Search engine competition with network externalities

Argenton, C.; Prüfer, J.

Published in:
Journal of Competition Law and Economics

DOI:
10.1093/joclec/nhr018

Publication date:
2012

Search Engine Competition with Network Externalities*

Cédric Argenton and Jens Prüfer
Tilburg University†

January 16, 2012

Abstract

The market for Internet search is not only economically and socially important, it is also highly concentrated. Is this a problem? We study the question whether "competition is only a free click away". We argue that the market for Internet search is characterized by indirect network externalities and construct a simple model of search engine competition, which produces a market share development that fits the empirically observed development since 2003 well. We find that there is a strong tendency towards market tipping and, subsequently, monopolization, with negative consequences on economic welfare. Therefore, we propose to require search engines to share their data on previous searches. We compare the resulting "competitive oligopoly" market structure with the less competitive current situation and show that our proposal would spur innovation, search quality, consumer surplus, and total welfare. We also discuss the practical feasibility of our policy proposal and sketch the legal issues involved.

JEL classification: L10, K23, L86
Keywords: Search engines, network externalities, query logs, antitrust, regulation

*We are grateful to Erik Brynjolfsson, Lapo Filistrucchi, Joshua Gans, Damien Geradin, Michael Grubb, Pierre Larouche, Catherine Tucker, and seminar or conference participants at Tilburg University and Northwestern University for comments on a previous draft of this paper. All errors are our own. We wish to disclose that TILEC has entered into a sponsorship agreement with Microsoft Inc., although neither this research nor any other of our respective research projects is funded under this program.

†Department of Economics, CentER, TILEC, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands; c.argenton@uvt.nl; j.prufer@uvt.nl.

1 Introduction

"There's always a concern that large private collections of data are not available to search engines... . We've taken a position, both in a religious and in a business perspective, that the world is better off if you take the information that you're assembling and make it accessible."
Eric Schmidt, then CEO of Google[1]

The market for Internet search has grown from virtually zero in the mid-1990s to a multi-billion dollar business nowadays. Internet search revenues in the United States totaled $5.7 billion for the first six months of 2010, an increase by 11.6 % from the same period in 2009 (PWC, 2010:13). On top of this direct economic significance, search engines are crucially important both for Internet users looking for information and for businesses who want to sell goods and services to those users. Although information is abundant on the World Wide Web, users' attention is limited. Search engines serve as gatekeepers between the information and services provided online and searchers' attention.

The market for Internet search is not only economically and socially important, it is also highly concentrated. In the US, Google had a market share of 66.2%, Yahoo of 16.4%, and Bing of 11.80%, as of November 2010. In the UK, just as in many other European countries, Google had a market share of 90.83 %, Yahoo of 3.21%, and Bing of 3.12%, as of December 2010.[2] Note that Bing is Microsoft's search engine and that in early 2010 Yahoo in effect sold its search and search advertising business to Microsoft.[3]

Is high market concentration a problem? For many, the main concern is Google's seemingly dominant position and the market power that comes with it.[4] Indeed, several competition authorities have opened (and in some cases, closed) investigations into the market behavior of Google. Most notably, on 30 November 2010, the European Commission launched a formal inquiry into allegations that Google has been manipulating the results of searches in ways that unfairly benefit it. Other matters up for investigation include the ad pricing scheme applied to rival providers of online services, as well as exclusivity obligations imposed on advertisers. (Related complaints have been brought before the French, Italian and German competition authorities.) On 8 April 2011, the US Department of Justice and Google proposed a settlement in the case of the intended acquisition of ITA Software Inc. by Google. The Department required Google to further develop, firewall, and license ITA's shopping software under a rather intrusive firewalling and monitoring regime. On 24 June 2011, the Wall Street Journal indicated that the US Federal Trade Commission was to open a formal investigation of Google's business practices, a move since confirmed by Google Inc. in regulatory filings.

On the other hand, given the characteristics of the search engine market, in particular the very modest switching cost for users, it is not obvious that high market concentration is detrimental to innovation and economic welfare and that government intervention should

[1] See `http://www.fastcompany.com/1695187/google-on-social-search-who-us-worried.`

[2] See `http://www.impactmedia.co.uk/blog/search-engine-news/search-engine-market-share-.`

[3] See `http://www.pcworld.com/article/204134/will_bings_backend_make_yahoo_better.html`

[4] There are other reasons for concern about the market for search engines, which we do not take as the primary objects of this study: media pluralism and privacy rights come to the forefront.

be considered. On 16 July 2010, following an extended series of articles on Google's market position, The Financial Times concluded: "it is better for different search engines to compete vigorously with each other to produce the best and most relevant results. Google may be highly successful in search but competition is only a free click away." This viewpoint draws on the economic theory of contestable markets, which holds that an incumbent with market power cannot exploit consumers as long as he has to fear that competitors would just step in and offer their services at lower prices.[5]

The appropriateness of this viewpoint is exactly the starting point of our research. In this paper, we do *not* try to answer the question whether Google has abused a dominant position. By contrast, we study the question whether it is true that "competition is only a free click away". In a crucial sense, this question is anterior to any discussion of monopolization, dominance, or abuse. If the market for search engines is highly contestable, then those discussions are moot: Google's high market share can then be taken as an indication that it is just the best search technology currently available.[6] Conversely, if the market is poorly or non contestable, then its structure, rather than the conduct of any of the participants, is likely to be problematic and to call for direct regulatory intervention rather than antitrust action.

