09/20

ARBEIDSNOTAT
WORKING PAPER

The role of data: A two-sided model of competition between Google and DuckDuckGo

Øystein Evenstad

SNF

Samfunns- og næringslivsforskning AS
Centre for Applied Research at NHH

SNF
SAMFUNNS- OG NÆRINGSLIVSFORSKNING AS

- er et selskap i NHH-miljøet med oppgave å initiere, organisere og utføre eksternfinansiert forskning. Norges Handelshøyskole og Stiftelsen SNF er aksjonærer. Virksomheten drives med basis i egen stab og fagmiljøene ved NHH.

SNF er ett av Norges ledende forskningsmiljø innen anvendt økonomisk-administrativ forskning, og har gode samarbeidsrelasjoner til andre forskningsmiljøer i Norge og utlandet. SNF utfører forskning og forskningsbaserte utredninger for sentrale beslutningstakere i privat og offentlig sektor. Forskningen organiseres i programmer og prosjekter av langsiktig og mer kortsiktig karakter. Alle publikasjoner er offentlig tilgjengelig.

SNF
CENTRE FOR APPLIED RESEARCH AT NHH

- is a company within the NHH group. Its objective is to initiate, organize and conduct externally financed research. The company shareholders are the Norwegian School of Economics (NHH) and the SNF Foundation. Research is carried out by SNF´s own staff as well as faculty members at NHH.

SNF is one of Norway´s leading research environment within applied economic administrative research. It has excellent working relations with other research environments in Norway as well as abroad. SNF conducts research and prepares research-based reports for major decision-makers both in the private and the public sector. Research is organized in programmes and projects on a long-term as well as a short-term basis. All our publications are publicly available.

SNF Working Paper No 09/20

The role of data:
A two-sided model of competition between Google and DuckDuckGo

by

Øystein Evenstad

SNF project no 10052

“Media Competition and Media Policy”

The project is financed by the Research Council of Norway

CENTRE FOR APPLIED RESEARCH AT NHH
BERGEN, NOVEMBER 2020
ISSN 1503-2140

Table of contents

Abstract

This thesis aims to analyze the consumer data's role as a revenue-shifting input in a two-sided competition-in-utility model between a search engine collecting personalized consumer data (Google) and one that does not (DuckDuckGo). Search engines are examples of platforms characterized by network externalities. They harvest the attention of users and resell this attention to advertisers. As such, standard market mechanisms typical of single-sided markets do not apply.

In our model, advertisers are willing to pay a higher premium for targeted search results utilizing consumer data. As such, a search engine like Google can command higher prices per sponsored link if it collects more personalized data from its users. This must be balanced against demand effects stemming from consumer aversion towards the disclosure of personal data. DuckDuckGo, a search engine that explicitly does not collect consumer data, can use investments in consumer aversion strategically to capture market shares from Google.

We find that Google has an incentive to collect consumer data and that the presence of DuckDuckGo will moderate the amount. We also find that as consumer aversion to advertisements increases, Google will choose to collect more consumer data as DuckDuckGo's incentives to respond with investments in privacy awareness are reduced.

1. Introduction

The internet is becoming the town square for the global village of tomorrow.

- *Bill Gates*

It used to be that a company's value could be measured by the number of machines, land, buildings, and other material goods it possessed. The digital economy has turned this image on its head. Gone are the brick-and-mortar days when the accumulation of tangible assets were the defining characteristics of successful companies. Today, many of the most valued companies in the world are primarily data-driven[1]. Companies such as Microsoft, Amazon, Facebook, and Google accumulate large amounts of data on their users, either as their primary purpose or as a by-product of their activities. It has been said that data is the new oil driving our digital economy.

Internet search engines[2] are prime examples of data-driven companies that have emerged on the scene in the last three decades resulting from technology developed during the digital revolution. The first search engine, Archie, was launched in the year 1990. Four years later, Yahoo was established. In 1998, the Stanford students Larry Page and Sergey Brin launched BackRub, the predecessor of Google[3]. Towards the end of the 90s, an array of internet-related start-ups emerged. Many did not survive the dot-com crash.

The technological development of the 2000s facilitated an increase in the scope and quality of internet search engines. The new millennium also sees a consolidation of market power. The economic literature generally attributes this consolidation to a natural consequence of data-enabled learning and network externalities[4], e.g., Hagiu and Wright (2020). In 2020, Google

[1] https://www.gfmag.com/global-data/economic-data/largest-companies [Last access: 30/07/20]
[2] A computer program that finds information on the internet by looking for words that one types in. https://dictionary.cambridge.org/dictionary/english/search-engine [Last access: 10/05/20]
[3] https://graphics.stanford.edu/~dk/google_name_origin.html [Last access: 10/05/20]
[4] An externality is a cost or benefit caused by a producer that is not financially incurred or received by that producer. An externality can be positive or negative and can stem from either the production or consumption of a good or a service. https://www.investopedia.com/terms/e/externality.asp [Last access: 09/07/20]

has a dominating position within the search engine market, with an astonishing 92 % market share[5].

Search engines work by using algorithms known as *search engine bots*, or *spiders*[6]. These crawl hundreds of billions of pages and add them to a *search engine index*. These indexes include all the discovered URLs, as well as relevant signals describing the content of each of the URLs. They determine the amount of content a user can access when typing in a word or a sentence into a search engine. Access to a more extensive index is, therefore, a component of the quality of a search engine.

Not only do search engines need to have an extensive database from which to show results. They also need to predict the relevance of the URLs to the particular user. If a search engine manages to show relevant URLs consistently, this increases the chances of retaining users over time (Calvino and Polo, 2020a). An essential task for a search engine is, therefore, to predict the taste of the customer. Relevant results motivate the user to keep using the same *platform*[7] and increase the price a search engine can charge to advertisers (Bajari, Chernozhukov, Hortaçsu, and Suzuki, 2018).

Search engines use statistical models known as algorithms to make predictions about user tastes and behavior. The input these algorithms feed on is *data* (Calvino and Polo, 2020a). Data comes in many shapes and forms but is generally created as a by-product of consumption. As stated by Acquisti, Taylor, and Wagman (2016, p. 3): *"The ascent of the so-called Web 2.0 (blogs, social media, online social networks) has rendered individuals no longer mere consumers of information, but public producers of often highly personal data".* These datasets are referred to as *Big Data*, characterized by high volume, high velocity, and various formats (Tucker & Wellford, 2014). The data is used both to train algorithms and to make predictions. We will broadly differentiate data by its degree of personalization[8].

[5] https://gs.statcounter.com/search-engine-market-share [Last access: 02/09/20]
[6] https://www.deepcrawl.com/knowledge/technical-seo-library/how-do-search-engines-work/ [Last access: 11/05/20]
[7] A platform is a service or business model which creates value by facilitating exchanges between two or more independent groups, usually consumers and producers. https://www.applicoinc.com/blog/what-is-a-platform-business-model/ [Last access: 23/06/20]
[8] We will thus depart From Calvino and Polo's division of data by the degree of substitutability, complementarity and returns to scale, as a division between personalized and aggregate data is more relevant for our model.

When a user of a service like Google searches for something, Google collects this information. The search history of a consumer and other information such as her IP address, location, and search duration are also collected. If the user is logged into her Google account, Google will also know her age, name, and gender. The user's personal information, combined with her search history, can be a strong predictor of the user's wants and needs, and serve to provide her with relevant search results. For a company to be able to harvest this kind of personalized data, they need to be able to track the user. This information harvest is facilitated by the use of cookies[9], and/or by forcing or encouraging the user to log into a profile when using a service.

The other type of data that we will define is non-personalized data. This form of information can be collected without the need to track the user. For example, Google uses information such as the popularity of displayed links for a particular keyword and common spelling mistakes to improve its search result quality.

Google uses a program called Google AdSense to auction keywords. This auction is an intertemporal process that takes place every time someone searches on Google or visits a site showing ads. According to Google, together, three main factors determine which ads appear, and the order of their appearance[10]. These are:

i) *The bid:* The maximum amount an advertiser is willing to pay for a click on their ad.
ii) *The quality of the ads:* As mentioned, search engines must keep the users engaged, so the relevance and quality of the particular ads are also taken into consideration when determining the outcome.
iii) *The expected impact from the ad extensions and other ad formats:* When an ad is created, the advertiser is given the option to add additional information such as contact details and links to specific sites. These details are called extensions, and if they are deemed highly relevant by Google's algorithms, it increases the chances of the ad being shown first on the result page.

Although the use of personalized data can lead the user of a search engine to experience the results as more relevant for her particular taste by tracking and targeting, concerns have been

[9] A cookie, or formally an HTTP cookie, is a package of data sent to a user's computer when she is using a website. The cookie is stored locally on your computer and sent back to the webpage, which enables it to track your movement on the internet. https://us.norton.com/internetsecurity-privacy-what-are-cookies.html [Last access: 16/06/20]

[10] https://support.google.com/google-ads/answer/6366577?hl=en [Last access: 16/06/20]

raised regarding the collection and misuse of personal information. Scandals such as the Cambridge Analytica-case have drawn scrutiny towards the potential for misuse of personal information by third parties[11]. Poorly managed information security can lead to what Solove (2006) refers to as *insecurity*, which he defines as the carelessness in protecting consumers' personal data. There is also a growing concern that personalized searches based on user data can lead to *filter bubbles*, a term coined by the internet activist Eli Pariser (2011). The term describes the intellectual isolation that can occur as a result of websites using algorithms to selectively assume the information a user would want to see based on previous searches and clicks. The harvesting of users' contact information by third parties has also led to a surge in spam[12]. The cost stemming from productivity loss due to spam was estimated to be of $130 billion, in a study by Ferris Research (Jennings, 2009)[13]. Improper disclosure of user information has also led to an increase in identity thefts, which had an estimated cost of $61 billion in 2006 (Acquisiti et al., 2016)[14]. Users' concerns towards the disclosure of private information seems to be growing over time, as shown by Goldfarb and Tucker (2012) and Stutzman, Gross, and Acquisti (2013).

With technology evolving at a fast pace while regulation is struggling to keep up, industry competition has led to the development of both privacy-invasive and privacy-enhancing technologies in the form of social networks, e-commerce websites, web browsers, and search engines. We refer to privacy as the control or protection of personal information (Acquisti et al., 2016). DuckDuckGo was founded in 2008 to address the growing concern over the lack of consumer privacy on the internet[15]. The company describes itself as *the internet privacy company.* With 93 employees, DuckDuckGo is a David to Goliath compared to Google's more than 100 000 strong staff[16]. Although DuckDuckGo has a modest market share of only 0,53 %[17], the number of searches typed into its browser more than doubled between 2018 and 2019[18].

---

[11] https://www.theguardian.com/technology/2019/mar/17/the-cambridge-analytica-scandal-changed-the-world-but-it-didnt-change-facebook [Last access: 14/05/20]
[12] The indiscriminate use of electronic messaging systems for unsolicited advertisement to customers (Acquisiti et al., 2016).
[13] Others have found this cost to be considerably lower. Rao and Reiley (2012) estimated an overall social cost of spam of $20 billion.
[14] Others have estimated a lower cost. Anderson, Durbin, and Salinger (2008) estimate the total cost of identity theft to be somewhere around $16 billion in 2005. https://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.22.2.171 [Last access: 26/06/20]
[15] https://duckduckgo.com/about [Last access: 16/06/20]
[16] https://edition.cnn.com/2019/04/29/tech/alphabet-q1-earnings/index.html [Last access: 17/06/20]
[17] Ibid., 5.
[18] https://www.webfx.com/blog/seo/2019-search-market-share/[Last access: 30/07/20]

While DuckDuckGo initially survived on donations, it eventually developed a business model similar to that of Google, financing their service through the auctioning of ad words[19]. DuckDuckGo claims to neither track nor store the personal data of its users[20]. Consequentially, it cannot base the results shown to its users on their search history, as opposed to Google. As the founder Gabriel Weinberg puts it: "We don't need to know about you or your search history to deliver a lucrative ad[21]." Users who type the same words into the search engine will see the same results, regardless of age, gender, or search history. Although DuckDuckGo does not collect their users' personal data, it uses aggregated, non-personalized data to improve its services[22]. As a consequence, the quality of the service is improved as the user base increases.

DuckDuckGo donates substantial parts of its profits to organizations working towards raising the standards of trust online[23]. In addition, it has launched several publicity campaigns focusing on the lack of privacy offered by its competitors. As an example, DuckDuckGo placed a billboard in San Francisco's tech-heavy SOMA district with the text "Google tracks you. We don't"[24]. The effort incurred by DuckDuckGo is meant to influence and inform consumers about the scope and potential consequences of consumer data collection. Privacy is already an important issue among consumers. A survey by Wunderman Thompson Data shows that 58 % of 1 500 U.S. respondents are "very concerned about the privacy and security of their personal information and data[25]". The motives behind these concerns are ample. Acquisti et al. (2016) point to several different rationales, including price discrimination, revealing personal information such as disease or unemployment to an unwanted audience, spam, personalized advertisements, or even falling victim to identity theft or scams. DuckDuckGo also points out that the only way to guarantee that a user's personal information will not be handed out to law enforcement as part of a legal investigation is to use sites that do not store personal data[26].

[19] DuckDuckGo has an additional source of revenue, which is *affiliate marketing*. If a user clicks on a link to eBay or Amazon after a search with DuckDuckGo, they collect a fee https://fourweekmba.com/duckduckgo-business-model/ [Last access: 16/06/20]. In fact, they use the ad service of Microsoft and Yahoo to generate ads through the Microsoft Yahoo search alliance. https://www.forbes.com/sites/forbestreptalks/2016/02/19/the-founder-of-duckduckgo-explains-how-to-get-customers-before-you-have-a-product-and-why-challenging-google-isnt-insane/#59c305c24e89 [Last access: 19/06/20]
[20] https://spreadprivacy.com/how-anonymous-is-duckduckgo/ [Last access: 23/06/20]
[21] https://www.wired.co.uk/article/duckduckgo-anonymous-privacy [Last access: 18/06/20]
[22] https://duckduckgo.com/privacy [Last access: 18/06/20]
[23] https://duckduckgo.com/donations [Last access: 23/06/20]
[24] https://www.wired.com/2011/01/duckduckgo-google-privacy/ [Last access: 23/06/20]
[25] https://www.mediapost.com/publications/article/344668/consumers-worry-more-about-privacy-than-any-other.html [Last access: 25/06/20]
[26] https://duckduckgo.com/privacy [Last access: 18/06/20]

DuckDuckGo is by no means the only search engine challenging Google by competing on a dimension other than pure search quality. Niche engines such as CC search, which focuses on copyright-free content, and Ecosia, a German start-up that uses its ad revenues to plant trees, are some examples. Additionally, there are larger and more established competitors such as Microsoft's Bing and Russian Yandex. Although these have more significant market shares, we will focus on the competition between DuckDuckGo and Google due to the different approaches and philosophies they have towards the collection of personalized user data.

This thesis will look at how Google is affected by the competition of a rival who refrains from collecting the users' personalized data in a two-sided framework where consumers are averse to giving away their personal data. At the same time, advertisers are willing to pay a higher fee for targeted ads than for non-targeted ads. Some of the elements upon which this analysis will be made are the difference in the amount of personalization of search results, the amount of personal data collected by Google, investments by DuckDuckGo aimed at raising privacy awareness as well as consumer aversion to sponsored links.

The sheer amount of research papers, seminars, and newspaper headlines concerning the role of data and privacy, as well as recent changes in regulations and the increasing importance of data-driven companies in the marketplace, all point to the importance of developing a deep understanding of the economic consequences of user data. The multi-sided nature of the search engine market makes it an interesting arena for studying the effects of data. Search engines such as Google and DuckDuckGo must capture the interest of at least two different groups in order to operate successfully: users and advertisers. The interplay of these groups and the externalities that exist between them must be taken into account when the platforms device their strategic behavior and decisions. Notably, we will study the tradeoff Google faces when choosing how much data to collect from its users. Google must balance the additional gains in the form of increased revenue per advertisement shown, while at the same time taking into consideration how collecting more data will lead some users to switch to the rival platform. At the same time, DuckDuckGo can invest in campaigns aimed at increasing consumer privacy awareness. For a given level of data collected by Google, this will cause a greater disutility for Google's users, causing some of them to switch to DuckDuckGo. This investment is costly for DuckDuckGo. In optimum, it must strike the ideal balance between capturing customers from Google and keeping their costs at a reasonable level. Another interesting characteristic of this particular marketplace is that consumers do not pay in *monetary* terms for using these services. Instead, they “pay” with their eyeballs and, in the case of Google, with their personal information. Consumers have a limited amount of attention, so the "price" they pay to the search engines is through exposure to sponsored links, which is converted into revenue for the search engines when advertisers compete for this attention bottleneck[27] through the auctioning of advertisement words.

Taking these concerns into consideration, we have developed the following research questions:

How can a search engine influence and exploit sentiments towards privacy to capture customers from a search engine collecting personal consumer data?

How are the incentives of Google to collect personal consumer data and the incentives of DuckDuckGo to influence consumer privacy awareness affected by competitive intensity and aversion to advertisements?

Through the rest of the thesis, we will first provide a review of the relevant economic literature. After that, we present a game-theoretic model of competition in a two-sided market with horizontal differentiation. The model contains three groups: users, advertisers, and platforms. An analysis of the model follows, after which we will present some conclusions. Finally, we will provide some critique concerning limitations to the model.

[27] An attention bottleneck is a result of an attention broker with market power limiting the amount of advertisements shown to consumers in order to raise prices (Armstrong, 2006).

