Should Google Charge Like a Taxi Driver?*

Jin Huang

Preliminary and Incomplete

Abstract

This paper models the interplay among three agents: a news aggregator that chooses the proportion of news to display on its own website, and the amount of advertisements it exposes readers to; a newspaper that chooses the level of effort exerted to produce news; and heterogeneous readers who choose whether to go to the newspaper directly or through the news aggregator or not at all, depending on their own loyalty to the newspaper. It shows that the presence of a news aggregator can moderate the moral hazard problem of the newspaper producing low quality news. However a long snippet on the news aggregator’s website discourages the newspaper from exerting efforts. This paper also provides a microfoundation for the heated policy debate today on news aggregators, and it finds regulators should tax the news aggregator when the price of newspaper is low, and vice versa.

1 Introduction

1.1 News Aggregator

With the rapid rise of the Internet and other technological advancements at the beginning of the 21st century, people nowadays have begun to do numerous activities through electronic means, one of which is reading news content online. This has led to a transition from a paper-based media industry to an Internet-based one. Same as the paper-based, one of the main sources for revenues of news providers is online advertising. However as reported in the State of the News Media 2013\textsuperscript{1}, the growth of online advertising cannot offset the decline of print ads in newspapers. In France, not a single national newspaper is profitable, despite around €1.2 billion in direct and indirect government subsidies\textsuperscript{2}.

The unfavorable circumstances for newspaper industry arouse numerous debates on newspapers’ new competitors: News aggregators. News aggregators are services that bring together headlines from multiple news sources all into one place. Most news aggregators are crawler-based: they use crawlers\textsuperscript{3} to find pages, and they have algorithms to assemble news together to

\*Special thanks go to Gerard Llobet. Also I would like to thank Javier Suarez, Guillermo Caruana, Rafael Repullo and all the audiences at Madrid Microeconomics Graduate Workshop. All remaining errors and omissions are my own.

\textsuperscript{1}It is an annual report on American Journalism.

\textsuperscript{2}“Newspapers versus Google: Taxing times”. The Economist, Nov 10th 2012

\textsuperscript{3}A Web crawler is an Internet bot that systematically browses the World Wide Web, typically for the purpose of Web indexing.
either effectively make a custom newspaper for its users (like Google News’ homepage), or to work as search engines to find pages that match users queries and rank the news according to the algorithms. According to the investigation of Google News SEO Ranking Factors 2011, the important Google News Ranking Factors contain the level of authority, social sharing (Facebook, Twitter), uniqueness and so on. To some extent the ranking is an objective signal of the quality of the news or the matching extent with what a reader is searching for. The part of the news that news aggregators show on their own websites contains the headline and sometimes a short summary along with thumbnail images. Usually, publishers can opt-out of inclusion at any time.\footnote{They are even given more freedom, for example since 2009, Google has introduced robots.txt code to allow publishers to insert into their sites to tell Google’s crawlers to crawl a page only for search, only for news, for neither or for both.}

Generally, there are two business models of news aggregators. The first model, represented by Google News\textsuperscript{5}, is one where a news aggregator only provides links without including advertisements on its webpage; however, it attempts to attract users to other services like Gmail or Google+, where they will be exposed by finely targeted advertisements. The second model, represented by Yahoo! News, is one where the aggregator pays newspapers to include original content on its own website instead of setting a link to the information source, with advertisements displayed along. This paper only focuses on the business model of Google News.

1.2 Oppositions

Since 2007, when a Belgian court ruled that Google did not have the right to display the lead paragraph from French-language Belgian news sources, there have been successive disputes against Google. In October 2012, newspapers accounting for 90% of the circulation in Brazil

\footnote{\textsuperscript{5}Bing!News, powered by Microsoft Corporation, is another news aggregator who has the same mechanism as Google News.}
(all 154 member newspapers in Brazil’s National Association of Newspapers) abandoned Google News when they decided to deny it the ability to repost headlines and the first few sentences of articles from their websites. On January 31st 2013, after three months of tense negotiations, Google and the French press hammered a deal: Google agreed to create a €60m digital publishing innovation fund to help support transformative digital publishing initiatives for French readers, and help French publishers to increase their online revenues using Google's advertising technology. Recently, the Spanish Cabinet approved a draft reform of the Intellectual Property Law, which includes a so-called Google tax on the use of fragments of information, opinion and entertainment grouped together, for example, on search engines. Those policy debates arise naturally from the two effects news aggregators bring to newspapers: “business-stealing effect” and “market-expansion effect”. Here I quote two representative statements from both sides to illustrate the two effects. For the “business-stealing effect”, Carlos Fernando Lindenberg Neto, the president of Brazil’s National Association of Newspapers, says:

“Staying with Google News was not helping us grow our digital audiences; on the contrary, by providing the first few lines of our stories to Internet users, the service reduces the chances that they will look at the entire story in our websites.”