Therefore, in this paper, we do not study, or take a stance on, the allegedly abusive behavior of search engines in general or of Google in particular. Rather, we try to understand what determines the quality of a search engine as perceived by users. On this basis, we provide a simple model of quality-based search engine competition. Interestingly, our model can be interpreted in terms of R&D investment and innovation. We argue that the current market structure is not stable and that the search engine market displays a strong structural tendency towards monopolization. This has negative effects on expected average search quality, the rate of innovation, consumer surplus, and total welfare.

Our key insight is that the production of search quality is characterized by a peculiar (intertemporal) type of *indirect network externalities*. We argue that such indirect externalities arise on the market for search engines because users will not consider, when deciding whether to run another query, that the results of their query and subsequent clicking behavior on suggested links are stored by the search engine. Currently, this information, also known as *query logs* or *search logs*, is not public. Only the search engine that is used to run the query can aggregate it with the information gained from other users who entered a similar search keyword. Thereby, it can improve its guess as to what future users—on average or with certain revealed characteristics such as geographical location or language—are looking for when they enter a certain keyword. This translates into higher search quality perceived by users.

The importance of large amounts of query log data for producing search engine quality has been widely acknowledged by the computer science literature (Spink et al., 2001; Jansen, 2006; Grimes et al., 2007; Bar-Yossef and Gurevich, 2008; Levy, 2009; Silvestri, 2010): Access to more search log data today leads to higher perceived search quality. Higher perceived search quality

[5] See Baumol, Panzar and Willig (1982) for a classical exposition.

[6] Whether the behavior of a market participant is susceptible to give rise to antitrust liability remains of course of prime *legal* interest even in that case but there is no reason to expect the enforcement of competition law to lead to a significantly different market outcome in the medium run.

leads to more demand for searches tomorrow, which in turn creates even more search log data tomorrow than today. This mechanism is at the core of the model that we propose to analyze competition in the search engine market.

This implies for a competitive market, taking everything else equal, that a firm that had a modest lead in market share at some point in time can increase that advantage more and more, while the other firms' market shares decrease more and more. That is, the market 'tips'.

To remedy this situation, we put forward and analyze the following policy proposal: *All search engines should be required to share their (anonymized) data on clicking behavior of users following previous search queries.*

We study the resulting market structure, coined a "competitive oligopoly, and show that it dominates alternative market structures in all the above mentioned dimensions. If we compare it to today's market structure, the disadvantage of the competitive oligopoly is that the costs of producing quality ('investments') are multiplied. However, we can show that the competing firms would have no incentive to decrease their quality below the quality of today's dominant firm (Google). In addition, *all* users would benefit from high quality, and *all* search engines' incentives to innovate would remain consistently high.

Although the implications of the externality we uncover have not been stressed in the academic literature so far, the point itself is perceived by practitioners. For instance, when the US Department of Justice cleared the alliance between Yahoo and Microsoft in February 2010, it stated that Microsoft's access to Yahoo's search query data would speed up Bing's automated learning, helping the search engine return more relevant results, in particular with respect to rare queries.[7] However, to our mind, the ultimate conclusions from this line of reasoning have not been drawn.

The next section discusses search engine quality, learning effects and network externalities in more detail and relates our paper to the existing literature. In Section 3, we introduce a simple model of search engine competition and solve for the equilibria of a row of market structures that are likely to follow upon each other without governmental intervention. In Section 4, we state our main policy proposal and show that (and why) a competitive oligopoly dominates market structures without intervention. There, we also discuss the technical feasibility, legal basis and practical implementation of our proposal. We conclude by outlining the trade-offs involved in any public intervention in this high-tech sector in Section 5. Technical details and computations are relegated to an Appendix.

2 Search Engine Quality, Network Externalities, and Related Literature

2.1 Search Engine Quality and Network Externalities

To develop an appropriate model of search engine competition, we will first try to understand the question, what actually is search engine quality? In a recent survey asking more than 1,100

[7] See US Department of Justice (2010).

software testers from more than fifty countries for the "most important attribute in choosing a search engine", 71 % chose "overall accuracy of search results".[8] Other important attributes mentioned were "page load speed" and "real-time relevance". In the ranking made up in that survey, Google came first in all three categories, followed by Bing (second) and Yahoo (third).