2. Literature review

In this section, we will start by introducing the seminal literature on two-sided markets, upon which the thesis is built. An extensive body of work has been developed within this field during the last two decades. We will then address the literature or attention brokers, which includes services such as search engines, after which we will focus on the role of consumer data. Finally, we will discuss the economic literature on privacy. Although the debate on data and privacy is not a recent occurrence, the interest in these interrelated topics has exploded in recent years. Consequentially, much of the literature that we will discuss is very contemporary, with several important contributions being produced during the months in which this thesis has been written.

2.1 Two-sided markets

Three seminal articles are often credited with the initiation of the literature on two-sided markets: Rochet and Tirole (2003), Anderson and Coate (2005), and Armstrong (2006). Similar models had already been described earlier by Parker and Van Alstyne (2000), Caillaud and Jullien (2001), and Caillaud and Jullien (2003), but under a different name, *cibermediaries.* Rochet and Tirole (2003) coined the term *two-sided markets,* describing a market where network externalities are present, along with a platform with the ability to cross-subsidize between different categories of end-users. We refer to cases where there are more than two sides causing externalities as *multi-sided markets.* According to Rochet and Tirole (2006), the theory of two-sided markets is conceptually related to theories of network externalities.

2.1.1 Network effects and externalities

Katz and Shapiro (1985) differentiate between two kinds of network effects: Direct and indirect. Direct network effects apply when the value a user derives from a service or a product increases as a result of more people joining the service or consuming the product. An example of a service where direct network effects are present is Google Maps. When more drivers are using Google Maps, all other users of the map service receive more information about the current traffic situation, thus increasing the value of the product for all users. Calvano and Polo (2020a) further divide direct network effects into the beforementioned direct network effects and network-like effects, the latter describing a situation where the link between an increased user base and increased used value is more subtle, for example as a result of economies of scale.

Indirect network effects, also known as cross-side externalities, occur when the value that a consumer derives from a good or a service increases with every additional user of an identical and/or interoperable complementary good (Veljanovski, 2007). The presence of indirect network effects characterizes the video gaming industry. There are few incentives for video game producers to invest in the production of a video game for a given console if they expect few users to adopt the console. Similarly, video gamers have little incentive to buy a console if they expect that few video games will be produced for the console. This mechanism can give rise to coordination problems where a platform must lure one side to join or create an expectation that there will be users on both sides. This is known as the *chicken-and-egg problem* (Rochet and Tirole, 2003). Evans (2003) further divide two-sided markets into *market-makers, audience-makers,* and *demand coordinators.* Search engines such as Google and DuckDuckGo fit into the second category, audience-makers, whose role is to match advertisers to audiences. The network effects between these different customer groups (audience and advertisers) are internalized by the platform.

The role of expectations can be a critical aspect of the coordination mechanism when dealing with a chicken-and-egg problem. If a service provider can create the expectation that one or both sides will join a particular platform, these expectations can turn out to be rational in the sense that the outcome will be as expected. Caillad and Jullien (2003) developed the notion of *focality*, a formalization of the intuition that incumbents face favorable beliefs over entrants, potentially resulting in a self-fulfilling prophecy.

According to Parker and Van Alstyne (2002) and Armstrong (2006), in cases where the externalities from one side to the other are stronger than the other way around, it is common that the side who cares less about the other is subsidized. Such is the case in the internet search engine market, where users do not pay for each search. Parker and Van Alstyne (2002) point out that zero-prices or even negative prices are by no means an indication of a lack of market power in two-sided markets. One must take all sides into the equation and look at the overall price levels that include changes on all sides of the market.

Network externalities often tilt the market towards one platform. It can be hard for an entrant to match the utility consumers derive from using the product or service of the incumbent if there are strong direct and/or indirect externalities present. Consequentially, even an entrant with a product of superior quality will risk failing. The incumbent can further reinforce this natural tendency towards concentration through the generation of *switching costs*, as shown by Farrell and Saloner (1986), Katz and Shapiro (1992), as well as Fudenberg and Tirole (2000). Switching costs are costs incurred by consumers who switch to a competing product or service. These costs can stem from having to learn how to operate a different technology when adopting a new product, incompatibility of complementary products, or time spent customizing a profile that is not transferable to the new product, as is the case in social media. In the latter case, personal data acts as a switching cost for the users.

There are cases in which an entrant can use network externalities to its advantage. Caillaud and Jullien (2003) and Jullien (2011) describe how a less efficient entrant in a multi-sided market can capture the whole market from a more efficient incumbent by attracting one of the sides. By setting a low price, the network externalities from this side to the other can potentially force the other side to join. This tactic is known as *divide-and-conquer*. Calvano and Polo (2020a) point to the fact that the possibility of conquering the market by "luring one side on board" can work as a mechanism to reduce market power significantly.

The theory of two-sided markets is also conceptually related to the theory of (market or regulated) multi-product pricing. This theory's contribution is the focus on price structure and the idea that price structures are less likely to be distorted by market power than the price levels. The distinction between the price level and price structure is that the former refers to the total price charged by the platform to the two sides, while the latter refers to the decomposition or allocation of the total price between the buyer and seller (Rochet and Tirole, 2006).

2.1.2 Frameworks within two-sided markets

Several frameworks have been applied to the two-sided market analysis. Some examples are Hotelling, the representative consumer model, and Salop's model of a circular city. The first and last frameworks account for consumer heterogeneity, while the representative consumer framework assumes homogeneity among consumers.

2.1.3 Differentiation

To understand these models, we must understand the concept of differentiation. Following Lancaster (1979), we distinguish between horizontal and vertical differentiation. In short, vertical differentiation is characterized by diversification in a dimension that can be objectively graded from best to worst. Quality is an example of such a dimension. Horizontal differentiation is more subjective, in the sense that it depends on the preferences or location of the consumer. Sweetness in wine is an example of horizontal differentiation. Some consumers prefer bitter wine, while others prefer their wine sweet. A mixed differentiation approach, where products differ both in subjectively and objectively diversifiable dimensions, is also possible.

The inclusion of some sort of differentiation is often considered a necessary assumption in order to obtain a sustainable equilibrium with more than one firm, as we know from the Bertrand setting with fixed costs. Pure *vertical* differentiation may be insufficient to sustain many firms in equilibrium due to the finiteness property (Lahmandi-Ayed, 2004), even when there are no fixed costs present. Resultingly, horizontal differentiation is a common assumption in two-sided markets when dealing with more than one platform. Calvano and Polo (2020b) show that the presence of consumer heterogeneity is not strictly necessary for more than one platform to co-exist. Two originally identical platforms in a two-sided market can co-exist in optimum by choosing different pricing structure and catering to different sides of the market.

2.1.4 The Hotelling model

The Hotelling model developed by Harold Hotelling (1929) describes a duopoly competition between horizontally differentiated firms, although the model can be expanded to include more than two competitors. There are many interpretations of what the differentiation parameter represents. Some examples from the economic literature are physical location, political orientation, or physical attractiveness.

The typical example is that of two ice-cream vendors who are identical in all aspects except for their physical location. Imagine both being located on a popular beach, but on different locations along the beach. Consumers are uniformly distributed along the beach and base their purchasing decisions solely on the distance they have to walk to reach the vendor. When walking towards a vendor, they incur a traveling cost. Consumers take this cost into account when optimizing their purchasing decisions, and vendors take the customers' optimization into account when deciding on their location. Hotelling shows that maximal differentiation is not a stable equilibrium, and the vendors ultimately move towards the center to capture customers from each other. This has become known as the *Principle of Minimum differentiation*.

Tirole (1988) describes two opposing effects that influence the localization decision of the firms. *The demand effect* stems from the firms' incentive to capture customers from their competitors by moving closer to them, leading to less horizontal differentiation. *The strategic effect,* on the other hand, describes the incentive firms have to differentiate themselves from their competitors to soften the competition and consequentially charge higher prices. Firms differentiate by moving away from their competitors. Which effect dominates depends on the specifications of the model.

Several authors have contributed to the economic literature by modeling competition of two-sided markets in a Hotelling setting. Armstrong (2006) uses this framework to create a model of competitive bottlenecks, allowing for multi-homing[28] and turning consumers non-indifferent to advertisements. Jullien (2011) uses a two-sided framework to study price discrimination in a setting with uniform network effects. Kind et al. (2013) use a similar framework to describe the effect of a tax on newspapers on the producers of media content's endogenous localization decision. D'Aspemont et al. (1979) question the Principle of Minimum Differentiation and prove that by substituting Hotelling's linear transportation costs with quadratic costs, maximum differentiation turns out to be the equilibrium strategy of the firms.

2.1.5 The representative consumer model

The representative consumer approach is another way to model competition in a two-sided market. This method was used by Spence (1976) and Dixit and Stiglitz (1977). As the name suggests, this model assumes that the behavior of consumers with different tastes can be aggregated into one representative consumer. Kind, Nilsen, and Sørgård (2009) used this approach to model how much time consumers spend on different media outlets. The model has been criticized for its inability to account for consumer heterogeneity. We will not discuss this further, as this particular model is not used in this thesis.

2.1.6 Salop's Circular City

The last approach we will mention is Salop's Circular City (Salop, 1979), which looks at a situation in which consumers are uniformly located along the edges of a circle. As in the linear city model, consumers incur traveling costs when moving from their location towards the location of a firm. The novelty of this model is an endogenization of the number of firms that

[28] Multi-homing means that some or all groups can consume the product of more than one platform.

enter the market. In order to enter the market, the firms incur a fixed cost, which will be the limiting factor for the final number of firms who decide to establish themselves along the circle.

Due to the endogenization of the number of firms in the market, this model is well suited for analyzing contestable markets, entry, and exit decisions. The model has been criticized because of how the circularity of the differentiation parameter does not allow for extreme positions, making the model unfit for analyzing things such as political views. We will not discuss this framework further, as it is not used in this thesis.

2.2 Attention-brokers and search engines

The two-sidedness in the search engine market stems from the fact that most search engines depend on at least two groups of clients: users and advertisers. Calvano and Polo (2020a, p. 3) describe these markets as a setting where platforms "…harvest the attention of their customers by providing valuable content and then resell that attention to advertisers." In the case of search engines, there are positive network externalities from customers to advertisers.

Whether the network externalities from advertisers to customers should be seen as positive or negative, depends on the industry. Anderson and Coate (2005) find that TV-commercials in the U.S. likely represent a negative externality for viewers. The notation is further supported by Ambrus, Calvano, and Reisinger (2016) and Anderson, Foros, and Kind (2016). Kaiser and Song (2009) find that readers in many magazine segments *appreciate* advertising.

An exciting development since Rochet and Tirole (2006) is the observability of user behavior and attention regarding advertisements. Rochet and Tirole point to the inability of advertisers to measure the actual *transaction* in the sense that it used to be impossible to know whether the user carefully read the ads, thereby generating potential sales, or if she simply scrolled through them. With ads displayed in search engines, the relationship has become measurable. The click-through rate of an advertisement is observed, as well as the scroll speed[29]. The media's purchase price and the advertising fees have thus become more variable since Rochet and Tirole (2006), who described this as being closer to a fixed cost relationship. As a result of this technological development, it is now possible for advertisers to pay directly for user attention, as opposed to paying for user *exposure* to advertisements.

[29] https://www.globalmediainsight.com/blog/scroll-speed-effective-kpi/ [Last access: 05/06/20]

In principle, there could have been, and indeed has been, made attempts at charging users for access to a search engine[30]. Although this possibility exists, the majority of today's search engines are based on a freemium model where users are given free access to the service, while advertisers pay to be among the results displayed in the search. Advertisers are thus dependent on users, with positive network externalities from users to advertisers. Armstrong (2006) rationalizes zero prices by pointing to a situation where competition for users is intense (t is small), or where the advertising revenue is significant. If we apply a non-negativity constraint on prices in this setting, then zero prices can be rationalized, which is what we observe in the search engine market. Search engines that charge users a subscription fee are often catering to niche markets, which would imply a non-intensive competition for users, which can provide a rationale for abstaining from the display of advertisements.

The mutual dependence between advertisers and users place most search engines into the category of *attention brokers*. Google and DuckDuckGo compete for and harvest the attention of their users and then resell this attention to advertisers by exposing the consumers to ads. Wu (2019) essentially states that we should consider all platforms reselling attention as being in the same market, regardless of their *functional definition*. This would broaden the relevant market for search engines to include social media and perhaps even news sites.

The first model of *competition for attention* was created by Anderson and Coate (2005). They shed light on the tradeoff between the quantity of advertisements and the number of readers for two competing news outlets, assuming that the readers see advertisements as a nuisance. In their model, the quantities of advertisements are strategic complements, a result replicated in other papers such as, e.g., Dietl, Lang, and Lin (2013) and Kind et al. (2013). Anderson and Coate (2005) also study the efficiency of the equilibrium quantity of advertisements, as well as the welfare effects of a merger between the two firms. As could be expected from the fact that advertisement quantities are strategic complements, a merger leads to an increase in advertisement quantities. Whether the increase is socially efficient or inefficient depends on the *nuisance cost* of advertising[31].

[30] There are several subscription-based search engines operating today, such as Academic Search, Compendex, EconLit, Merck index and more.
[31] If the nuisance cost is low, there will be an under-provision of advertisements in a competitive market. An increase of advertisement volume as a result of a merger would be efficient in this case. The opposite is true if the nuisance cost is high.

More recent literature has expanded the original model to allow for multi-homing: some or all sides can use several platforms. One example is Athey, Calvano, and Gans (2018), who find that the consumer's ability to multi-home complicates the strategic choices of advertisers, which leads to lower advertisement prices and publisher profits. The authors argue that the accessibility of internet attention platforms, combined with the *freemium* models that give users free access to the platforms, make multi-homing a common phenomenon in the post-internet world. Ambrus et al. (2016) find that shared eyeballs, the attention of users who multi-home, are less lucrative than exclusive, single-homing eyeballs.

Argenton and Prüfer (2012) argue that the search engine market is characterized by a robust and structural tendency towards monopolization, which negatively impacts search quality, innovation, consumer surplus, and total welfare. Prat and Valletti (2019) discuss the effect of increased concentration among attention brokers, in a setting where the brokers have proprietary information about the users' product preferences and use this information to sell targeted ad space to retail product industries. They are able to explain a puzzling behavior among major corporations, who pay for specific brand keywords, despite showing up in organic results[32] for the same words. They conclude that the motivation is to push down competitors' links in organic search results and lock them out from sponsored search results. Our model has some contextual similarities with that of Prat and Valletti. They propose a model in which firms can buy non-targeted ads from traditional mass-media or targeted ads from attention brokers. In our case, we are modeling the competition between search engines with different ability to sell targeted ads. On a theoretical level, our model differs significantly from Prat and Valletti by including advertisements as nuisances to consumers, and by endogenizing the consumers' choice of which platform to use (to mention some).

2.3 The role of data

A keyword lurking in the back- or foreground of many economic papers on attention brokers is *personal data* or *user data,* defined as any data the user creates or owns[33]. Henceforth we will refer to this only as *data*, but in a narrower sense than the conventional use of the word[34]. We can look at data as traces we leave behind intentionally or unintentionally while using electronic equipment. Facebook, for example, owns information about its users' location, age, friends,

[32] By organic results, we mean results that are not paid for by advertisers.
[33] https://www.yourdictionary.com/user-data [Last access: 24/06/20]
[34] Individual facts, statistics or items of information. https://www.dictionary.com/browse/data?s=t [Last access: 24/06/20]

sexual orientation, and gender. Google collects data on what videos their users watch, the ads they click on, their physical location, device information, IP address, and cookie data[35]. The potential areas of use are immense. Data can be used to personalize search results and ads based on preferences and history, improve the quality of algorithms by learning from past mistakes, sold to third parties, or used to improve complementary products (Shaefer et al., 2018).

The economic literature on data is a very contemporary field where a lot has happened in recent years, even months. The concept of firms gathering information about their customers and the market environment they operate in is not a new idea *per se*. However, the scale, scope and speed of data collection today is unprecedented, as is the importance data has for many of the largest and most important firms of today.

2.3.1 Economic concerns regarding data

In their recent paper, de Cornière and Taylor (2020) attempt identifying in which cases data as an input is unilaterally pro- or anti-competitive, by which they mean whether data increases or reduces total welfare in their models. Based on previous work in the economic literature, they highlight four main concerns regarding data and competition.

The first concern is that data can work as an entry barrier or increase existing entry barriers, thereby inducing a winner-takes-it-all situation due to network effects (Furman, Coyle, Flecher, McAuley, and Marsden, 2019, according to de Cornière and Taylor, 2020). The second concern is that a dominant firm may engage in exclusionary conduct related to data, either by refusing other firms access to the data or by signing exclusivity contracts, a concern shared by European antitrust agencies (Autorité de la Concurrence and Bundeskartellamt, 2016). The third concern is broader and is related to exploitative behavior, such as when a firm either uses its dominant position to collect excessive amounts of data or uses its data to extract surplus from consumers through price discrimination (Miranda, 2018 and Gu, Madio, and Regani, 2017, as quoted in de Cornière and Taylor, 2020). The fourth concern is that antitrust authorities lack an understanding of the effects of the increasing number of mergers in the digital sectors where data is involved (Stucke and Grunes, 2016, according to de Cornière and Taylor, 2020).

