further decompose the volume effect according to news characteristics. They find that post-shutdown, treatment users read less breaking news, hard news and news that is not well covered on their favorite news publishers.

Calzada and Gil (2017) use data at the domain level from news outlets in Spain, France and Germany. Hence, their data are complementary to the data used by Athey, Mobius and Pal (2017). They study the Google News shutdown in Spain by using French outlets as a control group. They find that the shutdown reduced on average the number of daily visits to Spanish outlets by 14%. This finding is consistent with that of Athey, Mobius and Pal (2017). Calzada and Gil (2017) find that this effect varies from no effect (business outlets), medium size effect (national and regional news outlets) and large effect (sports and Catalan language news outlets). They also find that the impact was larger in lower-ranked domains and domains with lower proportion of international visitors, which is quite consistent with the finding of Athey, Mobius and Pal (2017).

Calzada and Gil (2017) also study how the impact of the shutdown evolved over time until reaching a steady state. They find that the effect across all news outlets stabilizes around 13.8% seven weeks after the shutdown. They also try to decompose the total effect into a market-expansion effect and a substitution effect by studying the impact of the shutdown on the outlets' traffic sources. They find that the percentage of search visits decreased whereas the percentage

of direct visits increased. They interpret the former as an evidence of the market-expansion effect and the latter as an evidence of the substitution effect.

Whereas Athey, Mobius and Pal (2017) limit attention to the impact of the shutdown on traffic, Calzada and Gil (2017) study the impact on advertisement revenues as well. They focus on the online editions and separate those outlets that are above the median advertising revenues from those that are below the median. They find that after the shutdown, the daily revenues of above median outlets decreased significantly relative to those below the median. When they study the sources of this decrease in revenue, they find decreases in advertising intensity, revenue per advertiser and revenue per unit of advertising intensity. However, it seems that their finding on advertising revenue is hard to reconcile with the finding of Athey, Mobius and Pal (2017) that the shutdown did not change the overall traffic but increased the traffic to landing pages for top 20 outlets whereas it reduced the overall traffic without affecting the traffic to landing pages for below top 20 outlets. Suppose that top 20 outlets have advertising revenues above the median and that landing pages are more important than individual articles in terms of advertising revenue. Then, the shutdown should increase advertising revenue for top 20 outlets while reducing it for below top 20 outlets, which is opposite to the finding of Calzada and Gil. It would be nice to have a better understanding of the impact on advertising revenue.

3.1.3 Google News opt-in Policy in Germany

After the introduction of the opt-in policy in Germany, Google News continued to index all news outlets but could complement the links with long excerpts and images only from those outlets that had opted in. Calzada and Gil (2017) study the impact of VG Media's decision not to opt in. They find a negative but non-significant effect of the opt-out decision on the visits to the VG Media outlets relative to all other German outlets that did not belong to VG Media. But when they focus on the 10 outlets Axel Springer controlled, which are part of the VG media outlets, they find a negative and significant reduction in daily visits of around 8% in Axel Springer outlets relative to all other German outlets. This explains the fact that Axel Springer and the other VG Media outlets that had initially stayed out decided to opt in. The scenario that some outlets opt in while others do not is hard to be sustained as an equilibrium as the latter has competitive disadvantage because the traffic they would receive from Google News with opt-in is likely to be directed to the former. What happened in Germany is consistent with the prediction of Jeon and Nasr (2016).

3.1.4 Other studies on Google News

Chiou and Tucker (2017) study the removal of the content of Associated Press (AP) from Google News that occurred from December 23, 2009 until sometime in February 2010. They use Yahoo! News as a control since it continued to host the AP content. They study whether the removal leads to a shift away from Google News and whether traffic to news sites from Google News falls after the removal. They find that the removal does not affect the traffic to Google News. In the case of the effect on downstream news sites, they find that the odds of visiting a news site on Google News relative to a non-news site on Google News decreased by 28% compared to the odds of visiting a news site on Yahoo! News relative to a non-news site on Yahoo! News. This result suggests that the presence of AP articles in Google News prompted users to seek further information at news sites.