Based on these survey results, search engine quality may be proxied by the expected time a user needs to obtain a satisfactory result to his search query. What determines this expected time? It depends on the inputs into search, namely, (i) the sophistication of the search algorithm (algorithm quality), (ii) the computer power of the server farms searched by the algorithm (hardware quality), and (iii) the amount of potentially relevant data that the algorithm can search through (data quality). Data quality has several parts, among them (iii.a) the amount of raw data available online (virtually all data stored on the Internet, at least on the World Wide Web) and (iii.b) the context-specific data created by previous searches of a keyword and subsequent clicks by other users.[9] Notably, whereas everybody who has sufficient hardware quality can virtually copy the Internet, which makes (iii.a) *public* information, the context-specific data created by previous searches is kept secret by the search engines and, thus, constitutes valuable *private* information.[10]

Hardware quality (ii) and the amount of raw data available online (iii.a) can be taken to be equal across competitors with sufficiently deep pockets. Consequently, differences in users' perceived quality of search engines, as reported above, have to be rooted in differences in algorithm quality (i) and access to context-dependent data (iii.b).

Let us take the sophistication of the search algorithm (i) fixed and consider a market where there is some private data (iii.b) to be searched. Note that (i) is potentially unequal across search engines (and probably is in practice). In this paper, however, we want to isolate the effect that the existence of some private data (iii.b) has on the market equilibrium and abstract from the additional effects of heterogenous algorithms. Moreover, it is an open issue how important access to query log data (iii.b) is for the production of search quality relative to the other input factors (i)-(iii.a). However, we think we made an important step towards answering this question by showing that a model that entirely focuses on differences in access to query log data—and holding all other dimensions of competition equal—can generate several empirically observed developments in the search engine industry.[11] We will return to these points in the

[8]See the detailed report at `http://www.utest.com/bugbattle/q309`.

[9]Search engines also seek to gather other types of data such as information about a user's profile and preferences by monitoring click-through rates to contents or online purchasing behavior, a point stressed by Spulber (2009). To some extent, Google's business model has recently consisted in providing users with free ancillary applications (email, calendar, personal page management system, browser, etc) which generate traffic in the first place and can be used to collect such individual-level information. We do not think, however, that this information is subject to network externalities of the type we stress in this paper, as it is specific to a particular user.

[10]See Levy (2009) for a lively account of the importance of such data. Spink et al. (2001) noted that query logs "provide a snapshot for comparison of public behavior while searching, a behavior that can also serve as a clue for improvement of search engines." It is technically possible to sample search engines' drop-down suggestion menus as a way to indirectly mine their query logs of a particular engine but this is no substitute for directly processing those logs. See e.g. Bar-Yossef and Gurevich (2008).

[11]See subsection 3.6 for a comparison of several key predictions of the model with reality.

conclusion, Section 5.

We posit that the context-specific data created by searches of a keyword and subsequent clicks by users is not only private but is also characterized by indirect network externalities. If a search engine user decides whether to make another query, she trades-off the expected private benefits (getting accurate search results) and the expected private costs (time/opportunity costs) of that query. However, the user will not take into account that her query and her subsequent clicking behavior on this or on that search result provides the search engine with information about user preferences in general. This information allows the search engine to produce search results to the same keyword that are more likely to meet future users' preferences and, thereby, to increase its *perceived* search quality.

Based on this insight, we study the following questions. What are the effects of network externalities on competition among search engines? What are the effects of competition among search engines on economic performance (rate of innovation, consumer surplus, and welfare)? Do these results warrant intervention of public authorities? If so, what type of intervention would be both appropriate and feasible?

2.2 Related Literature

On a theoretical level, there is now a voluminous literature on network externalities.[12] This literature stresses early-mover advantages in the launch of network goods, natural market dominance (competition *for* the market and "market tipping"), and the strategic importance of compatibility decisions. Markets with network goods are affected by various instances of market failures: consumers might coordinate on the wrong platform or fail to switch to a more efficient one when it becomes available; conversely, firms might fail to coordinate on a unique standard, which prevents consumers from fully taking advantage of network externalities. In addition, dominant firms may fail to provide the right network scope. Public authorities may try to correct or alleviate those market failures but they face an especially hard task, given the specific nature of network industries.[13] In this paper, we will indeed stress the complexity of the sector but argue that a specific, limited public intervention is possible to restore a level playing field and increase economic welfare.

Direct network externalities usually arise on the demand side, as in the usual example of a communication network. Consumers derive more utility from the network, the higher the number of other consumers who choose to connect. Interestingly, the indirect intertemporal externalities we describe can also be viewed as a special version of the *learning curve hypothesis*, a supply-side phenomenon. Economists speak of a learning curve when the current cost of production negatively depends on the cumulative amount produced in the past. This idea dates back at least to Arrow (1962), and the Boston Consulting Group helped popularize it in the 1970s by stressing the importance of being a first mover and investing in market share (Boston Consulting Group, 1972). It is a well-known phenomenon, early observed, for instance, in the

[12]Good points of entry include Besen and Farrell (1994), Economides (1996) or Shapiro and Varian (1999).

[13]See Economides (2010) or Gandal (2002) for detailed analyses of the policy implications of this line of research.

production of aircrafts.[14] For search engines, previous queries are instrumental in decreasing the cost of producing quality today. An interesting twist is that here, contrary to the case of aircraft manufacturing, it is not the *intangible* process of production of earlier units in a particular plant that gives rise to learning but *recorded* information about past searches. Thus, we have a new type of learning effects whose very nature is not firm- or worker-specific but can actually be shared. This feature motivates our policy proposal (see Section 4).