[35] https://www.cnbc.com/2017/11/20/what-does-google-know-about-me.html [Last access: 24/06/20]

The first of these four concerns is the most commonly cited in recent economic literature. The potential problems stem from the mechanism by which digitization, connectivity to cloud-based infrastructures, combined with cheaper storage and more effective use of data such as improvements in machine learning algorithms, enable firms to improve their products by learning from customer data. In this way, they can improve their product quality, thereby attracting more customers and so on *ad infinitum*. A common concern is that this self-reinforcing cycle, known as *data-enabled learning,* can inhibit market entry (Hagiu and Wright, 2020). This problem is not a new one in the economic literature and is conceptually related to our previous discussion on the structural tendency towards market tipping in two-sided markets. However, there are some subtle differences between market tipping as a result of data compared to other network externalities. One of these differences is the role of customization: the ability of firms to improve their products for each *individual user* based on their particular user experience. Hagiu and Wright link this within-user learning to increased switching costs[36], as opposed to across-user learning, which would be more related to regular network effects. Biglaiser, Calvano, and Crémer (2019) provide an interesting discussion on the differences in competitive advantage created through across-user learning and within-user learning as part of a broader discussion of different ways by which a firm may enjoy an incumbency advantage, including through the access to more data.

The welfare effect of this tendency towards concentration is hard to measure. On the one hand, if concentration leads to increased product quality due to network effects, this can benefit both firms and consumers. The problem occurs if this concentration leads to the firm being able to charge monopoly prices without the disciplining threat of entrants. Which one of these two effects dominates, determines whether market concentration is welfare-enhancing or not.

2.3.2 Data as a source of competitive advantage

Lambrecht and Tucker (2015) discuss whether *proprietary* data can be a source of competitive advantage. They find little evidence for such a view, concluding that big data is neither inimitable nor rare due to the vast availability of alternative sources. The idea that data are non-rival and non-excludable and that having access to data does not *per se* give cause to anti-competitive concerns is shared by Tucker and Wellford (2014) and Varian (2018). These papers

[36] We can imagine a user of a search engine like Google who switches to another search engine. If Google is basing their results on the history of this user, then switching to another search engine may reduce the quality of the results. In this case, within-user learning has created an endogenous switching cost.

also provide an interesting distinction between data and other resources. Due to the non-rival nature of data, Varian states that we should focus on *data-access* instead of ownership.

2.3.3 Data and search results

To study the effect of data on search results, Chiou and Tucker (2017) exploit a policy change in European data retention rules which led to a shortening of data retention time[37] to conduct a natural experiment. The aim was to study the effect of historical data on search accuracy. The authors find little evidence of any improving effect of historical data on the accuracy of search results. An opposite conclusion is reached by Goldfarb and Tucker (2011b), who study the effect of the EU ePrivacy Directive on hypothetical advertisement effectiveness, concluding that the directive caused a decrease in effectiveness through the reduction in the amount of data firms are allowed to store.

2.3.4 Data and scalability

Bajari et al. (2018) research the scalability of data[38]. By measuring the accuracy of weekly sales of 36 different product classes as predicted by the use of data in two dimensions (time and units sold), they find that the prediction accuracy increases with more input data, although at a diminishing rate. This result indicates diminishing returns to scale, thereby providing evidence against the existence of a feedback loop in which big companies get bigger by harvesting proprietary data. Varian (2018) also concludes that data has diminishing returns to scale. The scalability of data is also explored by Schaefer, Sapi, and Szabolcs (2018). They discuss the role of data in improving the quality of recommendation systems, such as internet search engines. Like Bajari et al. (2018), they find evidence indicating diminishing returns to scale from data. They also find that the quality of search results displayed to queries is higher in those cases where the search engine had access to more personalized information.

2.3.5 Data and price discrimination

Valentino-Devries, Singer-Vine, and Soltani (2012) discuss price discrimination facilitated by personal data collection. They conclude that some online retailers may provide different prices

[37] The time a company like Google can retain the users' IP addresses and related search query logs.

[38] Varian (2018) provides three different categories of scale:

I) Classical supply side returns to scale through decreasing average cost.
II) Demand side returns to scale stemming from network effects.
III) Improvements in quality or decrease in costs through learning by doing.

to customers based on the physical distance to a rival brick-and-mortar store. Mikians, Gyarmati, Arramilli, and Laoutaris (2012, 2013) also find evidence of systematic price discrimination, concluding that a price difference of 10 % to 30 % for identical products stems from the revelation of users' search history. Vissers, Nikiforakis, Bielova, and Joosen (2014) find similar evidence of variations in airline ticket prices based on the profiling of customers, but find no evidence of systematic price discrimination, counter to a common belief that airline ticket prices increase as one repeatedly enters a search for a given destination. Belleflamme, Lam, Man, and Vergote (2019) study the effect of a firm's ability to profile their customers in a Bertrand duopoly situation. They conclude that the firms can both make a profit if they are able to profile their customers, but with different abilities. In their model, firms have an incentive to share customer data.

2.3.6 Data and profitability

Various papers have studied the effect of the ability to differentiate, segment, and provide customized results to users on the price a firm can charge for advertisements and products, as well as its effect on profits. Shiller (2013) provides empirical evidence indicating that access to web browsing data is better suited to analyzing a customer's willingness to pay compared to "classical" demographic data. In his study, Shiller concludes that access to such information improves profits by around 12,2%. McKinsey and Company (2016) provide a similar conclusion in a report indicating that more detailed data analytics will improve a firm's ability to segment customers and thereby improve profitability. Bajari et al. (2018) find that sponsored links on personalized searches commands a higher price than on non-personalized (organic) searches. Farahat and Bailey (2012) estimate that targeted advertising generates twice the revenue per ad compared to non-targeted advertisements. Similar results are found by de Cornière and Nijs (2014), and Schaefer et al. (2018). Mayer and Mitchell (2012), as well as Lambrecht and Tucker (2013) and Blake, Nosko, and Tadelis (2015) present different viewpoints. Although the literature diverges somewhat, we will follow the idea that targeted ads command a higher price than non-targeted ads.

Several other authors have discussed alternative effects stemming from the use of data. Prufer and Schottmüller (2017) model a dynamic feedback loop where the marginal cost of quality improvements is declining in the number of previous sales, an example of across-user learning. They also study how data from one market allows a firm to enter connected markets. Farboodi, Mihet, Philippon, and Veldkamp (2019) study the effects of data-enabled learning in a situation where data helps a firm choosing an optimal production technique, thereby increasing the quality of its products. Hagiu and Jullien (2011) discuss the incentive of platforms such as search engines to commit to actions that ultimately reduce consumer utility. They find that informational intermediaries have an incentive to provide a suboptimal result for the consumer in order to increase advertising revenues.

2.3.7 Switching costs

Gehrig and Stenbacka (2007) demonstrate how information sharing effectively reduces switching costs in a repeated-interaction framework between lenders. Proprietary information allows banks to offer relational benefits to customers. These benefits act as switching costs. Customers with high switching costs are lucrative for the banks if they can be targeted by price discrimination in the second period. With information sharing, competitors can poach credit-worthy clients. As such, the initial competition for these customers will be lower, and the benefits offered to the clients are reduced.

2.4 Privacy

The concerns raised as a result of the exponential growth of the importance of data in today's competitive environment are not only related to its effect on concentration and market power. The economic literature on *privacy* addresses the economic value and consequences of protecting and disclosing personal information. The issue of privacy is by no means a new concept, but the extensive development in information technology has brought it to the forefront of the public debate. Acquisti et al. (2016) summarize the literature on the economics of privacy. At its core, the economics of privacy concern the balancing of public and private spheres between individuals, organizations, and governments.

Acquisti et al. (2016) point to the heterogeneity in privacy issues, making it challenging to develop an all-encompassing framework to fit all cases. As described by the authors, the issue of privacy is delicate for several reasons. The value of privacy is hugely heterogeneous among different consumers. The opaqueness of privacy terms means that imperfect and/or asymmetric information will hinder users from making informed choices. These issues make it hard to quantify the tradeoff between protecting the rights of the consumers while at the same time allowing firms to harvest data in order to improve products or cut costs.

According to Acquisti et al. (2016), Chicago school scholars such as Stigler and Posner are credited with initiating the economic debate on privacy. These economists did not explicitly formulate models of privacy. Instead, they presented arguments around its positive and potentially negative impact on consumers. Posner (1978, 1981) argues that excessive privacy protection is inefficient because it leads to asymmetric information. Stigler (1980) comes to a similar conclusion in which he states that any regulatory interference regarding privacy will be ineffective. He argues that privacy allows people to selectively display favorable signals while hiding negative traits, potentially leading to a well-known problem in the economic literature, that of *adverse selection*. Daughety and Reinganum (2010) look at this from the opposite perspective. In their model, consisting of agents whose actions create positive or negative externalities, the optimal level of activity is reached in a regime of privacy. In contrast, a public regime gives individuals an incentive to distort their actions in order to enhance their reputation.

Hirshleifer (1971, 1980) counters the arguments delivered by the Chicago scholars by stating that the private benefits of collected data may outweigh social benefits, leading to an (inefficient) over-investment in the harvest of personal information from other parties. More recently, similar arguments to those of Hirshleifer are fronted by Hermalin and Katz (2006), Burke, Taylor, and Wagman (2012) and Wagman (2014).

Tamir and Michell (2012) discuss the advantages enjoyed by consumers as a result of the provision of their personal data and find that the disclosure of such information can lead to psychological advantage, or as they put it, it is *intrinsically rewarding*. White, Tatonetti, Shan, Altman, and Horvitz (2013) discuss positive synergies between pharmaceutical companies, which can happen when firms give each other access to data. Even more relevant considering the COVID-19 pandemic, Dugas et al. (2012) show how data sharing can provide early alerts for epidemics.

3. The model

In this section, we will model a duopoly competition in the search engine market using a two-sided Hotelling framework. This framework allows us to account for the differences in consumer preference on a one-dimensional, horizontally differentiated spectrum. In our case, it represents the degree of personalization of search results. The two-sidedness allows us to study network externalities between users of the platform and advertisers. The model is an extension of Armstrong (2006), which again extends on Gabszewicz, Laussel, and Sonnac (2001), who propose a newspaper industry model where the readers are single-homers[39] while advertisers are multi-homers. In Armstrong (2006), readers are non-indifferent to advertisements, which will also be the case in our analysis. The model depicts a strategic game in three stages, and the solution concept will be that of a subgame perfect Nash equilibrium. We assume that all players are rational and have access to perfect information. We will use the terms *sponsored links* and *advertisements* interchangeably.

In this model, we consider three types of agents: *consumers (users), platforms, and advertisers.*

3.1 The consumers

The number of consumers in our model is set to be $\theta \in \mathbb{R}^+$, with consumers uniformly distributed in the interval. In what follows, we will normalize θ to 1, providing us with a unit interval. We also assume market coverage and that consumers are single-homers: consumers value both services sufficiently to ensure that all potential consumers will choose one of the two search engines[40], while at the same time ensuring that no one will use *more* than one product.

Consumers differ in their taste for personalization of search results. Some may prefer generic search results, e.g., if they want to avoid ending up in a filter bubble[41]. Others may prefer more

[39] This is a strong assumption. Paw Research Center (2005) conducted a survey in which more than half the respondents answered that they use more than one search engine. We can justify our assumption by stating that we are talking about the installation of DuckDuckGo or Google as a default search engine on the browser, which would exclude the possibility of multi-homing.
[40] Technically, we introduce conditions $Q \geq \alpha A_d + xt$ and $Q \geq \alpha A_g + \gamma\sqrt{D} + (1-x)t$ to ensure market coverage.
[41] As an example, White, Zahay, Thorbjørnsen, and Shavitt (2008) and Turow, Kind, Hoofnagle, Bleakley, and Hennessy (2009) find that consumers react negatively to highly personalized messages. Similarly, Goldfarb and Tucker (2011a) find that obtrusive and targeted ads trigger privacy concerns among consumers to a larger degree than obtrusive but non-targeted ads. A survey by the Digital Advertising Alliance (2013) come to the opposite conclusion: that people prefer getting ads directed towards their interests. It should be noted that Google uses a

personalized search results based on their particular profile. When choosing their preferred platform, consumers incur a transportation cost $t \in \mathbb{R}^{+}$ with horizontal differentiation. *t* is a measure of the degree of differentiation. A low *t* means that the products are close substitutes, which implies intense competition. At the extreme, when *t* equals zero, there is no horizontal differentiation, and the products are seen as perfect substitutes[42]. At the opposite end, a large *t* implies low competitive intensity with little substitutability between the products. For the consumer located at point $x \in [0,1]$, the utility from entering a search into DuckDuckGo can be modeled in the following way:

$$U_d = Q - \alpha A_d - xt, \qquad U_d \in [0, \infty\rangle \qquad (1.1)$$

Q is a measure of the quality of the service. It is exogenously given, and as a simplification, we will assume that quality is equal for both search engines. $\alpha \in \mathbb{R}^{+}$ is a measure of consumer aversion towards advertisements, or the *nuisance cost.* The modeling choice is inspired by Dietl et al. (2013), but also draws inspiration from Armstrong (2003), as well as Anderson and Coate (2005). $A_d \in \mathbb{R}^{+}$ denotes the quantity of sponsored links shown per search, and together with α, it quantifies the impact of advertisements on utility. *x* is the location of a given customer, while *t* is the beforementioned horizontal differentiation parameter.

Similarly, the utility from entering a search into Google is modeled in the following way:

$$U_g = Q - \alpha A_g - \gamma\sqrt{D} - (1 - x)t, \;\; U_g \in [0, \infty\rangle \qquad (1.2)$$

In this utility function, we introduce $\gamma \in \mathbb{R}^{+}$, a consumer's aversion against giving the platform her personal data, and $\sqrt{D}$, where $D \in \mathbb{R}^{+}$, which is the amount of data collected per search by the search engine. We know that the collection of personal data by internet sites and firms is an important issue for many consumers[43], which is why we include it in the utility function of Google's users. The functional form of *D* implies that the subjective cost incurred by the

program called AdSense to filter content to ensure contextually targeted unobtrusive ads, which could reduce the privacy concerns of users (Acquisti et al., 2016).

[42] Not taking the subjective user cost of exposure to sponsored links and the collection of personal data into account.

[43] In a report for Pew Research Center for Internet and Technology, Rainie et al. (2013) show that 86 % of internet users have taken steps to remove or mask their digital footprints, while 68 % believed that the current laws were insufficient to protect consumer privacy.

consumer when parting with personal data is declining as the amount of data grows. We can imagine that as the amount of data collected increases, the consumer grows more indifferent towards giving away additional personal data at the margin[44]. Changing the functional form of *D* does not change the direction of our results.

Consumers in our model are fully aware of the amount of data collected by Google. De Cornière and Taylor (2020) propose a somewhat similar model, with consumers incurring a disutility when departing with their personal data to a private firm. A difference between our model and that of de Cornière and Taylor (2020) is that in our model, disutility is not an opaque feature *uncovered* by privacy awareness investments. Instead, it acts as a disutility *created* through DuckDuckGo's investments in privacy awareness. Although de Cornière and Taylor's model is more realistic in their description of data disutility, this is primarily relevant for welfare analysis and for determining the unilateral anti-competitiveness of data, the objective of de Cornière and Taylor's model, but not ours.

In order to determine the demands faced by the different search engines, we need to find the location of the *marginal consumer*, the consumer who is exactly indifferent between the two search engines. We find her location by equalizing the utility functions for a consumer using DuckDuckGo (Equation 1.1) and the utility function of a consumer using Google (Equation 1.2). Solving for $\hat{x}$, the location of the customer, we find that

$$\hat{x} = \frac{t+\gamma\sqrt{D}-\alpha(A_d-A_g)}{2t} \qquad (1.3)$$

With a market size normalized to 1, the demand for each product is

$$y_d = \hat{x} = \frac{t+\gamma\sqrt{D}-\alpha(A_d-A_g)}{2t}, \qquad y_d \in [0,1] \qquad (1.4)$$

$$y_g = (1-\hat{x}) = \frac{t-\gamma\sqrt{D}+\alpha(A_d-A_g)}{2t}, \qquad y_g \in [0,1] \qquad (1.5)$$

Where $y_g = y_d - 1$.

---

[44] Although the accuracy of this is questionable, it was necessary to include $\sqrt{D}$ in this functional form to achieve a closed-form solution to the model.

3.2 The advertisers

In this model, we follow Anderson and Coate (2005) and Armstrong (2006) by assuming that customers are single-homers while advertisers are multi-homing. In our model, advertisers are price takers and will pay a price equal to their own utility from an ad. As such, they receive zero profit. Albeit a simplification, it is in line with Armstrong (2006), who points out that it is characteristic that the single-homing side is treated well while the interest of the multi-homing side is neglected.

Search engine advertisement works by the auctioning of keywords. We model the value of a keyword to a particular advertiser in the following way:

$$V = \delta \times \sigma(D) \qquad (1.6)$$

Where $\delta \in \mathbb{R}^{+}$ denotes the inherent value of a click, while $\sigma(D)$ denotes the average number of clicks received by advertises per search for a given ad word. Personalized searches based on a consumer's search history increase the chance of a consumer clicking on a sponsored link. This is associated with higher revenues per ad, as shown by Farahat and Bailey (2012), de Cornière and Nijs (2014) and Schaefer et al. (2018). Following this logic, $\sigma(D)$ should be increasing in D. In this model we have normalized δ to one. We provide a simple, linear expression for $\sigma(D)$ to facilitate the solving of the model. From now on, we will assume that

$$\sigma(D) = 1 + D \qquad (1.7)$$

With this in mind, we can simplify the value expression to

$$V = 1 + D \qquad (1.8)$$

which will also be the price an advertiser will pay for an ad. We follow Anderson and Coate (2005) by assuming that advertisers' willingness to pay to reach a given user is independent of the number of users reached.