One striking feature of how AP content was featured on Google News is that in general quite a large amount of news content was displayed rather than merely a snippet. In light of this, the result that Google News increases traffic to downstream news sites is surprising. It is even more surprising in view of the finding of Dellarocas et al. (2016) that a longer snippet reduces the probability of clicking on the link (see Section 3.3).

Athey and Mobius (2012) study a case where Google News added local content to its home page for those users who chose to enter their location. By comparing the consumers who use this feature with controlled users, they find that users who adopted the feature increased their usage of Google News, which in turn led to additional consumption of local news. They conclude that their results support the view that news aggregators are complementary to local news outlets.

George and Hogendorn (2013) use a major redesign of Google News on June 30, 2010 that placed a permanent strip of geo-targeted local news headlines and links onto the Google News front page and find that adding geo-targeted links increases both the level and share of local news consumed online.

3.2 Facebook as a news aggregator

Sismeiro and Mahmood (2018) study how an outage of Facebook affected traffic to a news website. They have traffic data from the second largest online news website operating in a major Western European country. They take advantage of the exogenous variation in Facebook traffic created by a global Facebook outage that lasted four hours in the early morning of Monday, October 21, 2013. During the outage, it was not possible to add new posts, comment on previous posts and there were no newsfeed updates although users could access the information previously loaded on their device. Their data cover the period of October 13, 2013 to October 29, 2013 (17 days).

They observe a 38% decrease in visitors per hour and a 44% reduction in the total number of page views during the outage and a drop of about 9% of page views even after the outage. The results suggest that Facebook helps news websites to attract visitors and leads to more page requests. More importantly, they find that Facebook has an effect that goes beyond the traffic originating from clicks on the links to the news site posted on Facebook. This is because an hourly decrease of 3,956 page views originates directly from Facebook during the outage, which is substantially lower than the reduction in total page views during the outage (about 170,000 pages). More precisely, they find that during the outage hours, referrals from search engines and undefined referrals (i.e., people directly typing the URL, using their own bookmarks, or copying and pasting URLs) decreased far more than Facebook referrals. They find 29,470 fewer referrals from search and 142,020 fewer undefined referrals. This seems to be an interesting finding which shows a main difference between Facebook and Google News in terms of how each affects traffic to news sites.

However, the result may be due to the so-called "dark traffic" problem, which arises when a huge proportion of referral traffic is listed as "direct". Research from the analytics firm Chartbeat, as well as confirmation from major publishers, shows that Facebook's mobile apps are largely responsible for the swathes of dark traffic being directed toward websites.[4] Hence, most undefined referrals are likely to be originated from Facebook.

They further look at the performance of different news categories during the outage. The news categories they study include local news, sports, women issues and health. They find a reduction in traffic of all news categories during the outage. In contrast, after the outage, traffic recovery varies by category. Sports and local news see a significant increment after the outage whereas women issues and health-related sections remain below the baseline. They speculate that this difference arises because the first two categories are more time sensitive than the last two.

They also find that during the outage, a decrease in the number of home page views per user of 0.71 and an increase in the number of content page views per user of 0.52. These correspond to a reduction of 66% and an increase of 37% compared to their baselines. This suggests that Facebook introduce a selectivity bias by attracting shallower users (i.e., users who read mostly headlines from the home page and do not read many articles) to the site.

I think that the result that the Facebook outage reduced the traffic to the news site is much less surprising than the findings from the Google News shutdown in Spain as the former is about a temporary shock while the latter is about a lasting or permanent shock.

[4] See http://uk.businessinsider.com/facebook-mobile-app-responsible-for-dark-traffic-2014-12?r=US&IR=T

3.3 Experiments on attention allocation between a news aggregator and original articles

Dellarocas et al. (2016) study how readers allocate their attention between a news aggregator and the original articles it links to. They run field experiments on a Swiss news aggregator application called Newscron. The app has two client versions, an iPhone version and an iPad version. The two versions provide distinct user interfaces with different limitations and hence they conducted separate experiments on each version.