This proposal consists in mandating exchange of query logs across search engines, which amounts to setting up an interorganizational information system, and is related to the literature on electronic data interchanges (EDI's). Prominent papers have focused on the case where a single buyer tries to organize information exchange with several suppliers. In our case, the exchange would be *reciprocal* between all search engines. Of importance is the finding that competitive effects from sharing information typically preclude universal voluntary adoption by rival firms (Riggins, Kriebel and Mukhopadhyay, 1994, and Wang and Seidmann, 1995). Search logs have long been recognized as a rich source of information that can be used to improve search efficiency (among other things): see Jansen (2006) or Silvestri (2010) for accessible introductions.

The search engine industry, although new, has already given rise to a number of academic studies. See Varian (2006) and Spulber (2009), in particular. Evans (2008) and Ratliff and Rubinfeld (2010) look at the larger online advertising industry. Most technical papers have looked at some of the newest aspects of the sector (as compared to standard yellow pages), in particular the disclosure of sponsored links alongside so-called organic results (Ghose and Yang 2009, Taylor 2010, Yao and Mela, 2011), the keyword auctions introduced to price them (Chen and He, 2006; Edelmann, Ostrovsky and Schwarz, 2007; Varian, 2007; Athey and Ellison, forthcoming), or the substitutability between search engine advertising and the more traditional types of advertising (Goldfarb and Tucker, 2011a,c).

Gandal (2001) and Telang, Mukhopadhyay, and Rajan (2004) were early attempts at predicting the general dynamics of the industry, which was then still in its infancy. Those papers do not focus on, or identify, network effects. Neither does Pollock (2010), who argues that the search engine industry will continue to evolve down its path towards monopoly with deleterious consequences on welfare. Pollock also calls for regulatory oversight, in particular the vertical separation of the "software" and "service" divisions of search engine firms. Such an intervention, affecting the organizational or capital structure of the industry, has wide implications and we believe that our more modest proposal is easier to implement.

[14]See Wright (1936) or, more recently, Benkard (2000). The literature on the learning curve is voluminous. Recent noticeable contributions include (but are not limited to) Adler and Clark (1991), Zangwill and Kantor (1998) or Wiersma (2007).

3 The Model

3.1 From Stylized Facts to Modeling

Search engines do not charge end-users for running queries. Nevertheless, some are highly profitable.[15] Most of their revenues come from selling (targeted) advertisements related to search queries and displayed as "commercial results" or "sponsored links" next to the so-called "organic links", which are the results to a search query generated by the search engine's algorithm.

Over time, successful search engines have developed particular ways to select and price advertisements. The celebrated Google keyword auction (AdWords) is a prime example. Clearly, search engines are in the business of matching potential buyers (searchers) with sellers (advertisers). Thus, the theory of intermediaries, or *two-sided markets* in the economic jargon, may be of relevance.[16] We do not want to downplay this important feature of the industry. However, it seems to us that the fact the price on users' side is always set to zero has become a strong norm in the industry. In addition, in the literature, it is a controversial question whether sponsored links bring direct positive utility to consumers (Chen and He, 2006; Athey and Ellison, forthcoming) or not (Edelmann, Ostrovsky and Schwarz, 2007; Taylor, 2010).[17] As a result, the two-sidedness of the market may not be pronounced.[18] We thus follow an agnostic approach and assume that consumers neither derive positive nor negative utility from sponsored links.[19] As we are not primarily interested in the advertising side, we assume that the advertising revenue of a search engine is proportional to its market share, which is broadly in line with industry's figures. That is, advertisers are willing to pay a given price for "eyeballs" to search engines. This does not detract from the fact that ever better targeted advertising is an important margin of innovation and competition in the industry. It is simply not the focus of attention in this paper.

The search engine market is characterized by two additional important features: some degree of horizontal product differentiation, and incomplete information of consumers about search engine quality with respect to a specific keyword. In a recent report, users were asked how they perceive the major search engines to be different.[20] Google, for instance is perceived to have a very simple design that completely focuses on their search box. In contrast, Bing's "feature-filled image and video search" is recognized. Moreover, "as opposed to Google", Yahoo's "auto-recommendation bar [...] offers smart recommendations". In addition, Jansen, Zhang, and

[15] Not all search engines are profitable. As of March 2010, "reports indicate that Redmond [Microsoft] has spent as much as $5bn on its nine-month-old search revamp, and the company continues to pour cash into the would-be Google killer" (`http://www.theregister.co.uk/2010/03/02/ballmer_on_bing/`). Compare this to the performance of search engines 2 and 3 in our model below, as displayed in Figure 2.

[16] Varian (2006) and Evans (2008) insist on that point. See Rochet and Tirole (2006) for a technical overview of the theory. Rochet and Tirole (2007) introduce several less technical papers.