However Google’s Public Policy Director, Marcel Leonardi claims Google News expands the market and channels a billion clicks to news sites around the world, and compares requiring Google to pay is like:

“Taxing a taxi driver for taking tourists to eat at a particular restaurant.”

1.3 Literature Review

Concerned with those disputes, this paper studies the most popular news aggregator: Google News. It aims at the following: first, modeling the fact that the presence of Google News benefits readers through ranking and displaying part of the content, while places a double-edged sword on the newspaper; second, answering how the presence of Google affects the choices of a newspaper and readers, and providing a micro-foundation for evaluating the policies on Google or newspapers. Few papers have investigated news aggregators in the economic literature. Dellarocas, Katona, and Rand (2010) study the economic implications of interrelated and strategic hyper-linking and content investments in settings where content sites compete with one another and also with aggregators, for traffic and revenue. They find that content sites can reduce competition and improve profits by forming links to each other, and aggregators have both positive and negative effects. On one hand, aggregators increases the total traffic; on the other hand, the market entry of aggregators takes away some of the revenue that would otherwise go to content sites. Besides, by placing links to only a subset of available content, aggregators further increase competitive pressure on content sites. Jeon and Esfahani (2012)

---


3
model the reading process in which a reader first reads a aggregator’s homepage and click on the issues that she wants to read more about. They study how the presence of a news aggregator affects competition among newspapers in the Internet and find that the presence of the aggregator may lead to specialized newspapers and result in an increase in the quality of newspapers and consumer surplus, with an ambiguous effect on newspapers’ profits.

There are also two interesting empirical papers on news aggregators. Chiou and Tucker (2011) exploit an unusual natural experiment: a breakdown in contract between Google and Associated Press while Yahoo! News still hosted Associated Press news content. By using a difference-in-differences estimation, they find that after the removal, relative to Yahoo users, Google users were less likely to visit other news websites. Their results imply that the presence of news aggregators is beneficial to newspapers. Athey and Mobius (2012) provides a case analysis where Google News added local content to its news homepage. By comparing users who adopted the localization feature to those who didn’t, they find that the former users increased subsequently their usage of Google News, which in turn led to additional consumption of local news. However though in the short run the traffic to local outlets increased, in the long run users relied more on Google and read a wider variety of outlets through Google, therefore the dispersion of user attention across outlets increased.

A recent paper by Dellarocas et al. (2013) conduct a series of field experiments, based on manipulating elements of the user interface of a Swiss mobile news aggregator. One of the key design parameters is the length of the text snippet that an aggregator provides, and they found a substitution relationship between the amount of information that aggregators offer about articles and the probability that readers read the full articles at the original news sites. This result completely align with what I am showing in my model.

Compared to the previous literature, this paper introduces two novel features: First, the endogenous choice of readers. Readers choose one out of three options depending on their own loyalty or ideological view matching to the newspaper: directly visit a newspaper site; visit Google News and click the link to the news site if it is highly ranked; visit Google News and only read the short excerpt displayed in Google News. Second, the endogenous choice of the proportion of news Google News displays and the “price” Google charges.

The rest of the paper is as follows. In section 2, I introduce the model and discuss two cases: without Google and with Google. In the subsection of with Google, I discuss three scenarios: the first best, no regulation and the second best. In section 3, I suggest and evaluate two policies that could be implemented on Google or the newspaper. Section 4 shows a numerical example of the model. Finally, I present my conclusions in section 5. Most of the mathematical derivations auxiliary to the paper are gathered in the appendix A, and some pictures of Google News and Yahoo! News are attached in the appendix B.
2 The Model

2.1 Set Up

I am modeling a typical single-issue reading process. A reader wants to read a specific issue; if he is loyal enough to the newspaper that produces news on this issue, the reader visits directly the website of the newspaper’s. If he is uncertain about the news quality in that newspaper, he searches in Google News and only clicks on the link of the high-ranked news to be directly to the original news site\(^8\). However, if Google News displays a large part of the news content instead of an excerpt on its own webpage, the reader may just end up staying in Google News without being directed to any content providers’ websites.

There are three agents: readers, a newspaper and Google News. Readers have three possible options together with an outside option: (1.) directly visit a newspaper site, (2.) visit Google News and click the link to the news site if it’s highly ranked, or (3.) visit Google News and only read the short excerpt displayed in Google News; not read the news at all. The newspaper chooses the level of effort \(e\) to exert on producing news; with probability of \(e\), the quality of news would be \(V_H\), with probability of \((1 - e)\) the quality would be \(V_L\), and \(V_H > V_L > 0\). Efforts can be thought of as the investments in recruiting highly qualified journalists. Google News chooses the levels of two variables: \(\sigma\) and \(P_G\). \(\sigma\) is the proportion of news Google News displays on its own website. \(P_G\) is the price Google News charges readers, and can be understood as the form of displaying finely targeted advertisements. Although there are no advertisements on the webpage of Google News, Google displays them in other services like Gmail or Google+.