They first consider topics that have a single article in the iPhone environment and find that click-through probabilities of individual articles decrease as snippet lengths increase and that the presence of an image is also associated with lower click-through rates. Experiments with the iPad version lead to the same results. These findings suggest that click-through rates are significantly affected by snippet lengths. However, one can expect that the snippet length which is optimal for the newspapers providing original articles is shorter than the one which is optimal for the aggregator. This may provide a rationale for regulating the snippet length as is done in Germany.

They also consider topics containing two or more snippets and where exactly one snippet was clicked and study how an article's snippet length and the presence of an image affect the click-through probability in the iPhone environment. As only one snippet in a topic is clicked, this study allows them to study how competition among snippets is affected by snippet length and image. They find that having longer than average snippets has a positive effect on the choice probability and that the presence of an accompanying image increases a snippet's within group choice probability. The effect of having an image is strong and comparable to moving from second to first position on the list of related articles.

This result on click-through probability in a competitive environment is consistent with the finding of Calzada and Gil (2017) that those newspapers which opted out (and hence whose articles had very short snippets on Google News) suffered from traffic loss. Because of this competitive disadvantage, they ended up opting in. The same analogy can be made to the Swiss news aggregator in Dellarocas et al. (2016): even if newspapers may collectively prefer short snippets, short snippets may not be sustained as an equilibrium when each news site can deviate by allowing the news aggregator to show longer snippet.

3.4 News slant of two Korean Aggregators

South Korea is unique in terms of the influence of news aggregators. In 2016, 60% of Koreans had access to news through Internet portal news aggregators while only 13% consumed news through home pages of newspapers. The two major news aggregators, Naver and Daum, each

had a share of 55.4% and 22.4% in the Internet news traffic in 2016 (Choi, 2017). The business model of Naver and Daum is similar to that of Yahoo News in that each of them pays to receive articles from a selected group of newspapers. In 2015, 59 newspapers supplied articles to both aggregators while 17 only to one aggregator and 86 (60 among them sports or entertainment newspapers) only to the other aggregator.

Choi (2017) studies news slants of the two Korean news aggregators by adopting the methodology of Gentzkow and Shapiro (2010). He has data about all news articles showed by both aggregators during 2015 and he finds that both of them exhibit almost no slant. Even if there is competition between the two Korean news aggregators, Choi (2017) finds little ideological difference between the two. This finding is very consistent with the theoretical prediction of Gabszewicz, Laussel and Sonnac (2001) that when newspapers are financed by advertising, they tend to have minimal ideological differentiation instead of the maximal differentiation, which occurs when they are financed by sales revenue (Mullainathan and Shleifer, 2005). A main difference between Choi (2017) and Gabszewicz, Laussel and Sonnac (2001) is that in Choi (2017), the slant of an aggregator is defined as the average slants of all articles shown by the aggregator, which are supplied by different newspapers which can have very strong ideological bias.

4 Theoretical studies

Most theoretical articles on news aggregators go beyond the empirical articles surveyed in Section 3 in the sense that they are not only interested in identifying different channels through which aggregators affect traffics and profits of newspapers but also interested in studying how the aggregators affect quality choices of newspapers, which is a very important question. Note also that most theoretical articles reviewed in this section consider a single-topic model whereas Jeon and Nasr (2016) consider a multi-topic model. How the results obtained in a single-topic model can be generalized to a multi-topic environment remains an open question.

In the model presented in Section 2, Jeon and Nasr (2016) study how the aggregator affects the newspapers' incentive to invest in quality. They find that depending on the value of δ, it can increase or decrease the quality since the readership-expansion effect becomes stronger as δ increases. In order to further pin down the prediction, they find a lower bound on δ from the empirical findings of Athey and Mobius (2012) and Chiou and Tucker (2017). For instance, Athey and Mobius (2012) find that after adding content from new local outlets to Google News, traffic increases not only to these new outlets but also to the old (local and non-local) outlets that have been indexed by Google News. Using the lower bound, they find that the aggregator increases the quality chosen by each newspaper. They also find that the result on quality choice is robust to introducing noise into the quality certification technology of the aggregator. However,

noise in the certification technology makes the business-stealing effect stronger relative to the readership-expansion effect, which tends to decrease newspapers' profits. This finding offers a possible explanation for newspapers' complaint against Google News: they may find Google's algorithm to select news articles too noisy, resulting in low profits for them.