[17] Goldfarb and Tucker (2011b) show that targeted advertising leads to negative externalities, only when obtrusive.

[18] The French Competition Authority's online advertising sector inquiry (*Avis* 10-A-29 of 14 December 2010) reached the same conclusion.

[19] Positive indirect network externalities on the users' side would only reinforce market tipping.

[20] See the detailed report at `http://www.utest.com/bugbattle/q309`.

Ying (2007) show that the relevance of results, as rated by users, is influenced by the name of the search engine, thus pointing to the importance of branding.

The perception of search engine differentiation leads to the fact that, before running a specific keyword/picture/video search, consumers do not exactly know which search engine delivers the highest search quality for the item they are interested in. This uncertainty partly stems from an asymmetric and unknown distribution of previous queries across search engines. Hence, before running a specific search it is impossible to say whether this or that search engine delivers the highest quality results. Consumers can only rationally expect search engine quality to be different on average. Therefore, an appropriate model of the search engine market has to deliver "smooth" outcomes, where a firm's market share is monotonically increasing in its perceived average search quality but where even a search engine with low perceived average quality has a positive market share.

3.2 A Contest Model

As shown in the introduction, today's search engine market features one dominant firm that competes with two notable competitors.[21] Therefore, we start by modeling a triopoly market with firms 1, 2, and 3.

On the demand side, in our model there is a unit mass of consumers, each of which has demand for one query. As has been business practice since the birth of the industry, nominal prices for using a search engine are zero. We assume that the market is fully covered, i.e. that quality is high enough for every consumer to use a search engine.

It is natural to appeal to a class of models that allow consumers to choose which search engine they want to use, on the basis of those engines' characteristics, in particular their perceived quality level, as well as consumers' own preferences. Discrete choice models of product differentiation allow consumers' utility to be randomly shocked in a way that makes sure that all engines capture at least a fraction of demand.[22] In such models, consumers do not know in advance which product they will consume and it is possible to compute clear measures of economic welfare, such as expected consumer surplus. To simplify the presentation, we take a shortcut and work with a model that fits the stylized facts outlined above (as well as the properties of the functional form for market shares that would be derived from standard discrete-choice models). We model competition as a contest among search engines with simultaneous bids x_i, where x_i is firm i's search quality; $i \in \{1, 2, 3\}$. That is, each search engine simultaneously and independently chooses its quality level. The market share of firm i is then given by $D_i = \frac{x_i}{\sum_{j=1}^{n} x_j}$, where $j \in \{1, ..., n\}$ and $n = 3$ firms are active in the market. Production of quality x_i comes at a cost, $C(x_i) = \frac{x_i}{N_i}$, where N_i is the "installed base" of firm i, i.e. the amount of previous search queries run on i. Without loss of generality, we assume $N_1 \geq N_2 \geq N_3 \equiv 1$. Moreover, each firm bears a fixed cost F for its operations.

[21] Since 2010 Yahoo has used Bing's search technology and does not invest in its own search technology anymore; see the reference in Footnote 3. In the analysis below, we discuss the incentives to exit the market and to merge for search engines 2 and 3.

[22] For an exposition of those models, see Anderson, de Palma, and Thisse (1992).

This formulation interprets x_i as the quality of search engine i *perceived* by consumers. The cost to create a certain level of perceived quality depends on the resources spent on improving the search engine's algorithm and on the amount of private data accessible, which makes it cheaper to produce any quality level. It is as if search engines learned to produce quality more cheaply by using their stock of past queries.

Following these assumptions, search engine i solves the following program:

$$max_{x_i} \quad \pi_i = \frac{x_i}{x_1 + x_2 + x_3} p - \frac{x_i}{N_i} - F, \tag{1}$$

where p is the exogenously given advertising revenue associated to one consumer.[23]

3.3 Analysis of the Triopoly Case

It is possible to characterize the behavior of each search engine and solve for the equilibrium. For details, we point to Appendix A.1, where we display the relevant computations. Figure 1 displays equilibrium quality levels and market shares as a function of N_1, the amount of private data to be searched by firm 1.

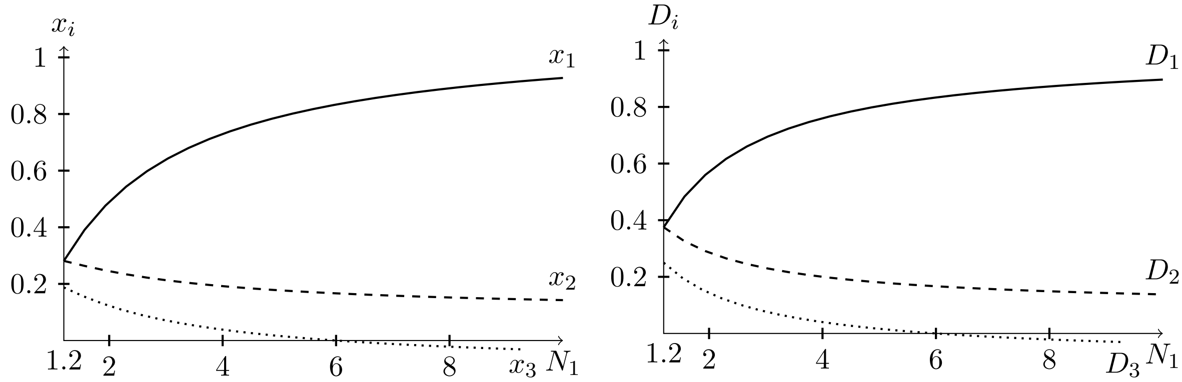

Figure 1: Numerical example for $p = 1, N_2 = 1.2$. LEFT: equilibrium quality; RIGHT: market shares.