Actually nowadays a big new fear of using Google services is to do with the privacy of users, which can be thought of as a cost as well.\(^9\) There are three exogenous parameters in this model: \(P_N\), price of the newspaper that could be interpreted as forms of fees or paywall\(^10\) or exposure to advertisements; \(d\), the nuisance factor, is used to capture the heterogeneity of readers, illustrating the loyalty or ideological view matching of a reader and a newspaper; and \(\theta\), a parameter of a newspaper, which is used to describe the efficiency of effort investment, and high \(\theta\) presents high efficiency.

The demands are pinned down after the interplay among the three parties. Denote \(T_N\) as the traffic from visiting newspaper directly without googling, \(T_{NG}\) as the traffic from visiting newspaper after searching on Google, and \(T_{GG}\) as the traffic from visiting Google and stays in Google instead of being directed to the news sites. Therefore, Google’s revenue is

\[
P_G \times (T_{GG} + T_{NG})
\]

\(^8\)The quality of news can be observed by Google users through the ranking(high ranking \(\approx\) high quality). Users can also infer the quality by reading the excerpt.


\(^10\)A paywall is a system that prevents Internet users from accessing webpage content without a paid subscription.
The newspaper's revenue, meanwhile, is:

\[ P_N \times (T_N + T_{NG}) \]

while the cost function of exerting efforts is \( \mathcal{C}(e, \theta) = \frac{e^2}{2\sigma} \). The cost function is convex in effort, and possesses the single crossing condition \( C_{e\theta} < 0 \). The condition implies that a more efficient newspaper incurs less cost for one extra unit of effort investment. Later in the numerical analysis, it will be shown that \( \theta \) is an important parameter for comparative statics.

Readers who neither go to the newspaper directly nor Google get a zero utility; those who visit the newspaper website directly can read the news at the cost of \( P_N \) and nuisance \( d \), the utility is

\[ U_N = V - P_N - xd, \text{ where } x \sim U[0, 1] \]

\( V \) represents \( V_H \) or \( V_L \); \( x \) is the heterogeneity among readers, illustrating the loyalty or ideological view matching of a consumer and a newspaper. Readers who go to Google News first and then go to the news site receive the following utility:

\[ U_{NG} = V - P_N - xd - P_G, \text{ where } x \sim U[0, 1] \]

Compared to the previous case, those readers can get assured of the quality of the news by googling but have to incur \( P_G \). For those who just stay in Google, they are informed of part of the news only at the cost of \( P_G \), so:

\[ U_{GG} = \sigma V - P_G \]

### 2.2 Timing

I introduce the timing of the game:

1. A newspaper chooses a effort level \( e \) to invest; simultaneously\(^{11}\) Google chooses \( \sigma \) and \( P_G \). Efforts are not observable and the final quality of the news is not publicly known.
2. \( \sigma \) and \( P_G \) are public information. If the newspaper anticipates there is no traffic to its site from Google, it chooses to opt out of Google.
3. Readers first choose whether to go to Google or not:
   - If they go to Google, they then decide whether to be directed to the newspaper or to stay in Google News;
   - If they don’t go to Google, they decide to visit the newspaper site directly or not.

\(^{11}\)It is reasonable to believe Google has commitment power and thus can act as a leader, while the newspaper moves as a follower. However effort investment is a long-run process, therefore it is also reasonable to assume a newspaper moves first and then Google moves sequentially. For simplicity, here I set the game as a simultaneously move game.
There are two cases that must be analyzed: the case with and without Google. A moral hazard problem will arise without Google due to asymmetric information. In contrast, with the presence of Google, the moral hazard problem can be moderated. I then compare the market equilibrium without any government intervention to the first best choice, and make inferences about the effective policies by the comparison.

2.3 Without Google

A1 (The Market for Lemons): \( V_H > P_N > V_L \)

Without Google, efforts are unobservable; hence, a newspaper will exert no effort, and produces low quality news with probability 1. A moral hazard problem arises. Readers can expect the newspaper is exerting a zero effort, hence no reader will visit/buy the newspaper under the assumption of the market for lemons\(^{12}\), the market collapses as a result.