Huang (2017) focuses on how a news aggregator alleviates the moral hazard of a newspaper in terms of investment in quality. In Jeon and Nasr (2016), the quality of each newspaper is known to consumers before they choose a newspaper to read. In her model, consumers do not observe the quality of a newspaper when they decide to visit its site or not. Hence, in the absence of the aggregator, the market collapses as the newspaper cannot commit to invest in quality: shirking is a dominant strategy for the newspaper. The aggregator alleviates this incentive problem as consumers can observe the quality of a newspaper by visiting the aggregator and can click on the link only if the quality is high. Depending on the degree of loyalty to the newspaper, a consumer can directly visit the newspaper or visit the newspaper indirectly by clicking the link at the site of the aggregator or visit only the aggregator. In addition, motivated by Dellarocas et al. (2016), she allows the aggregator to choose the length of snippet. She finds that the aggregator tends to choose a snippet length which is too long as it does not internalize the traffic directed to the newspaper. Therefore, it can be optimal to introduce a tax on the snippet length or a click-through subsidy. The information asymmetry problem she studies should be relevant for those newspapers with weak brand recognition. Then, the aggregator can help consumers discover interesting articles from these newspapers as is found by Athey, Mobius and Pal (2017). Regarding snippet length, it would be interesting to empirically study whether the snippet length chosen by aggregators is too long. Note that an aggregator has some incentive to limit snippet length of the articles shown at its home page for the same reasons as all newspapers limit snippet length of the articles shown at their home pages. The two major Korean news aggregators, Naver and Daum, are an extreme example since they show only one line for each article in the mobile home page. In fact, they do not even show the source of each article.[5]

While Jeon and Nasr (2016) consider homogenous consumers (but for their ideological taste), Rutt (2011) considers two types of consumers (loyal ones and searchers) and uses an all-pay auction model to study newspapers' choice of quality and prices. A loyal consumer reads only her preferred newspaper while a searcher uses an aggregator to read the highest quality one among *free* newspapers as searchers are assumed to be not willing to pay to access an article. Given the behavior of consumers, firms simultaneously decide on their price and quality investments. Firms face a trade-off in their pricing strategy between earning sales revenue from loyal consumers and

[5]For some major topics, clicking on a topic at the homepage opens a second page about the topic showing multiple articles. Even in this case, they use snippets of only one or two lines per article (but provide the sources of articles as well).

losing potential advertising revenue from searchers, which leads to a symmetric mixed strategy equilibrium. In the equilibrium, firms randomize between providing the article for free and charging for access to the article. There is a unique level of quality provided by the firms who charge for access to the article whereas there is a distribution of quality levels for articles which are free to access. He finds that as the fraction of searchers increases, the expected profit of each newspaper decreases, free newspapers choose higher quality while the rest choose lower quality. Although the results are interesting, I wonder how realistic the mixed strategy result is. The decisions regarding business models (free or paywall) and quality investments are core long-term decisions of a newspaper. For instance, the quality investment decision is strongly associated with the number of journalists to hire. I have difficulty in imagining a board taking these decisions in a random way.