Interestingly, although we model a simple one-shot game, N_i can be interpreted in a dynamic way: the market share of firm i in some period t influences the relative amount of private data that firm i has access to in period $t + 1$. Consequently, if firm 1 has access to more data than firm 2 and 3 at one point of time, the equilibrium predicts that *ceteris paribus* this advantage increases over time.[24]

[23] In a first approximation, the marginal cost of running an additional query can be taken to be zero. In any case, variable costs can always be subsumed into the p variable, which would then stand for net revenue per user. This formulation implicitly assumes that quality affects the fixed cost of production rather than the variable cost. This is likely: variable costs mostly come from the huge energy requirements needed to run server farms, whereas quality is directly related to the work of engineers and software developers.

[24] We are aware of the fact that a static model is not the best tool to study market dynamics! The dynamic extension of our model is straightforward to work out, especially if there is no quality persistence over time. We use the static model to illustrate our main ideas. The search engine industry arguably meets the conditions identified by Gans (2011) for static analysis to be a sufficient proxy for dynamic considerations in innovative industries.

This implies that the market is *tipping*, i.e. firm 1 is producing ever higher perceived quality—and gains ever higher market share—whereas firms 2 and 3 decrease their respective quality levels—and market shares—over time. However, the survival of the weakest search engine, firm 3, is called into question as soon as the data access advantage of firm 1 or 2, expressed by N_1 and N_2 respectively, is sufficiently large. It is indeed costly to produce quality and advertising revenues, and thus market share, must be high enough to maintain positive profits. Therefore, the model predicts that firm 3 sooner or later exits the market, and the market structure turns from a triopoly into a duopoly.

Before turning to the duopoly case, however, let us analyze the welfare effects of network externalities in the triopoly model. Equilibrium profits (π_i) are easily calculated by substituting equilibrium quality levels into (1). Producer surplus (PS) is the sum of all firms' profits. Expected consumer surplus (CS) is found by averaging equilibrium quality levels weighted with the market shares of active firms. Summing up producer surplus and consumer surplus, we get total surplus (W). Figure 2 displays a graphical presentation of firms' profits and total welfare as a function of N_1.

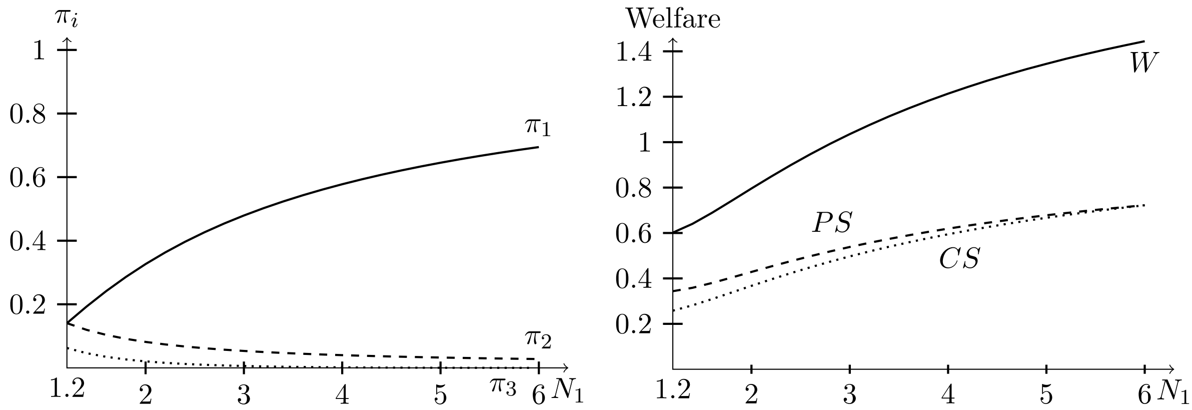

Figure 2: Numerical example for $p = 1, F = 0, N_2 = 1.2$. LEFT: equilibrium profits; RIGHT: welfare.

This figure shows that firm 1's profit increases in N_1, which implies that, due to the argumentation outlined above, its profit also increases over time. Instead, firm 2's and firm 3's profits decrease over time. Hence, in our model it is very profitable to be the market leader and to maximize the advantage over competitors in accessing (private) data about past queries.

Notably, consumer surplus and welfare are also increasing in N_1 (and in N_2). This effect stems from the fact that network externalities decrease the cost of producing quality: more and more consumers enjoy the increasing quality of the market leader.