Because DuckDuckGo collects no consumer data and we have normalized δ to 1, and because data is the only factor affecting the prices, the premiums that DuckDuckGo and Google can charge per sponsored link shown will be (respectively):

$$P_d = 1 \quad (1.9)$$

$$P_g = 1 + D \quad (1.10)$$

In a sense, our model is in line with de Cornière and Taylor (2020) who again follow Armstrong and Vickers (2001) by modeling data as a revenue-shifting input in a competition-in-utility model: For a given level of utility provided, more data commands a higher revenue from each customer.

3.3 The platforms

There are two distinct platforms in our model: Google and DuckDuckGo. The platforms differ in two ways: first, by the degree of personalization of the search results they display, and second, by the amount of consumer data they collect. Google chooses the amount of data, while DuckDuckGo does not collect consumer data at all. Both platforms offer their users free access to their service[45] and generate revenue through the sale of advertisements. DuckDuckGo's profit function can be modeled in the following way:

$$\pi_d = y_d A_d P_d - C(\gamma) \quad (1.11)$$

Simply put, the revenue is equal to their total amount of searches times the number of ads per search times the price DuckDuckGo can charge per ad. From (1.2) and (1.11), we can see that γ appears both in the utility function of Google's users and as a cost for DuckDuckGo. The inclusion of γ in (1.11) reflects that DuckDuckGo spends much of its revenue on campaigns aimed at increasing consumer privacy awareness[46]. By informing the public about the quantity of data they are giving away to companies like Google and the potential adverse effects from the disclosure of said data, DuckDuckGo can increase user aversion to data collection, represented by γ. We will use the terms user aversion to data collection and privacy awareness

[45] We are *assuming,* as opposed to deriving, that the platforms provide users access to the services free of charge as an equilibrium strategy.
[46] Ibit., 5.

interchangeably. The approach is inspired by Grini (2016), who used a similar approach to model how film studios can influence consumer aversion towards piracy. We use a simple, quadratic cost function for γ:

$$C(\gamma) = \frac{\gamma^2}{2}. \tag{1.12}$$

If we insert the demand function from (1.4) and the price from (1.8) as well as the cost function (1.12) into (1.11), the profit function can be re-written as

$$\pi_d = \left[\frac{t+\gamma\sqrt{D}-\alpha(A_d-A_g)}{2t}\right]A_d - \frac{\gamma^2}{2}. \tag{1.13}$$

Google has no costs in our model, and its profit thus only depends on its advertisement revenue. It can be modeled as follows:

$$\pi_g = y_g A_g P_g \tag{1.14}$$

By taking the demand function from (1.5) and the price from (1.8) and inserting them into (1.14) the profit function can be re-written

$$\pi_g = \left[\frac{t-\gamma\sqrt{D}+\alpha(A_d-A_g)}{2t}\right]A_g(1+D) \tag{1.15}$$

Our model differs from existing models by directly incorporating the tradeoff faced by a platform whose consumers are reluctant to give away their personal data, which ultimately affects demand. At the same time, access to user data has a positive effect on prices charged to advertisers. All this while competing against a horizontally differentiated rival who does not collect consumer data but can influence consumer aversion towards data disclosure.

It is important to note the limited role of data in our model. By including data only as a nuisance to the users, we imply that users get no benefit from departing with personalized user data. We are thus ignoring potential benefits such as quality improvements, as it is not the purpose of this model to study this.

3.4 Stages

The game consists of three different stages, which will be solved through backward induction. The timing of the different stages has been set to allow us to solve the model.

Stage I: Google decides on the amount of data it will collect from its users, balancing the demand and price effect resulting from this choice and anticipating and internalizing the strategic response of DuckDuckGo.

Stage II: DuckDuckGo observes the amount of user data Google collects. It then invests in campaigns aimed at increasing consumer aversion towards data collection, thereby affecting the disutility consumers experience from their data being collected by Google.

Stage III: Both platforms observe the choices made in the previous stages and choose the number of advertisements, or sponsored links, that are to be displayed to consumers when they enter a given search into the search engine. Consumers observe all the choices and decide which platform to join.

The timing of the model can be justified by the fact that Google has been in the market for much longer than DuckDuckGo. DuckDuckGo invests in campaigns aimed at consumer privacy awareness precisely as a *reaction* to the large amount of data collected by companies like Google. We are assuming that Google cannot change D after DuckDuckGo has moved, or that making such a change would be too costly to be profitable. The number of sponsored search results displayed by the platforms is a decision that can be altered rapidly and without substantial costs, which is why this is done at the last stage of the game. The timing is critical for obtaining a closed-form solution to the model.

4. Analysis and results

In this section, we will solve the model and derive insight from the different stages. Detailed descriptions of the calculations are found in the appendix, together with the conditions that must be satisfied in order to reach an interior, stable solution. We find the sub-game perfect Nash equilibrium by solving the game by the method of backward induction.

4.1 Stage III

In the last stage of the game, consumers choose their preferred search engine, and the platforms decide on the number of sponsored results that are to be displayed per search. We are interested in the equilibrium demand for each of the platforms. We find this by determining the location of the consumer who is exactly indifferent between the two platforms. We have already done this in the description of the model in the section above. All consumers to the left of the indifferent consumer will use DuckDuckGo, while all consumers to her right will be using Google.

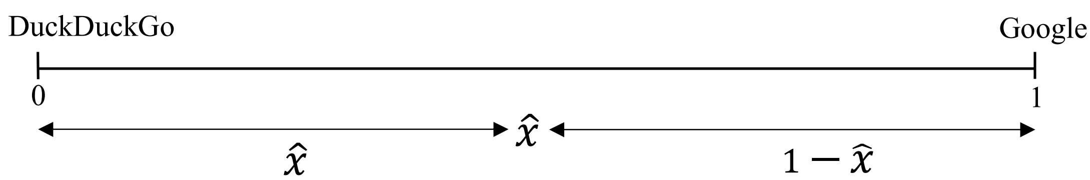

Figure 1: The distribution of users based on the location of the indifferent consumer, $\hat{x}$

For convenience, we repeat the equilibrium demand equations, as well as the profit expressions from the previous section:

$$y_d = \frac{t+\gamma\sqrt{D}-\alpha(A_d-A_g)}{2t} \quad (1.4)$$

$$y_g = \frac{t-\gamma\sqrt{D}+\alpha(A_d-A_g)}{2t} \quad (1.5)$$

$$\pi_d = \left[\frac{t+\gamma\sqrt{D}-\alpha(A_d-A_g)}{2t}\right]A_d - \frac{\gamma^2}{2}. \quad (1.13)$$

$$\pi_g = \left[\frac{t-\gamma\sqrt{D}+\alpha(A_d-A_g)}{2t}\right] A_g(1+D) \tag{1.15}$$

By taking the First Order Conditions (FOCs) of the profit functions with respect to the advertisement quantities, we find their optimal levels:

$$A_d^*(A_g) = \frac{t+\gamma\sqrt{D}}{2\alpha} + \frac{A_g}{2} \tag{2.1}$$

$$A_g^*(A_d) = \frac{t-\gamma\sqrt{D}}{2\alpha} + \frac{A_d}{2} \tag{2.2}$$

In equilibrium, each platform must balance the negative demand effect from a higher amount of advertisements with the positive marginal revenue effect. We see that $\partial A_d/\partial A_g > 0$ and $\partial A_g/\partial A_d > 0$. Advertisement quantities are strategic complements, a typical result in economic models of attention platforms, e.g., Anderson and Coate (2005), Dietl et al. (2013), and Kind, Schjelderup, and Stahler (2013). Given the assumption of market coverage, what the affects the equilibrium demands is the *difference* in advertisement volumes. As such, the platforms have an incentive to provide a quantity of advertisements moving in the same direction as that of their competitor.

If we insert the reaction functions (2.1) and (2.2) into each other, we find the equilibrium quantity of advertisements, equations (2.3) and (2.5). For modeling purposes, using equations (2.4) and (2.6) will make the analysis more straightforward.

$$A_d = \frac{t}{\alpha} + \frac{\gamma\sqrt{D}}{3\alpha} \tag{2.3}$$

$$A_d = \frac{2t}{\alpha} y_d \tag{2.4}$$

$$A_g = \frac{t}{\alpha} - \frac{\gamma\sqrt{D}}{3\alpha} \tag{2.5}$$

$$A_g = \frac{2t}{\alpha} y_g \tag{2.6}$$

From (2.3) and (2.5), we see that advertisement volumes increase in t and decline in α[47]. From (2.4) and (2.6) we can also see that the optimal amount of advertisements is directly related to the demands. The platform with the highest equilibrium demand will have the highest marginal benefit from an increase in advertisement volumes. Inserting the optimal quantities into the demand functions yields:

$$y_d = \frac{1}{2} + \frac{\gamma\sqrt{D}}{6t} \tag{2.7}$$

$$y_g = \frac{1}{2} - \frac{\gamma\sqrt{D}}{6t} \tag{2.8}$$

According to these demand functions, Google will reduce its market share as D increases for any $\gamma > 0$. This result is contrary to our discussion on page 21 on market tipping as a result of data accumulation, e.g., Hagiu and Wright, (2020). According to Hagiu and Wright, the accumulation of data can enable a firm to improve the quality of its products, which again leads to an increase in demand, which causes the firm to obtaining even more data. This self-reinforcing cycle is known as data-enabled learning. According to the writers, this effect can result in a natural tendency towards monopolization. In our model, we have included data only as a nuance factor in the demand function of the users and as a revenue-shifting factor in price. At the same time, we are assuming homogeneity in quality. Because of our modeling choices, DuckDuckGo will have a larger market share than Google in equilibrium. In reality, however, Google has a market share of close to 92 %, while DuckDuckGo has a market share of only 0,53 %[48].

We use equations (2.4), (2.6), (2.7) and (2.8) to find the platforms' profit functions after solving stage III:

$$\pi_d = \left(\frac{1}{2} + \frac{\gamma\sqrt{D}}{6t}\right)^2 \left(\frac{2t}{\alpha}\right) - \frac{\gamma^2}{2} \tag{2.9}$$

$$\pi_g = \left(\frac{1}{2} - \frac{\gamma\sqrt{D}}{6t}\right)^2 \left(\frac{2t}{\alpha}\right)(1 + D) \tag{2.10}$$

[47] A more detailed description of the effect of t and α on advertisement volume will follow from page 43.
[48] Ibid., 5

4.2 Stage II

In this stage, DuckDuckGo decides how much to invest in campaigns aimed at raising consumer privacy awareness. We find the optimal level of γ by taking the FOC of DuckDuckGo's profit with regards to γ. While an increase in γ has a positive demand- and advertisement effect for DuckDuckGo for any given level of D, the effect must be balanced against the cost of such investments. These effects are illustrated below:

$$\frac{\partial \pi_d}{\partial \gamma} = \underbrace{\left(\overbrace{\overbrace{\frac{\partial y_d}{\partial A_d}}^{(-)} \overbrace{\frac{\partial A_d}{\partial \gamma}}^{(+)} + \overbrace{\frac{\partial y_d}{\partial A_g}}^{(+)} \overbrace{\frac{\partial A_g}{\partial \gamma}}^{(-)}}^{Strat.\ effect\ (-)} + \overbrace{\overbrace{\frac{\partial y_d}{\partial \gamma}}^{(+)}}^{Dir.\ effect\ (+)} \right)}_{Demand\ effect\ (+)} A_d + \underbrace{\overbrace{\frac{\partial A_d}{\partial \gamma}}^{Dir.\ effect\ (+)} y_d}_{Advertisement\ effect\ (+)} - \underbrace{\overbrace{\frac{\partial C}{\partial \gamma}}^{Dir.\ effect\ (+)}}_{Cost\ effect\ (-)} \tag{2.11}$$

An increase in γ affects demand, advertisement volume, and cost. The demand effect is both *strategic* through changes in advertisement volumes, and *direct*, as increasing consumer privacy awareness will allow DuckDuckGo to poach customers from Google. Both the demand- and the advertisement effects are positive and must be balanced against the costs of raising consumer privacy awareness.

The solution to stage II is the following:

$$\gamma^*(D) = \frac{3t\sqrt{D}}{9t\alpha - D} \tag{2.12}$$

As can be seen from (2.12), $\partial\gamma/\partial D \geq 0$ as long as $D \geq 0$ and $D \leq 9t\alpha$.

Result 1: *DuckDuckGo will find it optimal to invest more in privacy awareness if Google collects more personalized data.*

4.3 Stage I

In the first stage of the game, Google decides how much data to collect from its users. An increased amount of data is associated with a higher revenue per sponsored link. However, as we have seen from the previous stage, increasing *D* will also lead DuckDuckGo to set a higher γ, lowering Google's demand for any given level of *D*. Google must consider both this strategic effect and the direct effects on both demand, advertisement levels and price when setting *D*.

We combine the solutions from stage III and II to analyze these direct and strategic effects on demand, advertisement volume, and price:

$$\frac{\partial \pi_g}{\partial D} = \underbrace{\left[\overbrace{\overbrace{\overbrace{\frac{\partial y_g}{\partial A_g}}^{(-)} \left(\overbrace{\frac{\partial A_g}{\partial \gamma}}^{(-)} \overbrace{\frac{\partial \gamma}{\partial D}}^{(+)} + \overbrace{\frac{\partial A_g}{\partial D}}^{(-)} \right)}^{(+)} + \overbrace{\overbrace{\frac{\partial y_g}{\partial A_d}}^{(+)} \left(\overbrace{\frac{\partial A_d}{\partial \gamma}}^{(+)} \overbrace{\frac{\partial \gamma}{\partial D}}^{(+)} + \overbrace{\frac{\partial A_d}{\partial D}}^{(+)} \right)}^{(+)} + \overbrace{\overbrace{\frac{\partial y_g}{\partial \gamma}}^{(-)} \overbrace{\frac{\partial \gamma}{\partial D}}^{(+)}}^{(-)}}^{Strat.\ effect\ (+)} + \overbrace{\frac{\partial y_g}{\partial D}}^{Dir.\ effect\ (-)} \right] A_g P_g}_{Demand\ effect\ (-)} + \underbrace{\left[\overbrace{\overbrace{\frac{\partial A_g}{\partial \gamma}}^{(-)} \overbrace{\frac{\partial \gamma}{\partial D}}^{(+)}}^{Strat.\ effect\ (-)} + \overbrace{\frac{\partial A_g}{\partial D}}^{Dir. effect\ (-)} \right] y_g P_g}_{Advertisement\ effect\ (-)} + \underbrace{\overbrace{\frac{\partial P_g}{\partial D}}^{Dir.\ effect\ (+)} y_g A_g}_{Price\ effect\ (+)} \tag{2.13}$$

From (2.13), we can see that an increase in *D* is associated with a reduction in demand- and advertisement levels, and an increase in price. Google maximizes its profit when these marginal effects are exactly equal. The demand- and advertisement effects are the subjective, marginal costs from raising *D*, while the price effect is the marginal gain. From the left side expression in (2.13) we can see how the *demand effect* goes through different channels: First, there is a strategic effect stemming from changes in advertisement levels as a response to changes in *D*[49]. There is also a demand effect from DuckDuckGo's response in γ, as well as a direct demand effect. For the rest of the analysis, it can be a good idea to keep (2.13) in mind as a simple way to visualize several mechanisms which Google must take into account when setting *D*.

Because Google has perfect information and anticipates DuckDuckGo's reaction when setting *D*, we can insert the optimal amount of γ, equation (2.12), into Google's profit function (2.10). By doing this, we obtain the following profit function for Google:

$$\pi_g = \left(\frac{9t\alpha - 2D}{18t\alpha - 2D}\right)^2 \left(\frac{2t}{\alpha}\right)(1 + D) \tag{2.14}$$

We find the optimal amount of data collected by Google by taking the FOC with respect to *D*.

The only solution satisfying the SOCs (Second Order Conditions) from stage II and III is the following (proof in the appendix):

$$D^* = \frac{45t\alpha - 3\sqrt{153t^2\alpha^2 + 16t\alpha}}{4} \tag{2.15}$$

[49] There is an additional effect stemming from the fact that advertisement levels are strategic complements which, for simplification, is not depicted in this equation.

From (2.15), we find that for *D* to be positive, it must be the case that $t\alpha > 2/9$. This condition must be fulfilled: if Google is to personalize searches, it must necessarily collect a positive amount of personalized data. This is a requirement for horizontal differentiation to occur. From (2.15), we can also see that $\partial D/\partial t > 0$.

Result 2: *With lower competitive intensity, Google will find it optimal to collect more personal data from its users.*

When *t* increases, the substitutability between the products is reduced. Consequentially, consumers are willing to accept a higher disutility from the disclosure of their personal data, without switching to the competitor. A more interesting result which can be seen from (2.15) is that $\partial D/\partial \alpha > 0$.

Result 3: *As aversion towards sponsored links increases, Google will find it optimal to collect more personal data from its users.*

This result is robust to changes in the functional form of *D* in the users' utility function[50]. For higher levels of advertisement aversion (α), the marginal demand loss associated with an increase in advertisement volume will be stronger. As such, the platforms find it optimal to reduce the number of advertisements displayed to their customers, thereby diminishing their revenues[51]. Consequentially, the marginal gains of DuckDuckGo from raising awareness of consumer privacy are reduced, while the cost $\gamma^2/2$ is independent of α. The optimal amount of γ for a *given* level of *D* is, therefore, decreasing in α, which we can see from (2.12) as $\partial\gamma^*(D)/\partial\alpha < 0$.