Dellarocas, Katona and Rand (2013) go beyond a standard model of media and aggregators by considering competition among content sites in a link economy. They consider a single-topic model. Each content site i can produce its own content of quality q_i and also provide a link to the content of another content site say j. Content site j cannot refuse the link of i. Content site i faces the following trade-off when providing a link to content of higher quality: it increases the anchor traffic to site i but a fraction $1-\rho \in (0,1)$ of the traffic will click the link and hence will not stay at site i meaning that site i obtains no advertising revenue from that traffic. After studying the equilibrium quality choice and link decision without aggregator, they introduce an aggregator who is defined as a content site which cannot create its own content but can provide a link. The aggregator provides the link to the highest quality content. By so doing, it increases the total anchor traffic to the media ecosystem (i.e., there is a market-expansion effect) but reduces the anchor traffic to each content site. In addition, a fraction $1-\rho$ of the anchor traffic of the aggregator ends up landing at the highest quality site. If $1-\rho$ is large enough, the aggregator increases the traffic to the highest quality site while always reducing the traffic to the lowest quality site. When the content sites cannot provide links (like most newspaper sites in real world), they find that the equilibrium content quality decreases with ρ. They also consider imperfect quality certification technology of the aggregator and find that as the technology becomes more accurate, there is more competition between the content sites such that the equilibrium quality becomes higher and the profit becomes lower. This result is opposite to the finding of Jeon and Nasr (2016). It would be interesting to dig deeper into the role of the algorithm used by the aggregator.

De Cornière and Sarvary (2018) study content bundling by social media, i.e., social media shows news content together with user-generated content (UGC). In the baseline model, they consider one newspaper. They are interested in studying how the content bundling affects the profit of the newspaper and its incentive to invest in quality. UGC quality is assumed to be

exogenous. Each consumer allocates a fixed total amount of attention between news and UGC and consumers differ in terms of their demand intensity for news. In the benchmark without content bundling, consumers optimally allocate their time between social media to consume UGC and the newspaper site to consume news. In order to understand the effect of content bundling, we can consider personalized content bundling: the social media knows each consumer's type and bundles a different amount of news content depending on the type. In this case, it is optimal for the social media to propose exactly the same amount of news content each type will consume in the benchmark without content bundling. This reduces the newspaper's profit because for any news consumed on the social media, the associated advertising revenue is shared with the social media. Even though this effect tends to reduce the incentive for the newspaper to invest in quality, however, there is an opposite effect which makes the overall effect on the investment incentive ambiguous. Namely, by investing more, the newspaper can induce those consumers who spend very small amount of attention on UGC to spend their entire attention directly on the site of the newspaper. This increases the advertising revenue of the newspaper in a discrete way as the newspaper captures all advertising revenue associated with news consumption on its site. They find qualitatively the same results on the profit and the investment incentive when the social media cannot personalize content bundling.

Calzada and Ordóñez (2012) study a newspaper's reaction to the aggregator in terms of versioning (and linking) decisions in the framework of a monopolist's second-degree price discrimination. George and Hogendorn (2012) consider a model of two-sided market in which news aggregators increase multi-homing viewers. They find that the switching of a given mass of viewers from single-homing to multi-homing is likely to reduce (increase) a news outlet's advertising revenue if the outlet initially has a high (small) share of exclusive viewers.

5 Concluding remarks

A big challenge for newspapers in the Internet environment is how they can attract attention of consumers who spend their limited attention among millions of different sites. For instance, Boik, Greenstein, and Prince (2017) find that for the period of 2008-2013, total time online at the primary home device has only modestly declined and that the concentration of sites visited and time spent in long sessions has remained remarkably stable. Their finding implies that the total amount of attention that consumers spend on the Internet is more or less fixed and is concentrated on a relatively small number of anchor sites. This puts newspapers in a vulnerable situation as they become dependent on major anchor sites such as Facebook and Google (Search and News) to attract traffic to their news sites. Such trend is observed by Boik, Greenstein, and Prince (2017) as they find that the period between 2008 and 2013 saw major changes in online

category shares, with social media and video experiencing significant increases while chat and news experienced significant declines.

A major empirical finding I surveyed is that news aggregators reduce traffic to newspaper home pages while increasing traffic to individual news articles. Even if all empirical articles agree on the statement that the business-stealing effect is dominated by the readership-expansion effect, if this comes with a reduced traffic to home pages, it can have a long-term consequence that is not captured by the empirical studies. For instance, if consumers using news aggregators do not pay much attention to the sources of original articles, this can reduce newspapers' incentives to build up reputation, which would make newspapers further depend on the reputation of the aggregators such as Google or Facebook. It would be interesting to study both empirically and theoretically how competition for attention among newspapers is done on a anchor site (such as news aggregators or social media) and how such competition is different from the competition among printed newspapers before the Internet. Is the competition on a news aggregator healthier than the competition among printed newspapers? One can further study how competition on a anchor site is affected by the site's algorithm (such as Facebook's newsfeed algorithm) and how the profit-maximizing algorithm differs from the welfare-maximizing one. For instance, Facebook recently announced an algorithm change which will de-prioritize videos, photos, and posts shared by businesses and media outlets in favor of content produced by a user's friends and family.