3.4 The Duopoly Case

If firm 3 exits the market, each of the remaining firms 1 and 2 solve the following simplified program, (1):

$$max_{x_i} \quad \pi_i = \frac{x_i}{x_1 + x_2} p - \frac{x_i}{N_i} - F \tag{2}$$

For simplicity and without loss of generality, we set $N_2 \equiv 1$. It is again straightforward to solve for a unique equilibrium in quality levels with positive market shares. Results (see Appendix A.2) show that the duopoly market is developing in the same way as the triopoly market: the market tips towards a single search engine. In line with the ever decreasing market share of firm 2, for every $F > 0$, the profit of firm 2 turns negative if N_1 is sufficiently high. It follows that our model predicts that a duopoly, too, is not a stable market structure in the long run if the market leader retains its advantage regarding access to private data. In effect, the industry has the character of a natural monopoly. Therefore, we are led to study equilibrium and welfare effects in the monopoly case.

3.5 The Monopoly Case

There are two variants of the monopoly model with potential applicability to the search engine industry. Again, see Appendix A.3 for computations of both. The benchmark result is the case of a *pure* or *uncontested monopoly.* We show that a monopolist who does not fear any competition sets the quality to a minimum (acceptable) level. Still, consumers would use this search engine, in the absence of alternatives, which would drive up the monopolist's profit. Unsurprisingly, this situation would be dismal for consumer surplus, total welfare, and the monopolist's incentive to invest in quality, i.e. to innovate.

Importantly, our results imply that the extent of network externalities virtually do not play any role as long as the market is monopolized and no potential entrant is in sight. The monopolist has no incentive to produce positive quality and, therefore, its cost advantage created by access to private data is not exploited in equilibrium.

The second—and more realistic—variant of the monopoly model is the case of a *contestable monopoly.* Here the monopolist faces a potential (re-)entrant but may be able to prevent entry by making use of an appropriate "limit strategy". To study this case we consider the following sequential game, which we solve in Appendix A.3. First, the incumbent, firm 1, sets quality x_1. Second, the potential entrant, firm 2, decides whether to enter the market for a cost $K \geq 0$, or not. Third, if firm 2 entered the market, it sets quality x_2. Finally, consumers choose a search engine, as before.

Given this structure, firm 1 can predict how firm 2 would behave—what quality it would produce—if it entered the market at the second stage. Therefore, firm 1 can calculate what entry would mean for its own profits and decide whether it wants to deter or accommodate it. The key trade-off that arises for firm 1 is whether it wants to set a rather high quality, which we denote by x_1^{lim} in the Appendix, and enjoy monopoly, or set a rather low quality, $x_1^{Stackelberg}$, which saves on cost but does not necessarily foreclose the market.

We find that, if the advantage in private data access of firm 1 is sufficiently large (i.e. if $N_1 \geq 2$), then firm 2 cannot enter the market *and* gain a positive market share even if firm 1 sets $x_1^{Stackelberg}$. Hence, in this case it is not even necessary for firm 1 to foreclose the market by making use of a limit quality strategy.[25]

[25] Even in this case firm 1 cannot set quality as low as in the absence of competition (i.e. $x_1 = \epsilon$) because then firm 2 could set a higher quality and capture the entire market demand, thereby making a profit.

In contrast, if firm 1's advantage in private data access is modest (for $N_1 < 2$), we find that firm 2's cost of market entry, K, becomes crucial. If K is rather low, firm 1 sets x_1^{lim}. In this case, the entrant knows that it would also have to set a rather high quality to compete with the incumbent. This high quality would be so costly for firm 2, however, that it could not profitably operate. It follows that firm 2 abstains from entering the market if the incumbent set the limit quality, x_1^{lim}. Finally, if K is high, firm 1's best choice is to set $x_1^{Stackelberg}$. This implies that firm 2 could indeed enter the market and gain a positive market share. However, it could not recoup the market entry cost with the modest operational profits this would bring. Understanding this, firm 2 would not enter the market in this case, either.

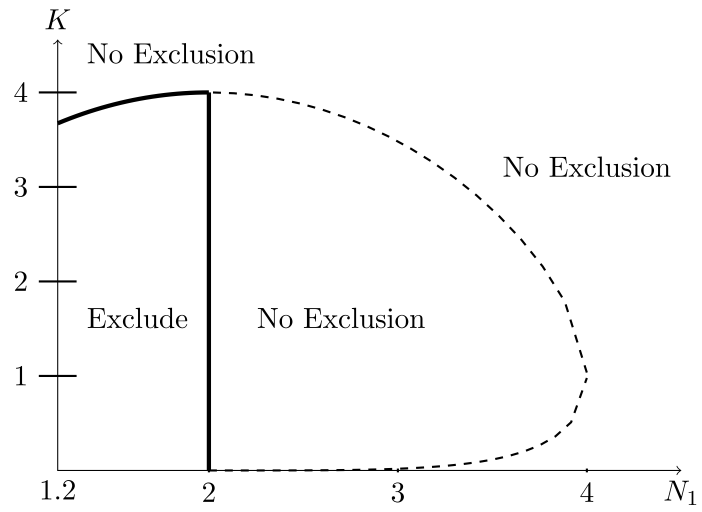

Figure 3: Numerical example for $p = 1, F = 0$. (Non)exclusion incentives in a contestable monopoly.