2.4 With Google

2.4.1 The First Best

The best reading scenario for readers (also for the social planner), is that everyone reads the whole news on Google News instead of going to the newspaper. The reason is straightforward. Google News can inform readers of the quality of news and save readers from incurring a nuisance \( d \) from going to newspapers. Therefore, the first best choice of \( \sigma \) should be \( \sigma_{FB} = 1 \), and the first best effort level will be the level that maximizes social welfare:

\[
Max \ e\sigma V_H + (1 - e)\sigma V_L - \frac{e^2}{2\theta}
\]

\[
FOC \ (with \ \sigma = 1): \ e_{FB} = \theta (V_H - V_L)
\]

It is straightforward to see that the first best level of effort increases in \( \theta \) and the difference between \( V_H \) and \( V_L \). A more efficient newspaper should exert more efforts, and the level of effort should be increased to ensure that readers can have high quality news with a higher probability when \( V_H \) is really high or \( V_L \) is really low.

At last, the optimal \( P_G \) could be any number small enough to encourage all the readers to go to Google in the first place, I just set \( P_{FB}^G = 0 \).

2.4.2 No regulation

Now consider how the market operates itself without intervention from the regulator. Following backward induction, I first analyze readers’ choice.

\(^{12}\)There is a lot of literature on adverse selection, one of the most classical one is George A. Akerlof (1970): The Market for “Lemons ”:Quality Uncertainty and the Market Mechanism.
**Third Stage** In the third stage, Google has already chosen $\sigma$ and $P_G$, and the newspaper has chosen $e$. Moreover, the newspaper has decided whether to opt out of Google or not. Given that the newspaper stays in Google in the second stage, the actions of readers are analyzed as follows:

First, suppose readers are in Google; if the news is in low quality, readers will just stay in Google due to A1. However if the news is in high quality, in order to decide to be directed to the newspaper or stay in Google News, readers need to compare the utilities:

- **Go to the newspaper:**
  $$U_{NH} = V_H - xd - P_N - P_G$$

- **Stay in Google:**
  $$U_{GG} = \sigma V_H - P_G$$

Hence, readers with
$$x \leq \frac{V_H(1-\sigma)-P_N}{d} = \overline{x}^1$$
will go to the newspaper when they observe high quality news instead of staying in Google.

For those $x \leq \overline{x}^1$, when they decide whether to go directly to the newspaper or go to Google before going to the newspaper in the first place, they compare the relevant two expected quantities:

**The expected utility of going to the newspaper directly:**
$$E[U_N] = eV_H + (1 - e)V_L - xd - P_N$$

**The expected utility of going to Google first and go to the newspaper if the news is in high quality:**
$$E[U_{NHG}] = e(V_H - xd - P_N) + (1 - e)\sigma V_L - P_G$$

Readers with
$$x \leq \frac{V_L(1-\sigma)+P_G-P_N}{d} = \overline{x}^2$$
will go directly to the newspaper instead of going to Google first.

For those $x > \overline{x}^1$, they also need to decide whether to go directly to the newspaper or first go to Google. They compare:

**The expected utility of going to the newspaper directly:**
$$E[U_N] = eV_H + (1 - e)V_L - xd - P_N$$

**The expected utility of going to Google first and staying in Google no matter the news is good or bad:**
$$E[U_{GG}] = e\sigma V_H + (1 - e)\sigma V_L - P_G$$

This implies that readers with
$$x \leq \frac{eV_H(1-\sigma)+(1-e)V_L(1-\sigma)+P_G-P_N}{d} = \overline{x}^3$$
will go directly to the newspaper instead of going to Google first.

It is easy to see all of $\overline{x}^1$, $\overline{x}^2$ and $\overline{x}^3$ are decreasing in $P_N$ and $d$. The reason is obvious, since when $P_N$ and $d$ are larger, going to the newspaper becomes less appealing to readers, therefore
the numbers of people going to newspaper directly and through Google will both decrease.

**Second Stage** One step back, there are two cases. The choices in equilibrium by the newspaper and Google determines the relevant case:

**Case 1** If \( \sigma \) is very high, that is when \( \sigma > 1 - \frac{P_G}{(1-e)(V_H-V_L)} \), then \( x^2 > x^1, \ x^3 > x^1 \), Google is appropriating too much. The scenario is described in the graph:

![Graph](image1)

Figure 2: Scenario 1

In this scenario, light readers\(^ {13} \) go to Google and stay in Google, and other readers all go directly to the newspaper. Since there is no traffic to newspaper through Google, then the newspaper will opt out of Google in the second stage. The situation goes back to the case “without Google” that I characterized, where a moral hazard problem arises.

**Case 2** If \( \sigma \) is not very high, that is when \( \sigma \leq 1 - \frac{P_G}{(1-e)(V_H-V_L)} \), then \( x^2 \leq x^1, \ x^3 \leq x^1 \). The scenario is described in the graph :

![Graph](image2)

Figure 3: Scenario 2

In this case, readers are divided into three regions. Light readers will go to Google and stay in Google, satisfied with just a part of the news. Normal readers search in Google News and go to the newspaper if it produces high quality news. Heavy readers directly go to the newspaper due to their high loyalties. The demands, hence, are: \( T_{GG} = 1 - x^1; \ T_{NG} = x^1 - x^2; \ T_N = x^2 \).