Google and Facebook launched respectively AMP (accelerated mobile page) project in 2016 and Instant Articles project in 2015. AMP and Instant articles host articles respectively on Google and Facebook for fast-loading of news in the mobile environment. It seems that Google's AMP project has received much wider support from publishers than Facebook's Instant Articles.[6] In fact, more than half of Facebook's launch partners on Instant Articles, including major newspapers such as New York Times and Washington Post, appear to have abandoned the format (Brown, 2018). The different outcomes may have to do with different business models embraced by Google and Facebook; while Google is attached to open web, Facebook is a closed system with the goal of getting people to spend more time inside its app in order to show more ads. However, even with the success of the AMP project, there are concerns about increasing dependence of media companies on the major platforms through "mediated advertising arrangements with accidentally enormous middlemen apps that have no special interest in publishing beyond value extraction through advertising (Herrman, 2015)".

Note that the decrease in traffic to newspaper home pages relative to traffic to individual news articles is a more general phenomenon, which is called the unbundling of journalism:

"It is a world of fragments, filtered by code and delivered on demand. For news organizations,

[6] https://digiday.com/media/how-google-amp-won-over-facebook/

said Cory Haik, senior editor for digital news at The Washington Post, the shift represents "the great unbundling" of journalism. Just as the music industry has moved largely from selling albums to songs bought instantly online, publishers are increasingly reaching readers through individual pieces rather than complete editions of newspapers or magazines."[7]

How will the unbundling of journalism affect the incentive to produce high quality journalism? A popular view is that the traditional way of selling bundle of news developed a cross-subsidy system which allowed to finance costly investigative journalism. For instance, according to a report prepared for FCC,

"A cross-subsidy system had developed: a consumer who bought the newspaper for the box scores was helping to pay the salary of the city hall reporter. Today, a reader can get a mobile app that provides only box scores (with second-by-second updates!). The bundle is broken—and so is the cross-subsidy. (Waldman et al., 2011, p. 13)."

Does the end of the cross-subsidy system imply the end of investigative journalism?

6 References

Athey, S. and M. Mobius (2012): "The Impact of News Aggregators on Internet News Consumption: The Case of Localization," Working Paper.

Athey, S., M. Mobius, and J. Pal (2017): "The Impact of Aggregators on Internet News Consumption," Working Paper.

Andre Boik, Shane Greenstein, and Jeffrey Prince (2017): "The Empirical Economics of Online Attention," Working Paper

Brown, P. (2018): "More Than Half of Facebook Instant Articles Partners May Have Abandoned It," February 2, Columbia Journalism Review

Calzada, J., and G. Ordóñez (2012): "Competition in the News Industry: Fighting Aggregators with Versions and Links," Working Paper 12-22, NET Institute.

Calzada, J., and R. Gil (2017): "What Do News Aggregators Do? Evidence from Google News in Spain and in Germany," Working Paper

Choi, D. O. (2017): "Internet Portal Competition and Economic Incentives to Tailor News Slant," Korean Journal of Industrial Organization, 25(2), 1-40 (in Korean)

Chiou, L., and C. Tucker (2017): "Content Aggregation by Platforms: The Case of the News Media," Journal of Economics and Management Strategy, 26: 782-805

De Cornière, A. and M. Sarvary (2018): "Social Media and News: Attention Capture via Content Bundling," Mimeo

[7]New York Times, "How Facebook Is Changing the Way Its Users Consume Journalism" by Ravi Somaiya, October 26, 2014.