In our model, Google does not incur any direct monetary cost from raising *D.* The cost of raising *D* comes from a loss of demand and thus advertisement volume, as can be seen from (2.13). If we keep γ constant and take the FOC of Google's profit function after stage III, (2.10), we can see that the optimal amount of data collected by Google is *independent* of α (proof in the

[50] Changing the functional form to allow for linearity or squaring *D* in the users' utility function does not change the direction of neither $\partial D/\partial t$ nor $\partial D/\partial \alpha$.

[51] $\partial A_d/\partial\alpha < 0, \partial A_g/\partial\alpha < 0$

appendix). As such, any change in D as a result of a change in α must stem from DuckDuckGo's reduced incentive to respond with a change in γ; the effect is purely strategic.

If we insert (2.15) into (2.12), we obtain the equilibrium level of γ:

$$\gamma^* = \frac{2t\sqrt{45t\alpha - \sqrt{153\ \ ^2\alpha^2+16t\alpha}}}{\sqrt{153t^2\alpha^2+16t\alpha}-3t\alpha} \tag{2.16}$$

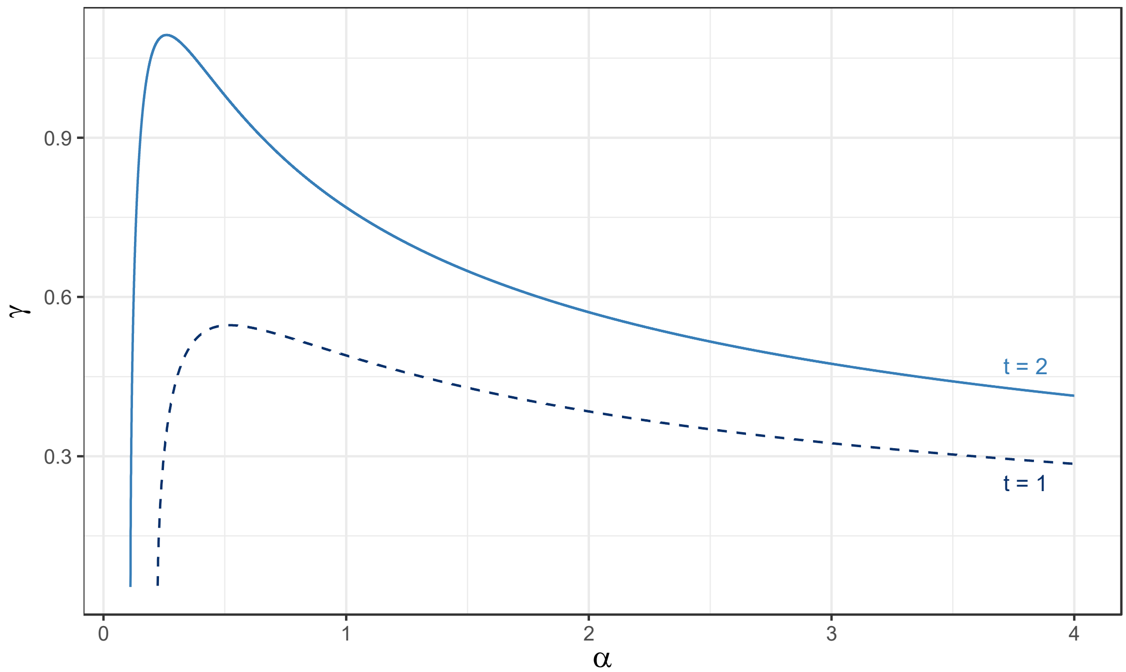

Figure 2: γ as a function of α

From (2.16) and figure 2, we can see that $\partial\gamma/\partial\alpha$ changes sign as α increases, being positive for low values of α and eventually becoming negative. The reason is that α affects γ through two channels, as illustrated below:

$$\frac{\partial\gamma}{\partial\alpha} = \overbrace{\frac{\partial\gamma}{\partial\alpha}}^{Direct\ effect\ (-)} + \overbrace{\overbrace{\frac{\partial\gamma}{\partial D}}^{(+)}\overbrace{\frac{\partial D}{\partial\alpha}}^{(+)}}^{Strategic\ effect(+)} \tag{2.17}$$

From the discussion above, we saw that $\partial\gamma^*(D)/\partial\alpha < 0$. *For a given level of D,* DuckDuckGo has less incentive to increase γ for higher levels of α because reduced advertisement revenue means that DuckDuckGo gains less from every customer captured by increasing privacy awareness. In contrast, the cost of γ is independent of α. This is the direct effect. The strategic effect stems from Google's reaction to an increase in α. For higher levels of α, Google raises *D*, and DuckDuckGo responds with an increase in γ. See (2.17) for an illustration of the direct and strategic effects. For low values of α, the strategic effect dominates the direct effect. However, as α increases, the direct effect will eventually dominate. This result stems from the functional forms of D and γ.

Result 4: *A higher level of aversion towards sponsored links is associated with an increase in γ for low levels of α, and a reduction in γ for high levels of α.*

From (2.16), we can see that $\partial\gamma/\partial t > 0$. From (2.12) we can see that $\partial\gamma^*(D)/\partial t < 0$. As the substitutability between the platforms is reduced, at the margin DuckDuckGo will capture fewer customers from Google by increasing γ for any given level of *D*. This is the direct effect. As seen in the discussion above, Google will respond to a higher level of *t* with an increase in *D*. DuckDuckGo will again respond to this with an increase in γ, which is the strategic effect. As $\partial\gamma^*(D)/\partial t < 0$ but $\partial\gamma/\partial t > 0$, then it must be the case that the strategic effect dominates the direct effect.

Result 5: *A lower competitive intensity is associated with more privacy awareness. This is a result of the strategic effect.*

4.4 Market shares and advertisement levels

We can insert (2.15) and (2.16) into (2.7), (2.8), (2.4) and (2.6) to obtain the equilibrium market shares and advertisement levels:

$$y_d{}^* = \frac{6t\alpha}{\sqrt{153t^2\alpha^2+16t} \; -3t\alpha} \tag{2.18}$$

$$y_g{}^* = \frac{\sqrt{153t^2\alpha^2+16t\alpha}-9t\alpha}{\sqrt{153t^2\alpha^2+16t\alpha}-3t\alpha} \tag{2.19}$$

$$A_d{}^* = \frac{12t^2}{\sqrt{153t^2\alpha^2+16t\alpha}-3t\alpha} \tag{2.20}$$

$$A_g{}^* = \left(\frac{\sqrt{153t^2\alpha^2+16t\alpha}-9t\alpha}{\sqrt{153t^2\alpha^2+16t\alpha}-3t\alpha}\right)\frac{2t}{\alpha} \tag{2.21}$$

From these four equations, we can see that

$$\frac{\partial y_d}{\partial t} > 0 \; for \; all \; t > 0 \tag{2.22}$$

$$\frac{\partial y_g}{\partial t} < 0 \; for \; all \; t > 0 \tag{2.23}$$

$$\frac{\partial A_d}{\partial t} > 0 \; for \; all \; t > 0 \tag{2.24}$$

$$\frac{\partial A_g}{\partial t} > 0 \; for \; all \; t > 0 \tag{2.25}$$

With a higher *t*, the competitive intensity is reduced. As we have seen in the section above, a higher *t* is associated with an increase in both γ and D. This is again associated with a decrease in the demand for Google and an increase in the demand for DuckDuckGo[52]. We find the value of the interaction term $\gamma\sqrt{D}$ by combining (2.15) and (2.16).

$$\gamma\sqrt{D}^{*} = \frac{3t(15t\alpha - \sqrt{153\ {}^{2}\alpha^{2} + 16t\alpha})}{\sqrt{153t^{2}\alpha^{2} + 16\ \ } - 3t\alpha} \tag{2.26}$$

With lower competitive intensity, DuckDuckGo experiences an increase in the marginal gains from advertisements. Below are the different mechanisms at work.

$$\frac{\partial A_d}{\partial t} = \overbrace{\overbrace{\frac{\partial A_d}{\partial \gamma}}^{(+)} \left(\overbrace{\frac{\partial \gamma}{\partial D}}^{(+)} \overbrace{\frac{\partial D}{\partial t}}^{(+)} + \overbrace{\frac{\partial \gamma}{\partial t}}^{(-)} \right) + \overbrace{\frac{\partial A_d}{\partial D}}^{(+)} \overbrace{\frac{\partial D}{\partial t}}^{(+)}}^{Strat.\ effect\ from\ \gamma\sqrt{D}\ (+)} + \overbrace{\frac{\partial A_d}{\partial t}}^{Dir.\ effect\ (+)} \tag{2.27}$$

We can see that the strategic- and direct effects are pulling in the same direction, unilaterally pointing towards an increase in DuckDuckGo's number of sponsored links per search.

With Google, the increase in the interaction term is associated with a loss of demand, which, *ceteris paribus*, would indicate a reduction in advertisement volume. However, there is a direct effect that dominates the effect of the interaction term. The mechanisms are depicted below:

[52] The demand effect is moderated by the *direct effect* of $\partial y_i / \partial t, i = d, g$

$$\frac{\partial A_g}{\partial t} = \overbrace{\overbrace{\frac{\partial A_g}{\partial \gamma}}^{(-)} \left(\overbrace{\frac{\partial \gamma}{\partial D}}^{(+)} \overbrace{\frac{\partial D}{\partial t}}^{(+)} + \overbrace{\frac{\partial \gamma}{\partial t}}^{(-)} \right) + \overbrace{\frac{\partial A_g}{\partial D}}^{(-)} \overbrace{\frac{\partial D}{\partial t}}^{(+)}}^{Strat.\ effect\ from\ \gamma\sqrt{D}\ (-)} + \overbrace{\frac{\partial A_g}{\partial t}}^{Dir.\ effect\ (+)} \tag{2.28}$$

Result 6: *With lower competitive intensity, both platforms will find it optimal to raise their advertisement volumes. The advertisement volume of DuckDuckGo will increase relatively more, as both the strategic and direct effects point in the same direction.*

From (2.18) to (2.21) we can also see that

$$\frac{\partial y_d}{\partial \alpha} > 0\ for\ all\ \alpha > 0 \tag{2.29}$$

$$\frac{\partial y_g}{\partial \alpha} < 0\ for\ all\ \alpha > 0 \tag{2.30}$$

$$\frac{\partial A_d}{\partial \alpha} < 0\ for\ all\ \alpha > 0 \tag{2.31}$$

$$\frac{\partial A_g}{\partial \alpha} < 0\ for\ all\ \alpha > 0 \tag{2.32}$$

The last two equations should come as no surprise: With higher consumer aversion to sponsored links, the optimal *level* of sponsored links should be lower for both platforms because the marginal loss of demand from advertisements increases with higher α. The effect is amplified because advertisement volumes are strategic complements.

The reasons why $\partial y_d/\partial \alpha > 0$ and $\partial y_g/\partial \alpha < 0$ are more intricate. The effects on the demands from a change in aversion to sponsored links α are illustrated bellow.

$$\frac{\partial y_d}{\partial \alpha} = \underbrace{\overbrace{\frac{\partial y_d}{\partial A_d}}^{(-)} \left(\overbrace{\frac{\partial A_d}{\partial \gamma}}^{(+)} \left(\overbrace{\underbrace{\overbrace{\frac{\partial \gamma}{\partial D}}^{(+)} \overbrace{\frac{\partial D}{\partial \alpha}}^{(+)} + \overbrace{\frac{\partial \gamma}{\partial \alpha}}^{(-)}}_{\substack{Advertisement \\ effect\ from\ \gamma\sqrt{D}}}}^{(-/+)} \right) + \overbrace{\frac{\partial A_d}{\partial D}}^{(+)} \overbrace{\frac{\partial D}{\partial \alpha}}^{(+)} + \overbrace{\frac{\partial A_d}{\partial \alpha}}^{(-)} \right) + \overbrace{\frac{\partial y_d}{\partial A_g}}^{(+)} \left(\overbrace{\frac{\partial A_g}{\partial \gamma}}^{(+)} \left(\overbrace{\underbrace{\overbrace{\frac{\partial \gamma}{\partial D}}^{(+)} \overbrace{\frac{\partial D}{\partial \alpha}}^{(+)} + \overbrace{\frac{\partial \gamma}{\partial \alpha}}^{(-)}}_{\substack{Advertisement \\ effect\ from\ \gamma\sqrt{D}}}}^{(-/+)} \right) + \overbrace{\frac{\partial A_g}{\partial D}}^{(-)} \overbrace{\frac{\partial D}{\partial \alpha}}^{(+)} + \overbrace{\frac{\partial A_g}{\partial \alpha}}^{(-)} \right)}_{Effect\ through\ advertisement\ volumes\ (-/+)} + \underbrace{\overbrace{\frac{\partial y_d}{\partial \gamma}}^{(+)} \left(\overbrace{\overbrace{\frac{\partial \gamma}{\partial D}}^{(+)} \overbrace{\frac{\partial D}{\partial \alpha}}^{(+)} + \overbrace{\frac{\partial \gamma}{\partial \alpha}}^{(-)}}^{(-/+)} \right) + \overbrace{\frac{\partial y_d}{\partial D}}^{(+)} \overbrace{\frac{\partial D}{\partial \alpha}}^{(+)}}_{Effect\ throu\quad interaction\ term\ (+)} + \underbrace{\overbrace{\frac{\partial y_d}{\partial \alpha}}^{(-)}}_{Direct\ effect\ (-)} \tag{2.33}$$

$$\frac{\partial y_g}{\partial \alpha} = \underbrace{\frac{\overbrace{\partial y_g}^{(-)}}{\partial A_g}\left(\frac{\overbrace{\partial A_g}^{(-)}}{\partial \gamma}\left(\underbrace{\overbrace{\frac{\overbrace{\partial \gamma}^{(+)}}{\partial D}\frac{\overbrace{\partial D}^{(+)}}{\partial \alpha}+\frac{\overbrace{\partial \gamma}^{(-)}}{\partial \alpha}}^{(-/+)}}_{\substack{Advertisement\\ effect\ from\ \gamma\sqrt{D}}}\right)+\frac{\overbrace{\partial A_g}^{(-)}}{\partial D}\frac{\overbrace{\partial D}^{(+)}}{\partial \alpha}+\frac{\overbrace{\partial A_g}^{(-)}}{\partial \alpha}\right)+\frac{\overbrace{\partial y_g}^{(+)}}{\partial A_d}\left(\frac{\overbrace{\partial A_d}^{(+)}}{\partial \gamma}\left(\underbrace{\overbrace{\frac{\overbrace{\partial \gamma}^{(+)}}{\partial D}\frac{\overbrace{\partial D}^{(+)}}{\partial \alpha}+\frac{\overbrace{\partial \gamma}^{(-)}}{\partial \alpha}}^{(-/+)}}_{\substack{Advertisement\\ effect\ from\ \gamma\sqrt{D}}}\right)+\frac{\overbrace{\partial A_d}^{(+)}}{\partial D}\frac{\overbrace{\partial D}^{(+)}}{\partial \alpha}+\frac{\overbrace{\partial A_d}^{(-)}}{\partial \alpha}\right)}_{Effect\ through\ advertisement\ volumes\ (+/-)}+\underbrace{\frac{\overbrace{\partial y_g}^{(-)}}{\partial \gamma}\left(\overbrace{\frac{\overbrace{\partial \gamma}^{(+)}}{\partial D}\frac{\overbrace{\partial D}^{(+)}}{\partial \alpha}+\frac{\overbrace{\partial \gamma}^{(-)}}{\partial \alpha}}^{(-/+)}\right)+\frac{\overbrace{\partial y_g}^{(-)}}{\partial D}\frac{\overbrace{\partial D}^{(+)}}{\partial \alpha}}_{Effect\ throu\quad interaction\ term\ (-)}+\underbrace{\frac{\overbrace{\partial y_g}^{(+)}}{\partial \alpha}}_{Direct\ effect\ (+)} \tag{2.34}$$

The direct effect (on the right) is stemming from the fact that the differences in advertisements are multiplied by $(\alpha/2t)$ in the demand functions (1.4) and (1.5). The effect is negative for DuckDuckGo and positive for Google because DuckDuckGo always has a larger advertisement volume than Google. In the middle, we can see the effect on demand for an increase in α through the interaction term. Although $\partial\gamma/\partial\alpha$ changes sign and becomes negative when α is high, *D* is always increasing sufficiently in α so as to ensure that $\gamma\sqrt{D}/\partial\alpha > 0$. As such, the demand effect through the interaction term is always positive for DuckDuckGo and negative for Google.