**First Stage** With the demands, I solve for the equilibrium of the simultaneous game. To facilitate solving for the equilibrium, I make the following assumption.

**A2:** Light readers can always get positive utility from staying in Google, that is, \( e\sigma V_H + (1-e)\sigma V_L - P_G \geq 0 \).\(^ {14} \)

Therefore Google’s problem is a maximization problem with two constraints; the first of them is A2, while the second is to prevent the newspaper from opting out in the second stage:

\(^{13}\)“light readers” are those who have high \( x \) (close to 1), which means they have low loyalty to the newspaper or their ideological views are not matching with the newspaper. In contrast, readers with low \( x \) (close to 0) are called heavy readers.

\(^{14}\)The assumption ensures the existence of readers who only go to Google for excerpts. As will be seen later in Google’s maximization problem, Google will make the traffic from Google to the newspaper close to zero; therefore Google will make sure the assumption holds to earn positive traffic from staying in Google.
\[
\text{Max} \quad \sigma, P_G \quad P_G(1 - \overline{x}^2)
\]
\[
\text{st.} \quad \frac{P_G}{eV_H + (1 - e)V_L} \leq \sigma \leq 1 - \frac{P_G}{(1 - e)(V_H - V_L)}
\]

The newspaper's problem is:
\[
\text{Max} \quad e P_N(\overline{x}^2 + e(\overline{x}^1 - \overline{x}^2)) - \frac{e^2}{2W}
\]

I focus on the interior equilibrium\(^{15}\):
\[
\sigma^G = 1 - \frac{d + P_N}{2V_H - V_L}
\]
\[
e^N = \frac{P_N \theta (d + P_N)(V_H - V_L)}{d(2V_H - V_L)}
\]
\[
P_G^G = \frac{(1 - e^N)(d - V_L(1 - \sigma^G) + P_N)}{2}
\]

When \(P_N\) and \(d\) increase, Google is in a more advantageous position compared to the newspaper than before, therefore Google can display a smaller \(\sigma\) to attract traffic. The smaller \(\sigma\) in turn leads to a higher level of effort from the newspaper, since now Google “steals” less traffic from the newspaper; hence, the newspaper has more incentives to exert high efforts.

To ensure the existence of the interior equilibrium, it must be \(\overline{x}^1 = \frac{V_H(1 - \sigma) - P_N}{d} > 0\), so an assumption should be made to ensure \(\frac{d + P_N}{2V_H - V_L} > \frac{P_N}{V_H}\).

**A3:** The exogenous parameters satisfy \(P_N < \frac{dV_H}{V_H - V_L}\).

For example when \(d = 0\), readers become homogeneous, therefore the only equilibrium is that every reader goes to the newspaper through Google News. A3 ensures a certain level of heterogeneity among readers.

### 2.4.3 The Second Best

The second best can be defined as the situation where the social planner regulates Google to choose \(\sigma\) and \(P_G\)\(^{17}\); simultaneously, the newspaper chooses efforts.

To have an idea how to improve the no-regulation equilibrium towards the first best\(^{18}\), I compare \(e^{FB}\) and \(e^N\), the quantities are specified below:

\[
e^{FB} = \theta(V_H - V_L)
\]
\[
e^N = \frac{P_N \theta (d + P_N)(V_H - V_L)}{d(2V_H - V_L)}
\]

\(^{15}\)The mathematical derivation is in Appendix A.

\(^{16}\)With \(V_H > V_L\), \(\overline{x}^1 < 1\).

\(^{17}\)An interesting avenue for future research would be the situation where the social planner can only regulate on Google on the choice of \(\sigma\) but not \(P_G\), then it becomes a game with four agents: the social planner, Google, the newspaper and readers.

\(^{18}\)If \(e\) is observable and verifiable, then constructing a contract between Google News and the newspaper is able to implement the first best effort level by transferring newspaper an amount of \(eV_H + (1 - e)V_L\). However, it is not the case here as things will become more complex if \(e\) is not observable.
This leads to the following proposition:

**Proposition 1.** For an exogenous $P_N$, if $P_N > \bar{P}_N$, where $\bar{P}_N = \frac{P_N (d+P_N)}{2V_H-V_L}$, the regulator should increase $\sigma$ to decrease the effort exerted; if $P_N \leq \bar{P}_N$, then $e^N < e^{FB}$, the regulator should decrease $\sigma$ to increase the effort exerted; Under A3, if parameter $d$ satisfies: $d < \frac{(2V_H-V_L)(V_H-V_L)^2}{2V_H^2-V_HV_L}$, then $P_N < \bar{P}_N$ and effort under-investing is always the relevant case.