Dellarocas, C., Z. Katona, and W. Rand (2013): "Media, Aggregators, and the Link Economy: Strategic Hyperlink Formation in Content Networks," Management Science, 59(10), 2360–2379.

Dellarocas, C., J. Sutanto, M. Calin and E. Palme (2016): "Attention Allocation in Information-Rich Environments: The Case of News Aggregators," Management Science, 62(9), 2543-2562.

FTC (2010): "Potential Policy Recommendations To Support The Reinvention Of Journalism," Discussion Draft, Federal Trade Commission.

Gabszewicz, J. J., D. Laussel, and N. Sonnac (2001): "Press Advertising and the Ascent of the 'Pensée Unique'," European Economic Review, 45(4-6), 641 – 651.

Gentzkow, M., and J. M. Shapiro (2010): "What Drives Media Slant? Evidence from U.S. Daily Newspapers," Econometrica 78(1), 2010, 35-71

Gentzkow, M., and J. M. Shapiro (2011): "Ideological Segregation Online and Offline," The Quarterly Journal of Economics, 126(4), 1799–1839.

George, L., and C. Hogendorn (2012): "Aggregators, Search and the Economics of New Media Institutions," Information Economics and Policy, 24(1), 40 – 51.

George, L., and C. Hogendorn (2013): "Local News Online: Aggregators, Geo-Targeting and the Market for Local News," Working Paper.

Google (2010): "Comments on Federal Trade Commission's News Media Workshop and Staff Discussion Draft on "Potential Policy Recommendations to Support the Reinvention of Journalism"," Discussion Paper.

Herrman, J. (2015): "The Next Internet is TV," February 5, the AWL. https://www.theawl.com/2015/02/th next-internet-is-tv/

Huang, J. (2017): "Should Google Profit Like a Taxi Driver?," Working Paper

Jeon, D.-S. and N. Nasr (2016): "News Aggregators and Competition Among Newspapers on the Internet," American Economic Journal: Microeconomics, 8(4), 91-114

Korea Press Foundation (2017): Digital News Report 2017 South Korea.

Mullainathan, S., and A. Shleifer (2005): "The Market for News," American Economic Review, 95(4), 1031–1053.

Murdoch, R. (2009): "From Town Crier to Bloggers: How Will Journalism Survive the Internet Age?," Speech Before the Federal Trade Commission's Workshop.

Outsell (2009): "News Users," Discussion Paper.

Pew Research Center (2017): State of News Media, www.pewresearch.org

Reuters Institute (2015): "Digital News Report," Discussion Paper, University of Oxford.

Rutt, J. (2011): "Aggregators and the News Industry: Charging for Access to Content," Working Paper 11-19, NET Institute.

Sismeiro C. and A. Mahmood (2018): "Competitive vs. Complementary Effects in Online Social Networks and News Consumption: A Natural Experiment," Management Science, Articles Published in Advance

Waldman S. and the Working Group on Information Needs of Communities (2011): The Information Needs of Communities: The Changing Media Landscape in a Broadband Age, Federal Communications Commission

7 Appendix

Given (s_1, s_2), the utility that a consumer with location x obtains from using the aggregator is given by:

$$\begin{aligned} U^{Agg}(x) &= u_0 + u_T + \mu(s_1 \cup s_2)\Delta u \\ &\quad - \left(\mu(s_1 - s_2) + \frac{1}{2}\left[\mu(s_1 \cap s_2) + (1 - \mu(s_1 \cup s_2)\right]\right) xt \\ &\quad - \left(\mu(s_2 - s_1) + \frac{1}{2}\left[\mu(s_1 \cap s_2) + (1 - \mu(s_1 \cup s_2)\right]\right) (1 - x)t, \end{aligned}$$

where $s_1 - s_2$ means $s_1 \cap s_2^c$. $u_0 + u_T + \mu(s_1 \cup s_2)\Delta u$ represents utility from reading gross of the transportation cost. The transportation cost depends on the composition of the articles covered by the aggregator, and is equal to the measure of articles from newspaper 1 multiplied by xt plus the measure of articles from 2 multiplied by $(1 - x)t$.