The term on the left shows the effect on demand from changes in the difference between the advertisement volumes. With a higher level of α, both platforms will set a lower advertisement volume. Still, as long as α is relatively small, DuckDuckGo will decrease its advertisement volume at a slower rate than Google, leading to a negative demand effect for DuckDuckGo and a positive demand effect for Google. However, as α becomes larger, the sign of $\partial\gamma/\partial\alpha$ changes from positive to negative. As the advertisement revenue decreases, DuckDuckGo's incentives to respond to an increase in *D* by raising γ is reduced, as mentioned in the discussion on page 38 and 39. The entire expression labeled *advertisement effect through* $\gamma\sqrt{D}$ changes sign and becomes dominating, eventually causing a decrease in the difference between the advertisement volumes, as illustrated in figure 3 below:

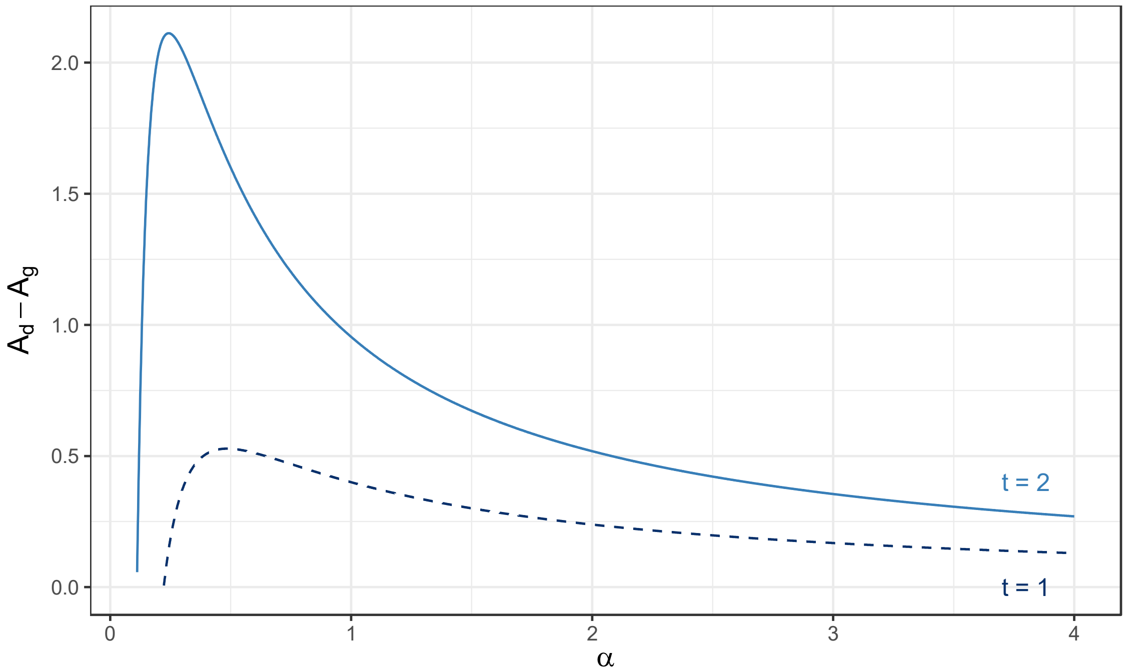

Figure 3: Differences in advertisement volumes as a function of α

We can see that the difference in advertisement volumes grows as α goes up for low values of α, and eventually starts to diminish as the effect through the interaction term begins to dominate. Although the difference in advertisement volumes diminishes, the total demand effect through the interaction term, the center expression in (2.33) and (2.34), will always dominate, leading to results 7 and 8.

Result 7: *A higher level of aversion to sponsored links is associated with a reduction in both advertisement volumes. The reduction in advertisement volumes will be lower for DuckDuckGo than for Google for low levels of aversion to sponsored links, while the reverse is true for high levels*

Result 8: *A higher level of aversion to sponsored links is associated with increased demand for DuckDuckGo and reduced demand for Google.*

4.5 Profit

By inserting (2.15) and (2.16) into the expressions for profit (2.9) and (2.10), we obtain the equilibrium profits as expressed by the parameters t and α:

$$\pi_d{}^* = \frac{3t\left(\sqrt{153t^2\alpha^2+16t\alpha}-3t\alpha\right)}{\alpha\left(81t\alpha+8-3\sqrt{153t^2\alpha^2+16t\alpha}\right)} \tag{2.35}$$

$$\pi_g{}^* = \left(\frac{2349t^2\alpha^2+315t \quad (189t\alpha+15)\sqrt{153t^2\alpha^2+16t} \;+8}{162t\alpha+16- \;\sqrt{15 \quad {}^2\alpha^2+16t}}\right)\frac{4t}{\alpha} \tag{2.36}$$

The profits of both platforms as a function of *t* are depicted below:

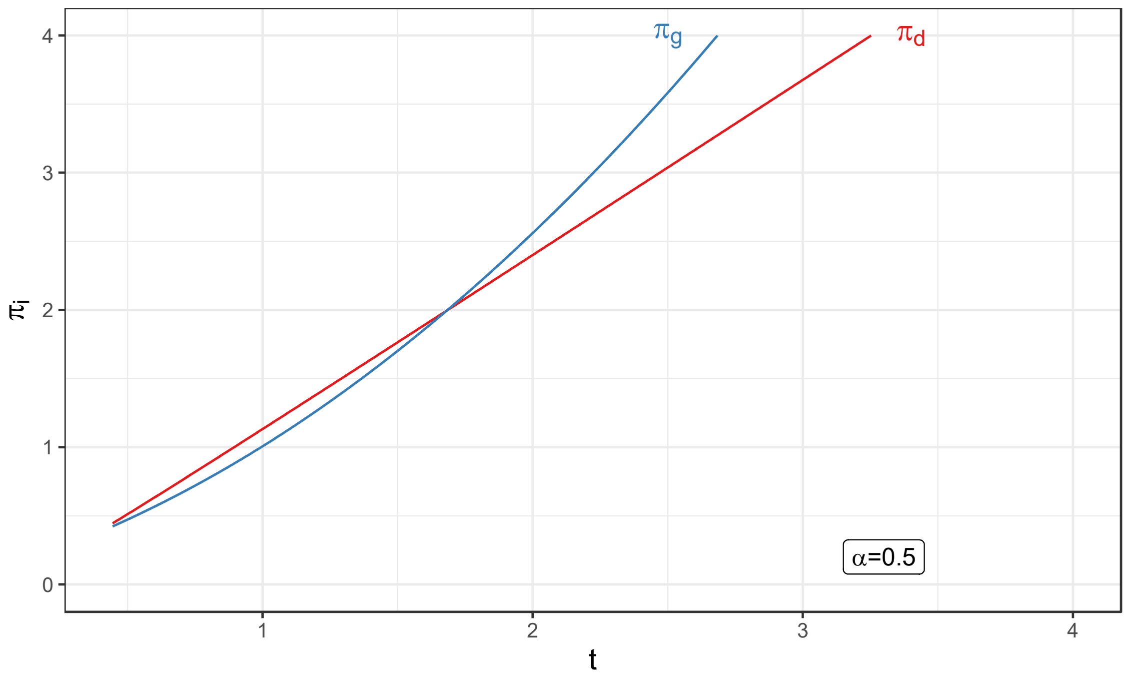

Figure 4: Profits as a function of t

From figure 4, we can see several dynamic effects. The inclusion of data as a revenue-shifting input and consumer aversion towards the collection of data does not change the typical result of the Hotelling model: that an increase in the transportation cost is associated with an increase in profits. With increased transportation costs, competitive intensity is reduced. The products are seen as less substitutable, and it is more costly for a consumer to use a platform that diverges from her horizontal preferences. As a result, the platforms can raise the number of advertisements shown per search, and Google can increase its advertising premium by collecting more personal data from its users. We can see that for low levels of *t*, DuckDuckGo's profit exceeds that of Google. When the competitive intensity is low, DuckDuckGo can capture a relatively large market share from Google by increasing γ. Because the products are seen as close substitutes, the impact of the disutility from data collection is substantial compared to the transportation cost, t. For higher levels of t, the marginal demand effect of data collection diminishes, and Google's profit rises above that of DuckDuckGo.

Result 9: *Both platforms benefit from higher transportation cost (t). For low levels of t, DuckDuckGo has a higher profit than Google. For higher levels of t, Google ultimately achieves a profit higher than DuckDuckGo.*

We will also look at the effect of consumer aversion to sponsored links on profits.

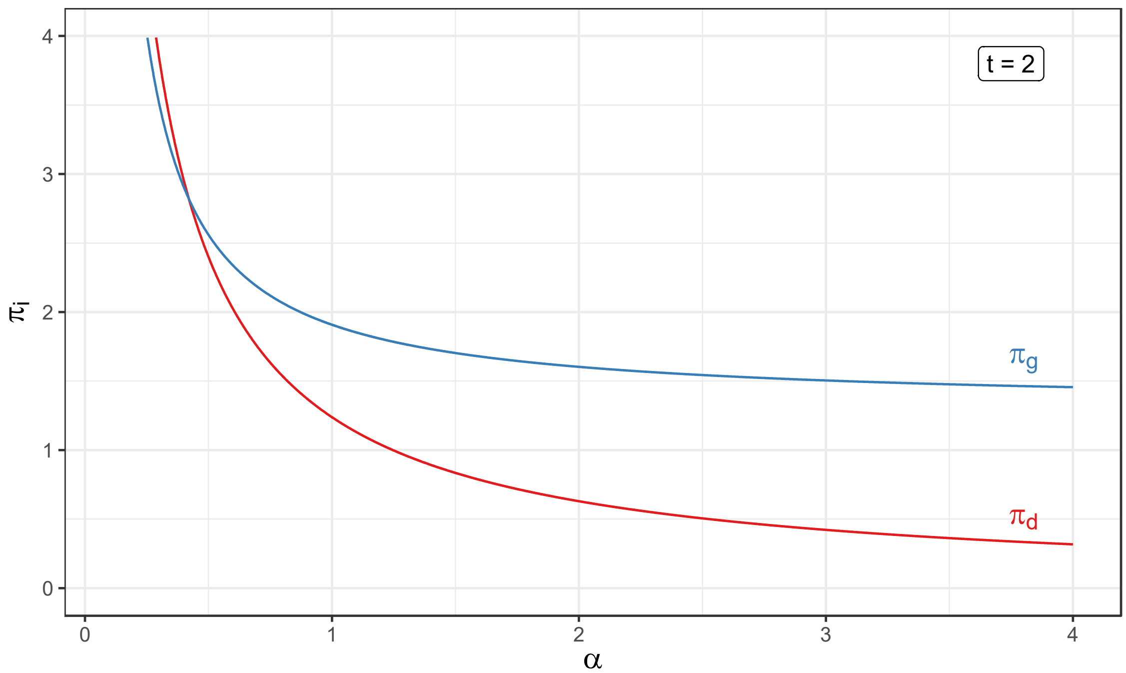

Figure 5: Profits as a function of a

Figure 5 shows the profit level of both platforms as a function of α. The profits of both firms decrease in α, which should come as no surprise; for higher levels of consumer aversion towards sponsored links, the platforms will set lower advertisement quantities, thus lowering their revenue. Google's profit function appears to be less sensitive to an increase in α than that of DuckDuckGo[53]. Although both platforms experience less advertisement revenue for higher levels of α, Google will reduce its losses by setting a higher level of D, as $\partial D/\partial \alpha > 0$. The demand effect of this increase in D will be limited when α is high, as $\partial \gamma/\partial \alpha < 0$ for moderate to high levels of α.

[53] The exception is for very low levels of α

Result 10: *The profits of both platforms are lower for higher levels of consumer aversion towards sponsored links. Google will experience less loss of profit by increasing the amount of data collected, which again will lead to an increase in advertisement premiums.*

4.6 Discussion on welfare analysis

This section contains a brief discussion on how our model is unfit for social welfare analysis. In our case, maximizing total welfare would be the same as maximizing consumer- and producer surplus, as advertisers pay a price equal to the utility derived from advertising. The main problem is that γ appears in the model as a cost for DuckDuckGo and as a disutility for the consumers. The socially optimal amount of γ is, therefore, 0. The implication is that consumers incur no cost from disclosing their data. As such, Google should set D as high as technically possible. In reality, by investing in privacy awareness, DuckDuckGo is not *generating* disutility for consumers. Instead, as modeled by de Cornière and Taylor (2020), by informing consumers of the value of privacy, DuckDuckGo's investment contributes to revealing the otherwise opaque consequences of departing with personal data. For a welfare analysis, we would need to know the actual cost customers incur as a result of data harvesting.

5. Conclusion

The uses of data are ample. Collected data can improve the quality of personalized searches and targeted ads, filter out non-relevant results, and improve user experience. Data represents a resource whose value is still being explored. At the same time, concerns regarding privacy and the protection of personal data (Rainie et al., 2013) and the concentration of market power in the digital sector (Moore and Tambini, 2018) are growing. Our model shows that these concerns create room for small platforms such as DuckDuckGo. By catering to the preferences of users who dislike targeted search results or are worried about leaving behind digital traces, such platforms can poach users from search engines like Google. The model shows that DuckDuckGo has an economic motive to invest in campaigns aimed at raising consumer awareness, and how this motivation affects the incentives of larger companies such as Google to collect consumer data. When Google decides on the amount of personal data to collect from its users, it must carefully balance how this affects their demand, revenues, and the response by competitors and third parties.

We have also discussed the interplay of data, data aversion, and the amount of sponsored links displayed per search. We have discussed how a change in consumer aversion to sponsored links can have surprising effects on the incentives of Google to collect personal data from their users. The result is due to the strategic effects caused by the cost structure in our model. Due to an absence of direct effects, consumer aversion to sponsored links will only influence the amount of data collected by Google if there is a competitor such as DuckDuckGo present in the market. We have also seen how higher levels of transportation costs is associated with an increase in both consumer privacy awareness and the amount of data collected by Google. Additionally, we have seen that the inclusion of data and data awareness as endogenous variables does not alter how profits are positively influenced by higher transportation costs and negatively influenced by higher consumer aversion towards sponsored links. This is a generic result in two-sided Hotelling models, e.g., Anderson and Coate (2005). Still, many questions are left unanswered.

If consumers are averse to giving away personal data, how can it be that a company collecting as much personal data as Google has a market share of close to 92%[54]? One possible explanation is that the quality of the search results in Google is simply better than that of other search

[54] Ibid., 5.

engines. Network externalities and learning may partly explain why Google captures most consumers in the real world. Another possible explanation is that consumers' attitudes and behavior differ. Attitudes are often expressed generically, whereas behavior is specific and contextual (Fishbein and Ajzen, 1975). Turow et al. (2009) provide an interesting example of this inconsistency of attitudes and behavior. They find that 66 % of Americans do not wish for marketers to tailor advertisements to their interests. Simultaneously, the vast majority of them use search engines and social media whose operations are based on enabling advertisers to target them by the use of their personal information. Consumer ignorance may also be at play. Web-users may be unaware of the digital traces they leave behind when navigating the world wide web. Cognitive and behavioral biases, such as immediate-gratification or status-quo bias, may also play a part (Acquisti, 2004., John, Acquisti, and Loewenstein, 2011, according to Acquisti et al., 2016).

As discussed in our analysis, our model is unfit for welfare analysis. It should still be pointed out that the traditional welfare analysis could be insufficient for analyzing social welfare in digital markets. Lynn (2013) highlights how the alleged failure to police tech-firms shows that new methods of thinking may be appropriate, with a focus on values such as liberty, democracy, community, sovereignty, and stability.

The general discussion proposed by this model is relevant for other markets where a dominant player is fortifying its competitive position by the use of personalized data. A smaller firm might be able to compete with a more powerful and resource-rich competitor by exploiting an existing skepticism or turn public opinion against firms collecting large amounts of consumer data. Even if the smaller firm is unable to enter the market, the potential threat of an entrant who can use privacy concerns to grab market shares can discipline the data-intensive company to act competitively, in line with the contestable market theory (Baunol, Pansar, and Willig, 1982). For future research, extensions to the model could allow for differences in quality, for example, by including data as an input to quality and include across-user network externalities. It would also be interesting to see whether the results would differ by allowing Google to endogenously choose their location, letting the amount of data determine the degree of personalization of its search results. Another possibility for future research could be to allow the model to account for diminishing returns to scale from access to proprietary data, or to include complementarities with related products, such as Google Maps or Google's development of artificial intelligence.

6. Limitations to the model

As mentioned on page 35, our model makes the strong assumption that the quality of DuckDuckGo and Google are homogenous. As a result of this modelling choice, and because data is seen only as a nuisance by the consumers, DuckDuckGo obtains an equilibrium market share which is higher than that of Google. As such, the resulting market shares contrast with the empirical evidence. The aim of this thesis is to focus on the strategic effects for DuckDuckGo of investing in privacy awareness to increase its demand at the expense of Google. To isolate this effect, we have excluded the obvious effect of superior search technology resulting from access to big data.

Another limitation to our model is the beforementioned implication that consumer disutility from providing personal data is *created* by DuckDuckGo, instead of being *revealed*. Our model also assumes that DuckDuckGo is the only party that can influence the user's perception of the cost of privacy. In reality, many parties are affecting this, including Google[55]. Our model also assumes that data aversion is the same for all users and that data aversion is not correlated with preferences for generic versus targeted searches.

The model operates with data appearing in a square root in the consumer utility function, implying that the cost of providing personal data to the search engine is declining in the amount of data disclosed by the consumer. The reverse may very well be the case. This choice was necessary in order to obtain a closed-form solution to our model. However, we did change the functional form of data to both a linear and a squared form, without changing the *direction* of any of the effects studied in this thesis.

Our model also allows the users to consume infinite amounts of advertisements, as demands are only affected by the *difference* in advertisement volumes, as opposed to the total levels. Additionally, our model implies a linear relationship between the price Google can charge per sponsored link and user data, while the empirical evidence implies diminishing returns to scale, as mentioned in our literature study.

[55] As an example, Google spent more than $21 million on lobbying in 2018. https://www.bloomberg.com/news/articles/2019-01-22/google-set-2018-lobbying-record-as-washington-techlash-expands. [Last access: 02/08/20]

7. Sources

Acquisti, A. (2004). Privacy in electronic commerce and the economics of immediate gratification. *Proceedings of the 5th ACM conference on Electronic commerce*, pp. 21–29.

Acquisti, A., Taylor, C., and Wagman, L. (2016). The economics of privacy. *Journal of economic Literature*, *54*(2), 442-92.

Ambrus, A., Calvano, E., and Reisinger, M. (2016). Either or both competition: A" two-sided" theory of advertising with overlapping viewerships. *American Economic Journal: Microeconomics*, *8*(3), 189-222.

Anderson, S. P., and Coate, S. (2005). Market provision of broadcasting: A welfare analysis. *The review of Economic studies* 72.4, pp. 947–972.