The first two parts of Proposition 1 are trivial to prove. When $P_N$ is higher, as mentioned before, Google will charge a smaller $\sigma$, which gives the newspaper more incentives to exert high efforts; moreover, now the newspaper can get more revenue from the traffic. Therefore the newspaper will increase the level of effort to either attract more people going directly to the newspaper or going to the newspaper from Google when news is good. In a word, high $P_N$ can lead to over-investing efforts. The proof of the last part is provided in Appendix A.

In the second best, the regulator acts on behalf of Google. To prevent the newspaper from opting out, the traffic from Google to the newspaper has to be positive. There are only four possible cases of the equilibrium:

In case 1, there are all three regions of readers. In case 2 there are two regions: directly go to the newspaper or go to the newspaper through Google. In case 3 there are also two regions: go to the newspaper through Google or stay in Google. In the final case, all readers go to the newspaper through Google.

**Proposition 2.** The equilibrium only lies in case 1 or 4. For small $d$ that $d \leq \frac{P_N}{V_L} - P_N$, the equilibrium may lie in case 4; meanwhile for large $d$ that $d > \frac{P_N}{V_L} - P_N$, the equilibrium lies in case 1.

The proof of Proposition 2 is provided in Appendix A. Intuitively as $d$ increases, readers incur more nuisance from going to the newspaper, which is why the regulator wants more people stay in Google, and the equilibrium moves from case 4 to case 1.

3 Policy Suggestions: The Third Best

In reality, what has been done to punish Google is a lump-sum payment from Google to newspapers. For example, in France Google created a €60m digital publishing innovation fund to
support French publishers. However in the context of this model, a lump-sum transfer is not going to be effective since it cannot distort agents’ decisions. Therefore in this section, I suggest two possible policies and compare the policy equilibrium with the no-regulation one with straightforward intuitions.

### 3.1 Tax on the Proportion of News Displayed

The policy imposes a tax $t\sigma$ on Google. Now Google’s problem becomes:

Google’s problem:

$$\max_{\sigma, P_G} P_G(1 - x^2) - t\sigma$$

subject to

$$v_H + (1 - e) v_L \leq \sigma \leq 1 - \frac{P_G}{(1 - e)(v_H - v_L)}$$

The newspaper’s problem, meanwhile, is:

$$\max_{\epsilon} P_N(x^2 + \epsilon(x^2 - x^2)) - \frac{\epsilon^2}{2\theta} + t\sigma$$

The mathematical derivation is similar to the no-regulation case. I find that if the tax $t$ is very small with $t < \frac{P_G v_L}{d}$, then Google will choose $\sigma = \bar{\sigma} = 1 - \frac{P_G}{(1 - e)(v_H - v_L)}$, the same as no regulation case; however if $t \geq \frac{P_G v_L}{d}$, then Google will be forced to choose the smallest $\sigma$, that is $\sigma = \bar{\sigma} = \frac{v_H + (1 - e) v_L}{(1 - e)(v_H - v_L)}$.

Compared to the no-regulation equilibrium, if Google is punished enough by paying for the excerpt displayed to Google users, Google will reduce the length of the excerpt and thus provides the newspaper with incentives to increase its effort level. Therefore, if initially the newspaper is under-investing in efforts, taxing Google on $\sigma$ can achieve a higher social welfare level.

### 3.2 Tax on Every Click on Google News

Now the policy requires the newspaper to pay Google $t$ per unit of traffic directed from Google, that is, the newspaper needs to transfer Google $t$ for each click on the link in Google News. The maximization problems now are:

Google’s problem:

$$\max_{\sigma, P_G} P_G(1 - x^2) + t(x^2 - x^2)$$

subject to

$$v_H + (1 - e) v_L \leq \sigma \leq 1 - \frac{P_G}{(1 - e)(v_H - v_L)}$$

Newspaper’s problem:

$$\max_{\epsilon} P_N(x^2 + \epsilon(x^2 - x^2)) - \frac{\epsilon^2}{2\theta} - t(x^2 - x^2)$$

Different levels of $t$ will have different effects on the equilibrium, there are three cases:

---

19Or, Google pays the newspaper for the click if $t$ is negative. I discuss this case as well.
1. $t \leq 0$: As mentioned in the footnote 19, when $t \leq 0$ it is Google paying the newspaper $t$ per click; In this case, Google has a high incentive to set the highest $\sigma$ to avoid traffic to the newspaper, therefore $\sigma = \bar{\sigma}$. Moreover, compared to the no regulation case, the newspaper is now motivated to reduce $e$ to increase the traffic to the newspaper through Google instead of directly to the newspaper, since the newspaper can earn the tax through the traffic directed from Google. Consequently, a loss in social welfare is expected.

2. $0 < t < \frac{P_NV_L}{\bar{e}(V_H-V_L)}$: In this case, $t$ is too small to compensate Google from choosing a low $\sigma$, hence Google still sets $\sigma = \bar{\sigma}$; However compared to the no-regulation case, the newspaper has more incentives to exert a high level of effort. The reason is that by increasing the effort level, more readers will go directly to the newspaper instead of through Google. In this way, the tax transfered to Google can be saved.

3. $t \geq \frac{P_NV_L}{\bar{e}(V_H-V_L)}$: In this case, Google is subsidized enough to set a low $\sigma$ to force readers to click on the link to read the news deeper, so $\sigma = \bar{\sigma}$. Due to the low $\sigma$ the newspaper has even more incentives to produce high quality of news compared to the second case.

To conclude, if a newspaper is under-exerting efforts, the third best can be implemented by a tax on $\sigma$ or the third case of a tax on the newspaper for every traffic directed from Google news. These two policies can improve social welfare to the highest extent by giving the highest incentives to the newspaper to produce high quality news, especially when the newspaper is efficient.

4 A Numerical Example

I set $P_N = 0.3$, $d = 0.45$, $V_H = 1$, $V_L = 0.25$. By comparing the no-regulation level of effort with the first best level, it is obvious to see that in this numerical example the newspaper is under-exerting efforts due to the low value of parameter $P_N$.

Consistent with what is expected from the previous section, if a policy requires Google to pay the newspaper $t$ per click, the newspaper will exert less efforts compared to the no-regulation case, and social welfare is reduced. In this numerical example, there is no permanent best policy. As it can be seen, when $\theta$ is small, the policy that requires a low tax on the newspaper for every click can implement the third best. However, as $\theta$ increases, that is, for a highly efficient newspaper, a tax on Google for $\sigma$ and a high tax on the newspaper for every click can improve social welfare to a higher level. Comparing these two, here the tax on Google for $\sigma$ can provide incentives to the newspaper that gives high efforts to the highest extent (in Figure 5), therefore it achieves the third best when the newspaper is highly efficient.

Lastly, it is important to point out that it is not clear if a regulator puts a higher weight on Google or on the newspaper. In my numerical example, I set equal weights for both, so that the regulator just maximizes total surplus and redistribute it at no cost.
Figure 5: $\sigma$ and $e$

Figure 6: Social Welfare
5 Conclusion

This paper studies how the presence of a news aggregator, like Google News, affects the behavior of a newspaper and readers. In the era without news aggregators, the quality of news is unobservable by the public unless readers pay for the newspaper, hence a moral hazard problem arises. With the presence of a news aggregator, readers benefit from a new option: search in the news aggregator and among all the search results read only the high quality ones. Therefore in order to attract the readers, a newspaper has to exert more efforts, which consequently leads to a moderation of the moral hazard problem. We are also able to use this simple model to provide a micro-foundation for analyzing whether Google charges both the newspaper and the readers too much or too little, and I suggest and evaluate two policies that can help improve social welfare.

This paper takes the price of the newspaper as an exogenous parameter. A possible extension in the future is to make the price endogenous, which implies either I simply allow the monopolist newspaper to choose the price variable itself or introduce competition among newspapers in this framework. However, by making $P_N$ endogenous, we need to take the competition for advertisements among newspapers and Google into consideration, since the price of a newspaper is strongly correlated with the amount of advertisements it can attract. It is not trivial to solve, though we expect that there will be no qualitative differences compared to the model here.

6 References


Appendix A

1. Mathematical derivation of the no-regulation equilibrium:

Google:

\[
\begin{align*}
\text{Max}_{\sigma, P_G} & \quad P_G (1 - \frac{P_G (1 - \sigma + \frac{P_G}{d} - P_N)}{d}) \\
\text{st.} & \quad \frac{P_G}{eV_H + (1-e)V_L} \leq \sigma \leq 1 - \frac{P_G}{(1-e)(V_H - V_L)}
\end{align*}
\]

The profit function is increasing in \(\sigma\), therefore Google will charge \(\sigma^G = 1 - \frac{P_G}{(1-e)(V_H - V_L)}\), and the interior optimal \(P_G = \frac{(1-e)[d-V_L(1-\sigma)+P_N]}{2}\).

Newspaper:

\[
\begin{align*}
\text{Max}_{e} & \quad P_N (x^2 + e(x^1 - x^2)) - \frac{e^2}{2g} = P_N eV_H (1-\sigma + (1-e)V_L(1-\sigma) + P_G - P_N) - \frac{e^2}{2g} \\
\text{FOC}(e) : & \quad P_N \frac{(1-\sigma)(V_H - V_L)}{d} = \frac{e}{g}
\end{align*}
\]

Therefore the optimal \(e^N = \frac{P_N \theta(1-\sigma)(V_H - V_L)}{d}\).