Anderson, S. P., Foros, Ø., and Kind, H. J. (2016). Competition for advertisers and for viewers in media markets. *The Economic Journal* 128.608 (2018): 34-54.

Anderson, K., Durbin, E., and Salinger, M. (2008). Identity theft. *Journal of Economic Perspectives* 22(2), 171–192.

Argenton, C., and Prüfer, J. (2012). Search engine competition with network externalities. *Journal of Competition Law and Economics*, *8*(1), 73-105.

Armstrong, M., and Vickers, J. (2001). Competitive price discrimination. *Rand Journal of economics*, 579-605

Armstrong, M. (2006). Competition in Two-sided Markets. *The RAND Journal of Economics* Vol. 37, No. 3 (Autumn, 2006), pp. 668-69. Retrieved from http://www.csef.it/WP/wp379.pdf [Last access: 06/06/20]

Athey, S., Calvano, E., and Gans, J. S. (2018). The impact of consumer multi-homing on advertising markets and media competition. *Management Science*, *64*(4), 1574-1590.

Autorité de la Concurrence and Bundeskartellamt. (2016). Competition law and data.

Bajari, P., Chernozhukov, V., Hortaçsu, A., and Suzuki, J. (2018). The impact of big data on firm performance: An empirical investigation. In *AEA Papers and Proceedings* (Vol. 109, pp. 33-37).

Baunol, W. J., Panzar, J. C., and Willig, R. D. (1982). Contestable markets and the theory of industrial structure. *Ilarcourt, Brace, JovanovLch, New York*.

Belleflamme, P., Lam, W., Man, W., and Vergote, W. (2019). Competitive Imperfect Price Discrimination and Market Power. *CESifo Working Paper, No. 7964*

Biglaiser, G., Calvano, E., and Crémer, J. (2019). Incumbency advantage and its value. *Journal of Economics & Management Strategy*, *28*(1), 41-48.

Blake, T., Nosko, C., and Tadelis, S. (2015). Consumer heterogeneity and paid search effectiveness: A large-scale field experiment. Econometrica 83(1), 155–174.

Burke, J., Taylor, C. R., and Wagman, L. (2012). Information acquisition in competitive markets: An application to the us mortgage market. *American Economic Journal: Microeconomics 4(4), 65–106.*

Caillaud, B., and Jullien, B. (2001). Competing cybermediaries. *European Economic Review 45 (4), 797–808.*

Caillaud, B., and Jullien, B. (2003). Chicken & egg: Competition among intermediation service providers. *RAND journal of Economics,* 309–328.

Calvano, E., and Polo, M. (2020a). Market Power, Competition and Innovation in digital markets: A survey. *Information Economics and Policy*, 100853.

Calvano, E., and Polo, M. (2020b). Strategic differentiation by business models: Free-to-air and pay-tv. *The Economic Journal*, *130*(625), 50-64.

Chiou, L., and Tucker, C. (2017). Search engines and data retention: Implications for privacy and antitrust (No. w23815). *National Bureau of Economic Research.*

d'Aspremont, C., Gabszewicz, J. J., & Thisse, J. F. (1979). On Hotelling's" Stability in competition". *Econometrica: Journal of the Econometric Society*, 1145-1150.

Daughety, A., and J. Reinganum (2010). Public goods, social pressure, and the choice between privacy and publicity. *American Economic Journal: Microeconomics 2(2), 191–221.*

De Cornière, A., and De Nijs, R. (2016). Online advertising and privacy. *The RAND Journal of Economics*, *47*(1), 48-72.

De Cornière, A., and Taylor, G. (2020). Data and Competition: a General Framework with Applications to Mergers, Market Structure, and Privacy Policy. *New online seminar on the Economics of Platforms*. Toulouse School of Economics.

Dietl, H., Lang, M., and Lin, P. (2013). Advertising pricing models in media markets: Lump-sum versus per-consumer charges. *Information Economics and Policy*, *25*(4), 257-271.

Digital Advertising Alliance (2013). Poll: Americans want free internet content, value interest-based advertising. Retrieved from https://digitaladvertisingalliance.org/press-release/poll-americans-want-free-internet-content-value-interest-based-advertising [Last access: 06/09/20]

Dixit, A. K., and Stiglitz, J. E. (1977). Monopolistic Competition and Optimum Product Diversity. *The American Economic Review, Vol. 67, No. 3,* pp. 297-308.

Dugas, A. F., Hsieh, Y. H., Levin, S. R., Pines, J. M., Mareiniss, D. P., Mohareb, A., ... and Rothman, R. E. (2012). Google Flu Trends: correlation with emergency department influenza rates and crowding metrics. *Clinical infectious diseases*, *54*(4), 463-469.

Evans, D. S. (2003). Some empirical aspects of multi-sided platform industries. *Review of Network Economics*, *2*(3).

Farahat, A. and M. Bailey (2012). How effective is targeted advertising? In *Proceedings of the 21st International Conference on WWW*, *2012.*

Farboodi, M., Mihet, R., Philippon, T., and Veldkamp, L. (2019). Big data and firm dynamics. In *AEA papers and proceedings* (Vol. 109, pp. 38-42).

Farrell, J., and Saloner, G. (1986). Installed Base and Compatibility: Innovation, Product Preannouncements, and Predation. *The American Economic Review*, *76*(5), 940–955.

Fishbein, M. and Ajzen, I. (1975). Belief, attitude, intention and behavior: An introduction to theory and research. Addison-Wesley. *Journal of business venturing 5, 177-189.*

Fudenberg, D., and Tirole, J. (2000). Pricing a network good to deter entry. *The Journal of Industrial Economics*, 48: 373-90

Furman, J., Coyle, D., Flecher, A., McAuley, D., and Marsden, P. (2019). Unlocking Digital Competition. *Report of the Digital Competition Expert Panel*. HM Government

Gabszewicz, J. J., Laussel, D., and Sonnac, N. (2001). Press advertising and the ascent of the 'Pensée Unique'. *European Economic Review*, *45*(4-6), 641-651.

Gehrig, T., and Stenbacka, R. (2007). Information sharing and lending market competition with switching costs and poaching. *European Economic Review* 51(1), 77–99.

Ghini, M. B. (2016). The Effects of Popcorn Time on Netflix in a Two-Sided Market. (*Master's thesis).* Norwegian School of Economics.

Goldfarb, A., and Tucker, C. (2011a). Online display advertising: Targeting and obtrusiveness. *Marketing Science 30(3)*, 389–404.

Goldfarb, A. and Tucker, C. (2011b). Privacy regulation and online advertising. *Management Science 57(1), 57–71.*

Goldfarb, A., and Tucker, C. (2012). Shifts in privacy concerns. *American Economic Review: Papers and Proceedings 102.*

Gu, Y. Madio, L., and Reggiani, C. (2017). Exclusive Data, Price Manipulation and Market Leadership. *CESifo Working Papers.* Munich Society for the Promotion of Economic Research.

Hagiu, A., and Jullien, B. (2011). Why do intermediaries divert search? *The Rand Journal of Economics*, *42*(2), 337–362.

Hagiu, A., and Wright, J. (2015). Multi-sided Platforms. *International Journal of Industrial Organization 43*, 162-174.

Hagiu, A., and Wright, J. (2020). Data-enabled learning, network effects and competitive advantage. *New online seminar on the Economics of Platforms*. Toulouse School of Economics.

Hermalin, B. E. and Katz, M. L. (2006). Privacy, property rights and efficiency: The economics of privacy as secrecy. *Quantitative Marketing and Economics 4(3)*, 209–239.

Hirshleifer, J. (1971). The private and social value of information and the reward to inventive activity. *The American Economic Review 61(4),* 561–574.

Hirshleifer, J. (1980). Privacy: Its origins, function and future. *Journal of Legal Studies 9(4),* 649–664.

Hotelling, H. (1929). Stability in Competition. *The Economic Journal* 39.153, pp. 41–57.

Jennings, R. (2009). Cost of spam is flattening–our 2009 predictions. *Ferris Research*.

John, L. K., Acquisti, A., and Loewenstein, G. (2011). Strangers on the plane: Context-dependent willingness to divulge sensitive information. *Journal of Consumer Research 37(5),* 858–873.

Jullien, B. (2011). Competition in Multi-Sided Markets: Divide and Conquer. *American Economic Journal*: *Microeconomics 3.*

Kaiser, U., and Song, M. (2009). Do media consumers really dislike advertising? An empirical assessment of the role of advertising in print media markets. *International Journal of Industrial Organization*, *27*(2), 292-301.

Katz, M. L., and Shapiro, C. (1985). Network Externalities, Competition, and Compatibility. *The American Economic Review, 75*(3), 424-440.

Katz, M. L., and Shapiro, C. (1992). Product Introduction with Network Externalities. *The Journal of Industrial Economics*, *40*(1), 55–83.

Kind, H. J., Nilssen, T., and Sørgård, L. (2009). Business models for media firms: Does competition matter for how they raise revenue? *Marketing Science 28*.6, pp. 1112–1128.

Kind, H. J., Schjelderup, G., and Stahler, F. (2013). Newspaper differentiation and investments in journalism: the role of tax policy. *Journal of Media Economics,* 13(4), 215-232.

Lahmandi-Ayed, R. (2004). Finiteness property in vertically differentiated markets: a note on locally increasing and decreasing returns. *Economic Theory*, *23*(2), 371-382

Lambrecht, A., and Tucker, C. (2013). When does retargeting work? Information specificity in online advertising. *Journal of Marketing Research 50(5),* 561–576.

Lambrecht, A., and Tucker, C. (2015). Can Big Data protect a firm from competition? *Available at SSRN 2705530*.

Lancaster, K. (1979). Variety, equity, and efficiency: product variety in an industrial society. New York City: *Columbia University Press*.

Lynn, B. (2013). Antitrust: A Missing Key to Prosperity, Opportunity, and Democracy. *Demos, policy paper*.

Mayer, J. R. and Mitchell, J.C. (2012). Third-party web tracking: Policy and technology. *Security and Privacy (SP),* 2012 IEEE Symposium on, pp. 413–427. IEEE.

McKinsey and Company (2016). The Age of Analytics. Competing in a Data-Driven World. *McKinsey Global Institute Report.*

Mikians, J., Gyarmati, L., Erramilli, V., and Laoutaris, N. (2012). Detecting price and search discrimination on the Internet. In *Proceedings of the 11th ACM Workshop on Hot Topics in Networks*, pp. 79–84.

Mikians, J., Gyarmati, L., Erramilli, V., and Laoutaris, N. (2013). Crowd-assisted search for price discrimination in e-commerce: First results. In *Proceedings of the Ninth ACM Conference on Emerging Networking Experiments and Technologies,* pp. 1–6.

Miranda, M. G. (2018). How Big Data and online markets will lead to higher. – not lower – prices. Retrieved from https://www.marketwatch.com/story/how-big-data-and-online-markets-can-lead-to-higher-prices-2018-05-19 [Last access 29/06/20]

Moore, M., and Tambini, D. (Eds.). (2018). Digital dominance: the power of Google, Amazon, Facebook, and Apple. *Oxford University Press*.

Pariser, E. (2011). The filter bubble: What the internet is hiding from you. Penguin Press, London.

Parker, G. G., and Van Alstyne, M. W., (2000). Internetwork externalities and free information goods. In: *Proceedings of the 2nd ACM Conference on Electronic Commerce.* ACM, pp. 107–116.

Parker, G. G., and Van Alstyne, M. (2002). Unbundling in the presence of network externalities. *Mimeo*.

Paw Research Center (2005). Search engine users. Retrieved from https://www.pewresearch.org/internet/2005/01/23/search-engine-users/ [Last access: 29/08/20]

Posner, R. A. (1978). The right of privacy. *Georgia Law Review 12(3), 393–422*.

Posner, R. A. (1981). The economics of privacy. *The American Economic Review 71 (2)*, 405–409.

Prat, A., and Valletti, T. M. (2019). Attention oligopoly. *New online seminar on the Economics of Platforms*. Toulouse School of Economics. Available at SSRN 3197930.

Prufer, J., and Schottmüller, C. (2017). Competing with big data. *Tilburg Law School Research Paper.*

Rainie, L., Kiesler, S., Kang, R., Madden, M., Duggan, M., Brown, S., and Dabbish, L. (2013). Anonymity, privacy, and security online. *Pew Research Center*, *5*.

Rao, J. M., and Reiley, D. H. (2012). The economics of spam. *The Journal of Economic Perspec- tives 26(3), 87–110.*

Rochet, J.-C., and Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the European economic association* 1 (4), 990–1029.

Rochet, J.-C., and Tirole, J., (2006). Two–sided markets: a progress report. In: *The RAND journal of economics 37.3*, pp. 645–667.

Salop, S. (1979). Monopolistic Competition with Outside Goods. *Bell Journal of Economics 10: pp 141-156.*

Schaefer, M., Sapi, G., and Szabolcs, L. (2018). The effect of Big Data on Recommendation Quality. The Example of Internet Search. *Düsseldorf university pres.*

Shiller, B. R. (2013). First degree price discrimination using big data (p. 32). *Brandeis Univ., Department of Economics.*

Solove, D. J. (2006). A taxonomy of privacy. *University of Pennsylvania Law Review 154 (3)*

Spence, M. (1976). Product selection, Fixed Costs, and Monopolistic Competition. *The Review of Economic Studies*, Volume 43, Issue 2, June 1976,

Stigler, G. J. (1980). An introduction to privacy in economics and politics. *The Journal of Legal Studies 9(4),* 623–44.

Stucke, M. E., and Grunes, A. P. (2016). Introduction: Big Data and Competition Policy. *Big Data and Competition Policy, Oxford University Press (2016)*.

Stutzman, F., Gross, R., and Acquisti, A. (2013). Silent listeners: The evolution of privacy and disclosure on Facebook. *Journal of Privacy and Confidentiality 4(2),* 2.

Tamir, D. I., and Mitchell, J. P. (2012). Disclosing information about the self is intrinsically rewarding. *Proceedings of the National Academy of Sciences*, *109*(21), 8038-8043.

Tirole, J. (1988). The theory of industrial organization. MIT press.

Tucker, D. S., and Wellford, H. (2014). Big mistakes regarding big data. *Antitrust Source, American Bar Association*.

Turow, J., King, J., Hoofnagle, C., Bleakley, A., and Hennessy, M. (2009). Americans reject tailored advertising and three activities that enable it. *Working paper. Available at SSRN 1478214*.

Valentino-Devries, J., Singer-Vine, J., and Soltani, A. (2012). Websites vary prices, deals based on users' information. *Wall Street Journal*, *10*, 60-68.

Varian, H. (2018). Artificial intelligence, economics, and industrial organization (*Working Paper No. w24839).* National Bureau of Economic Research.

Veljanovski, C. (2007). Network Effects and Multi-Sided Markets. Kings College, University of London. *Available at SSRN 1003447.*

Vissers, T., Nikiforakis, N., Bielova, N., and Joosen, W. (2014). Crying wolf? On the price discrimination of online airline tickets. *In Proceedings of the 7th Workshop on Hot Topics in Privacy Enhancing Technologies.*

Wagman, L. (2014). Good news or bad news? Information acquisition and applicant screening in competitive labor markets. *Available at SSRN: http://ssrn.com/abstract=2377849.*

White, R. W., Tatonetti, N. P., Shah, N. H., Altman, R. B., and Horvitz, E. (2013). Web-scale pharmacovigilance: listening to signals from the crowd. *Journal of the American Medical Informatics Association*, *20*(3), 404-408.

White, T. B., Zahay, D. L., Thorbjørnsen, H., and Shavitt, S. (2008). Getting too personal: Reactance to highly personalized email solicitations. *Marketing Letters 19(1),* 39–50.

Wu, T. (2019). Blind Spot: The Attention Economy and the Law. *Antitrust Law Journal*. 83

8. Appendix

We find the location of the marginal consumer from the utility functions, which gives us the demand of DuckDuckGo (and Google, as $\boldsymbol{y_g = 1 - y_d}$)

$$U_d = U_g \rightarrow Q - \alpha A_d - xt = Q - \alpha A_g - \gamma\sqrt{D} - (1-x)t$$

$$-2xt = \alpha A_d - \alpha A_g - \gamma\sqrt{D} - t$$

$$\hat{x} = \frac{t+\gamma\sqrt{D}-\alpha(A_d - A_g)}{2t} \qquad (1.3)$$

We insert the demand from (1.3) into the profit functions and take the FOC with respect to advertisements. This gives us the advertisement reaction functions:

$$\pi_d = \frac{t+\gamma\sqrt{D}-\alpha(A_d - A_g)}{2t}A_d - \frac{\gamma^2}{2}$$

$$\frac{\partial \pi_d}{\partial A_d} = -\frac{\alpha}{2t}A_d + \frac{t+\sqrt{\gamma}D-\alpha(A_d - A_g)}{2t} \times 1 = 0$$

$$A_d = \frac{2t}{\alpha}\left(\frac{t+\gamma\sqrt{D}-\alpha(A_d - A_g)}{2t}\right)\left(= \frac{2t}{\alpha}y_d\right) \qquad (2.4)$$

$$A_d{}^*(A_g) = \frac{t+\gamma\sqrt{D}}{2\alpha} + \frac{A_g}{2} \qquad (2.1)$$

We do the same for Google:

$$\pi_g = \left[\frac{t-\gamma\sqrt{D}+\alpha(A_d - A_g)}{2t}\right]A_g(1+D)$$

$$\frac{\partial \pi_g}{\partial A_g} = \left(-\frac{\alpha}{2t}A_g + \frac{t-\gamma\sqrt{D}+\alpha(A_d - A_g)}{2t} \times 1\right)(1+D) = 0$$

$$(1+D) \neq 0 \Rightarrow \left(-\frac{\alpha}{2t}A_g + \frac{t-\gamma\sqrt{D}+\alpha(A_d - A_g)}{2t} \times 1\right) = 0$$

$$A_g = \frac{2t}{\alpha}\,\frac{t-\gamma\sqrt{D}+\alpha(A_d - A_g)}{2t} = \frac{2t}{\alpha}y_g \qquad (2.6)$$

$$A_g{}^*(A_d) = \frac{t-\gamma\sqrt{D}}{2\alpha} + \frac{A_d}{2} \qquad (2.2)$$

We insert the reaction function of Google into that of DuckDuckGo.

$$A_d{}^*(A_g) = \frac{t+\gamma\sqrt{D}}{2\alpha} + \frac{A_g}{2} \Rightarrow A_d = \frac{t+\gamma\sqrt{D}}{2\alpha} + \frac{\frac{t-\gamma\sqrt{D}}{2\alpha} + \frac{A_d}{2}}{2}$$

$$A_d{}^* = \frac{t}{\alpha} + \frac{\gamma\sqrt{D}}{3\alpha} \qquad (2.3)$$

We insert the optimum quantity of advertisement from (2.3) into the reaction function of Google.

$$A_g{}^*(A_d) = \frac{t-\gamma\sqrt{D}}{2\alpha} + \frac{A_d}{2} \Rightarrow A_g = \frac{t-\gamma\sqrt{D}}{2\alpha} + \frac{\frac{t}{\alpha}+\frac{\gamma\sqrt{D}}{3\alpha}}{2} = \frac{3\left(t-\gamma\sqrt{D}\right)+3t+\gamma\sqrt{D}}{6\alpha}$$

$$A_g{}^* = \frac{t}{\alpha} - \frac{\gamma D}{3\alpha} \qquad (2.5)$$

To calculate the equilibrium demands after stage III, we insert the advertisement quantities (2.3) and (2.5) into the demand of DuckDuckGo (1.3).

$$y_d = \frac{t+\gamma\sqrt{D}-\alpha\left(A_d-A_g\right)}{2t}$$

$$y_g = 1 - y_d \Rightarrow y_d = \frac{t+\gamma\sqrt{D}-\alpha\left(\frac{2t}{\alpha}y_d-\frac{2t}{\alpha}y_g\right)}{2t} = \frac{1}{2} + \frac{\gamma\sqrt{D}}{2t} - \left[y_d-(1-y_d)\right]$$

$$3y_d = \frac{1}{2} + \frac{\gamma\sqrt{D}}{2t} + 1 \Rightarrow y_d = \frac{1}{2} + \frac{\gamma\sqrt{D}}{6t} \qquad (2.7)$$

$$y_g = 1 - y_d = 1 - \left(\frac{1}{2}+\frac{\gamma\sqrt{D}}{6t}\right) = \frac{1}{2} - \frac{\gamma\sqrt{D}}{6t} \qquad (2.8)$$

In stage II, we find the optimal level of γ by inserting the demand from (2.7) and the optimal advertisement quantities from (2.3) and (2.5) into DuckDuckGo's profit function. We then take the FOC with respect to γ.

$$\pi_d = \left(\frac{1}{2}+\frac{\gamma\sqrt{D}}{6t}\right)^2\left(\frac{2t}{\alpha}\right) - \frac{\gamma^2}{2}$$

$$\frac{\partial\pi_d}{\partial\gamma} = \frac{2\left(3t+\gamma\sqrt{D}\right)\sqrt{D}}{18t\alpha} - \gamma = 0$$

$$\gamma^*(D) = \frac{3t\sqrt{D}}{9t\alpha-D} \qquad (2.12)$$

From (2.12) we can confirm result 1.

$$\frac{\partial\gamma}{\partial D} = \frac{2t\sqrt{D}(9t\alpha+D)}{(9t\alpha-D)^2} > 0$$

We find the optimal amount of data collected by Google by inserting (2.12) into Google's profit function. We take the FOC with respect to *D*.

$$\pi_g = \left(\frac{1}{2}-\frac{\gamma\sqrt{D}}{6t}\right)^2\left(\frac{2t}{\alpha}\right)(1+D) = \left(\frac{3t-\frac{3t\sqrt{D}}{9t\alpha-D}\sqrt{D}}{6t}\right)^2\left(\frac{2t}{\alpha}\right)(1+D)$$

$$\pi_g = \left(\frac{9t\alpha-2D}{18t\alpha-2D}\right)^2\left(\frac{2t}{\alpha}\right)(1+D) \qquad (2.14)$$

$$\frac{\partial\pi_g}{\partial D} = 2\left(\frac{9t\alpha-2D}{18t\alpha-2D}\right)\left(\frac{-2(18t\alpha-2D)+2(9t\alpha-2D)}{(18t\alpha-2D)^2}\right)\left(\frac{2t}{\alpha}\right)(1+D) + \left(\frac{9t\alpha-2D}{18t\alpha-2D}\right)^2\left(\frac{2t}{\alpha}\right) = 0$$

From this, we can see that $D = 9t\alpha/2$ is a solution. But this does not satisfy the SOC (it is a minimum point; $\partial^2\pi_g/\partial D^2 > 0$ when $D = 9t\alpha/2$ (See *conditions* bellow)). The two other solutions are

$$\rightarrow D = \frac{45t\alpha \pm \sqrt{2025t^2\alpha^2 - 4 \times 2 \times (81t^2\alpha^2 - 18t\alpha)}}{4} = \frac{45t\alpha \pm 3\sqrt{153t^2\alpha^2 + 16t\alpha}}{4}$$

The only one of these solutions satisfying the SOC is the following (proof in *conditions of stability/SOC*)

$$\rightarrow D = \frac{45t\alpha - 3\sqrt{153t^2\alpha^2+16t\alpha}}{4} \tag{2.15}$$

By inserting the optimal quantity of data collected (2.15) into (2.12), we find the equilibrium quantity of γ:

$$\gamma = \frac{3t\sqrt{D}}{9t\alpha - D} = \frac{3t\sqrt{\frac{45t\alpha - 3\sqrt{153t^2\alpha^2 + 16t\alpha}}{4}}}{9t\alpha - \frac{45t\alpha - 3\sqrt{153t^2\alpha^2 + 16t\alpha}}{4}}$$

$$\gamma = \frac{2t\sqrt{45t\alpha - 3\sqrt{153t^2\alpha^2+16t\alpha}}}{\sqrt{153t^2\alpha^2+16t\alpha} - 3t\alpha} \tag{2.16}$$

We find the equilibrium demands by inserting (2.15) and (2.15) into (2.7) and (2.8)

$$y_d = \frac{3t+\gamma\sqrt{D}}{6t} = \frac{1}{2} = \frac{6t}{\sqrt{153t^2\alpha^2+16t\alpha} - 3t\alpha} \tag{2.18}$$

$$y_g = 1 - y_d = \frac{\sqrt{153t^2\alpha^2+16t\alpha} - 9t\alpha}{\sqrt{153t^2\alpha^2+16t\alpha} - 3t\alpha} \tag{2.19}$$

We find the equilibrium advertisement levels by inserting (2.18) and (2.19) into (2.4) and (2.6)

$$A_d = y_d \frac{2t}{\alpha} = \frac{12t^2}{\sqrt{153t^2\alpha^2+16t\alpha} - 3t\alpha} \tag{2.20}$$

$$A_g = (1 - y_d)\frac{2t}{\alpha} = \left(\frac{\sqrt{153t^2\alpha^2+16t\alpha} - 9t\alpha}{\sqrt{153t^2\alpha^2+16t\alpha} - 3t\alpha}\right)\frac{2t}{\alpha} \tag{2.21}$$

We find the value of the interaction term $\gamma\sqrt{D}$ by combining (2.12) and (2.15)

$$\gamma\sqrt{D} = \frac{3t\sqrt{D}}{9t\alpha - D}\sqrt{D} = \frac{3t\left(\frac{45t\alpha - 3\sqrt{153t^2\alpha^2+16}}{4}\right)}{9t\alpha - \left(\frac{45t\alpha - \sqrt{153t^2\alpha^2+16}}{4}\right)} = \frac{3t\left(15t\alpha - \sqrt{153t^2\alpha^2+16}\right)}{\sqrt{153t^2\alpha^2+16t\alpha} - 3t\alpha} \tag{2.26}$$

Inserting (2.16) and (2.26) into the profit function of DuckDuckGo gives us the equilibrium profit as expressed only by the parameters t and α:

$$\pi_d = \left(\frac{3t + \gamma\sqrt{D}}{6t}\right)^2 \frac{2t}{\alpha} - \frac{\gamma^2}{2} = \left(\frac{6t\alpha}{\sqrt{153t^2\alpha^2 + 16t\alpha} - 3t\alpha}\right)^2 \frac{2t}{\alpha} - \frac{\left(\frac{2t\sqrt{45t\alpha - 3\sqrt{153t^2\alpha^2 + 16t\alpha}}}{\sqrt{153t^2\alpha^2 + 16t\alpha} - 3t\alpha}\right)^2}{2}$$

$$\pi_d = \frac{3t\left(\sqrt{153t^2\alpha^2+16t\alpha} - 3t\alpha\right)}{\alpha\left(81t\alpha + 8 - 3\sqrt{153t^2\alpha^2+16t}\right)} \tag{2.35}$$

We find the equilibrium profit of Google by combining (2.15) and (2.19).

$$\pi_g = \left(y_g\right)^2 \frac{2t}{\alpha}(1 + D) = \left(\frac{\sqrt{153t^2\alpha^2 + 16t\alpha} - 9t\alpha}{\sqrt{153t^2\alpha^2 + 16t\alpha} - 3t\alpha}\right)^2 \frac{2t}{\alpha}\left(1 + \frac{45t\alpha - 3\sqrt{153t^2\alpha^2 + 16t\alpha}}{4}\right)$$

$$\pi_g = \left(\frac{2349\ {}^2\alpha^2+315t\quad (189t\alpha+15)\sqrt{153t^2\alpha^2+16\quad}+8}{162t\alpha+16-\ \sqrt{153t^2\alpha^2+16t\alpha}}\right)\frac{4t}{\alpha} \tag{2.36}$$

Conditions of stability/SOCs:

To ensure that we have found stable, local maximum points, we must ensure that the second order conditions of our optimizations give negative results.

From stage III: Checking the SOC for DuckDuckGo

$$\frac{\partial \pi_d}{\partial A_d} = -\frac{\alpha}{2t}A_d + \frac{t+\gamma\sqrt{D}-\alpha\left(A_d-A_g\right)}{2t}\times 1 = 0$$

$$\frac{\partial^2 \pi_d}{\partial {A_d}^2} = -\frac{\alpha}{2t}-\frac{\alpha}{2t} = -\frac{\alpha}{t} \leq 0 \text{ because } (\alpha \geq 0, t \geq 0)$$

From stage III: Checking the SOC for Google.

$$\frac{\partial \pi_g}{\partial A_g} = \left(-\frac{\alpha}{2t}A_g + \frac{t-\gamma\sqrt{D}+\alpha\left(A_d-A_g\right)}{2t}\times 1\right)(1+D) = 0$$

$$\frac{\partial^2 \pi_g}{\partial {A_g}^2} = \left(-\frac{\alpha}{2t}-\frac{\alpha}{2t}\right)(1+D) = -\frac{\alpha}{t}(1+D) \leq 0 \text{ because } (\alpha \geq 0, t \geq 0, D \geq 0)$$

From stage II: Checking the SOC for DuckDuckGo:

$$\frac{\partial \pi_d}{\partial \gamma} = \frac{2\left(3t+\gamma\sqrt{D}\right)\sqrt{D}}{18t\alpha} - \gamma$$

$$\frac{\partial^2 \pi_d}{\partial \gamma^2} = \frac{2D}{18t\alpha} - 1 \leq 0\ if\ D \leq 9t\alpha$$

This condition leaves us with only one alternative for D

$$D^* = \frac{45t\alpha - 3\sqrt{153t^2\alpha^2+16t\alpha}}{4} \tag{2.15}$$

From stage I: Checking the SOC for Google

$$\frac{\partial \pi_g}{\partial D} = 2\left(\frac{9t\alpha-2D}{18t\alpha-2D}\right)\left(\frac{-18t\alpha}{(18t\alpha-2D)^2}\right)\left(\frac{2t}{\alpha}\right)(1+D) + \left(\frac{9t\alpha-2D}{18t\alpha-2D}\right)^2\left(\frac{2t}{\alpha}\right) = 0$$

$$\begin{aligned}\frac{\partial^2 \pi_g}{\partial D^2} = 2\left(\frac{-18t\alpha}{(18t\alpha-2D)^2}\right)^2\left(\frac{2t}{\alpha}\right)(1+D) + 2\left(\frac{9t\alpha-2D}{18t\alpha-2D}\right)\frac{72t\alpha}{(18t\alpha-2D)^3}\left(\frac{2t}{\alpha}\right)(1+D) \\ + 2\left(\frac{9t\alpha-2D}{18t\alpha-2D}\right)\left(\frac{-18t\alpha}{(18t\alpha-2D)^2}\right)\left(\frac{2t}{\alpha}\right) + 2\left(\frac{9t\alpha-2D}{18t\alpha-2D}\right)\left(\frac{-18t\alpha}{(18t\alpha-2D)^2}\right)\left(\frac{2t}{\alpha}\right)\end{aligned}$$

$$= \frac{9t^2\left([45t\alpha+4]\left(\frac{45t\alpha-3\sqrt{153t^2\alpha^2+16t\alpha}}{4}\right) - 162t^2\alpha^2 - 9t\alpha\right)}{\left(\frac{9t\alpha-3\sqrt{153t^2\alpha^2+16t\alpha}}{4}\right)^4} \leq 0\ if$$

$$1377t^2\alpha^2 - 135t\alpha\sqrt{153t^2\alpha^2+16t\alpha} + 144t\alpha - 12\sqrt{153t^2\alpha^2+16t\alpha} \leq 0$$

(Which is true for any $\alpha \geq 0, t \geq 0$)

Other conditions

Because the demand appears in squared form in optimum, we must ensure that it is positive for both platforms. If not, the platforms could have a negative demand and a negative advertisement volume and receive a positive profit. Both demands are positive as long as $\gamma\sqrt{D} \leq 3t$, which we can see from (2.26) is always true.

$$\gamma\sqrt{D} \leq 3t \rightarrow \frac{3t\left(15t\alpha - \sqrt{153t^2\alpha^2 + 16t\alpha}\right)}{\sqrt{153t^2\alpha^2 + 16t\alpha} - 3t\alpha} \leq 3t \rightarrow -9t^2\alpha^2 \leq 2t\alpha$$

Ensuring that the optimal amount of consumer aversion to data is non-negative:

$$\gamma \geq 0 \rightarrow \frac{3t\sqrt{D}}{9t\alpha - D} \geq 0 \rightarrow D \geq \text{and } D < 9t\alpha \left(\text{in optimum: } t\alpha > \frac{2}{9}\right)$$

This must hold without the equality, because $t\alpha = \frac{2}{9}$ would give $D = 0$, which gives us an irrational solution for γ.

Ensuring that the optimal amount of data collected is positive:

$$D > 0 \rightarrow \frac{45t\alpha - 3\sqrt{153t^2\alpha^2 + 16t\alpha}}{4} > 0 \rightarrow t\alpha > \frac{2}{9}$$

It must hold in optimum, or Google will not be able to personalize searches. This is why all graphs in this thesis start at $t\alpha > 2/9$.

Proof of D's independence of α keeping γ constant:

$$\pi_g = \left(\frac{1}{2} - \frac{\gamma\sqrt{D}}{6t}\right)^2 \left(\frac{2t}{\alpha}\right)(1 + D)$$

FOC. Keeping γ constant:

$$\frac{\partial \pi_g}{\partial D} = 2\left(\frac{1}{2} - \frac{\gamma\sqrt{D}}{6t}\right)\left(\frac{\gamma}{12t\sqrt{D}}\right)\left(\cancel{\frac{2t}{\alpha}}\right)(1 + D) + \left(\frac{1}{2} - \frac{\gamma\sqrt{D}}{6t}\right)^2 \left(\cancel{\frac{2t}{\alpha}}\right) = 0$$

We see that the optimal solution is independent of α if we keep γ constant.

This thesis aims to analyze the consumer data's role as a revenue-shifting input in a two-sided competition-in-utility model between a search engine collecting personalized consumer data (Google) and one that does not (DuckDuckGo). Search engines are examples of platforms characterized by network externalities. They harvest the attention of users and resell this attention to advertisers. As such, standard market mechanisms typical of single-sided markets do not apply. In our model, advertisers are willing to pay a higher premium for targeted search results utilizing consumer data. As such, a search engine like Google can command higher prices per sponsored link if it collects more personalized data from its users. This must be balanced against demand effects stemming from consumer aversion towards the disclosure of personal data. DuckDuckGo, a search engine that explicitly does not collect consumer data, can use investments in consumer aversion strategically to capture market shares from Google. We find that Google has an incentive to collect consumer data and that the presence of DuckDuckGo will moderate the amount. We also find that as consumer aversion to advertisements increases, Google will choose to collect more consumer data as DuckDuckGo's incentives to respond with investments in privacy awareness are reduced.

SNF

Samfunns- og næringslivsforskning AS
Centre for Applied Research at NHH

Helleveien 30
NO-5045 Bergen
Norway

P +47 55 95 95 00
E snf@snf.no
W snf.no

Trykk: Allkopi Bergen