Finally solve the system of equations \(\sigma^G, P_G^G, e^N\) and get the interior equilibrium:

\[
\begin{align*}
\sigma^G &= 1 - \frac{d + P_N}{2V_H - V_L} \\
e^N &= \frac{P_N \theta(d + P_N)(V_H - V_L)}{d(2V_H - V_L)} \\
P_G^G &= \frac{(1-e)[d-V_L(1-\sigma^G)+P_N]}{2}
\end{align*}
\]

2. Proof of Proposition 1:

Proof. First solve for \(\tilde{P}_N\):

\[
\tilde{P}_N (d + \tilde{P}_N) \frac{dV_H}{dV_H - V_L} = 1 \Rightarrow \tilde{P}_N^2 + d\tilde{P}_N - d(2V_H - V_L) = 0
\]

\[
\Rightarrow \tilde{P}_N = \frac{-d + \sqrt{d^2 + 4d(2V_H - V_L)}}{2} > 0, \quad \text{the other root is negative}
\]

Under A3, \(P_N < \frac{dV_H}{V_H - V_L}\), substitute \(\frac{dV_H}{V_H - V_L}\) into the function, if:

\[
(\frac{dV_H}{V_H - V_L})^2 + d(\frac{dV_H}{V_H - V_L}) - d(2V_H - V_L) < 0
\]

\[
\Leftrightarrow d < \frac{(2V_H - V_L)(V_H - V_L)^2}{2V_H^2 - V_H V_L}
\]

Therefore if \(d < \frac{(2V_H - V_L)(V_H - V_L)^2}{2V_H^2 - V_H V_L}\), then \(P_N\) is always less than \(\tilde{P}_N\), and \(e^N\) is always less than \(e^{FB}\).
3. Proof of Proposition 2:

Proof. Firstly I prove case 2 is not the equilibrium: In case 2, with constraint \( x^1 \geq 1 \)

Regulator:
\[
\max_{\sigma, x^2} \int_0^1 (e \sigma V_H + (1 - e) V_L - exd)dx + \int_{x^2}^1 (e \sigma V_H + (1 - e) \sigma V_L - exd)dx - \frac{e^2}{2\theta}
\]

\[ \Rightarrow \sigma = 1 \]

However when \( \sigma = 1 \) then \( x^1 < 0 \), therefore the constraint \( x^1 \geq 1 \) is binding, it goes back to case 1.

Next I prove case 3 is not the equilibrium: In case 3, with constraint \( x^2 < 0 \)

Regulator:
\[
\max_{\sigma, P_G} \int_0^1 (e \sigma V_H + (1 - e) \sigma V_L - exd)dx + \int_0^{x^1} (e \sigma V_H + (1 - e) \sigma V_L)dx - \frac{e^2}{2\theta}
\]

\[ \text{FOC } (\sigma): (1 - e) V_L + \frac{d(1-\sigma)V_H+P_N}{d} e V_H = 0 \]

\[ \text{SOC } (\sigma): \frac{\sigma V_H^2}{d^2} > 0 \]

Therefore the optimal \( \sigma \) is the boundary, either \( x^1 = 0 \) where the suboptimal arises since efforts exerted by the newspaper will be close to zero due to zero traffic to the newspaper, or \( x^1 = 1 \) where it goes back to case 4.

Then in case 4, with constraint \( x^1 \geq 1 \Rightarrow \sigma \leq 1 - P_N - d \) and \( x^2 \leq 0 \Rightarrow \frac{V_L(1-\sigma)+P_N-P_N}{d} \leq 0 \):

Regulator:
\[
\max_{\sigma} \int_0^1 (e \sigma V_H + (1 - e) \sigma V_L - exd)dx - \frac{e^2}{2\theta}
\]

\[ \Rightarrow \sigma = 1 \]

Therefore \( \sigma = 1 - P_N - d \) should hold. If \( \frac{V_L(1-(1-P_N-d))-P_N}{d} \leq 0 \Leftrightarrow d \leq \frac{P_N}{V_L} - P_N \), the regulator can set a positive \( P_G \) to realize the equilibrium in this case; However, if \( \frac{V_L(1-(1-P_N-d))-P_N}{d} > 0 \Leftrightarrow d > \frac{P_N}{V_L} - P_N \), regulator can not set a \( P_G \) to get the equilibrium. \( \square \)
Appendix B

Figure 7: Google News’ Homepage

Figure 8: Directed to the Original News Source from Google News
Figure 9: Yahoo! News’ Homepage

Figure 10: Read the Whole News on Yahoo! News
Figure 11: Search on Google News