Figure 3 maps all applicable constraints. Firm 1 plays $x_1^{Stackelberg}$ unless both N_1 and K are low, in which case it actively forecloses the market by setting the limit quality, x_1^{lim}. The outcome of all parameter combinations is the same, though. A monopoly is a stable market structure.

3.6 Reality Check

The model generates several testable implications. First, it predicts market tipping (see the right panel of Figure 1): the market share of the dominant firm is expected to increase more and more, whereas the market shares of the other firms are expected to decrease. Second, the model predicts substantial profit growth for the market leader and decreasing profits for the following firms (see the left panel of Figure 2). Third, as a consequence of the second point the model predicts market exit of one follower.

These predictions are well reflected in the history of the search engine market since 2003, when Google became the market leader. The actual development of market shares of the top three search engines in the US is displayed in Figure 4. There, Google takes over the role of firm 1, whereas Yahoo and Bing stand for firms 2 and 3, respectively.

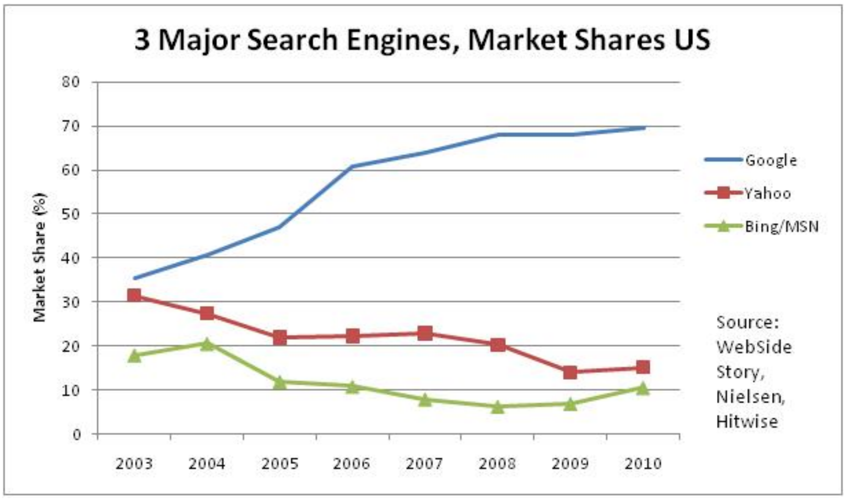

Figure 4: US Market Shares in Search Referrals to Other Websites.

Although starting from roughly the same market share as Yahoo in 2003, Google managed to increase its US market share up to nearly 70% by the end of 2010.[26] In line with that success, Yahoo's and Microsoft's search engines (first MSN, then Bing) halved their market shares—from a combined 50% in 2003 to 25% in 2010. The model's second prediction also corresponds to the empirically observed development of profits: while Google's profit rose by 36% to 2.51 billion dollars in the second quarter of 2011,[27] Bing seems to lose money.[28] Finally, the model's third prediction is reflected by the de facto market exit of Yahoo, which sold its search and advertising business to Microsoft in early 2010. We conclude that, although our model is simple and stylized, it can reproduce several key developments that are well in line with the latest developments in the search engine industry.

4 Policy Proposal

We have shown that, absent any other major changes, network externalities in the search engine market can be expected to drive out competitors of the dominant firm and to lead to a stable monopoly. Due to the adverse welfare effects that this development would have (see below for details), we state our policy proposal now:

All search engines should be required to share their (anonymized) data on clicking behavior of users following previous search queries among each other and among new entrants!

In the following, we will first analyze the economic case this proposal is based on. Second, we will discuss the technical and practical feasibility of the proposal. Finally, we will sketch the potential legal basis, provided by regulation and competition law, that could justify the

[26]The development of market shares on a global level followed a similar pattern in the last decade. The main difference was that Google's dominance was even more pronounced there, reaching about 85% market share at the end of 2010.

[27]See `http://financesjournal.com/taxes/record-profits-google-stock-shoots-7114.html`.

[28]See the reference in footnote 15.

proposed market intervention.

4.1 The Competitive Oligopoly Case

Consider the triopoly market studied above and assume, for now, that our policy proposal could be implemented without frictions. We call this the competitive oligopoly case.[29] Technically, our proposal implies that all search engines get access to all context-specific data available in the industry, N, where:

$$N \equiv N_1 + N_2 + N_3 \tag{3}$$

We solve this model for equilibrium quality levels and the resulting welfare measures in Appendix A.4.

Now we are in a position to compare the effect of our policy proposal on key variables with the respective effect of the market structures analyzed before. Figure 5 depicts the comparisons of consumer surplus and total welfare as a function of the installed base of the most successful search engine, N_1. There "monopoly" is based on the equilibrium values generated by the limit quality case.[30]

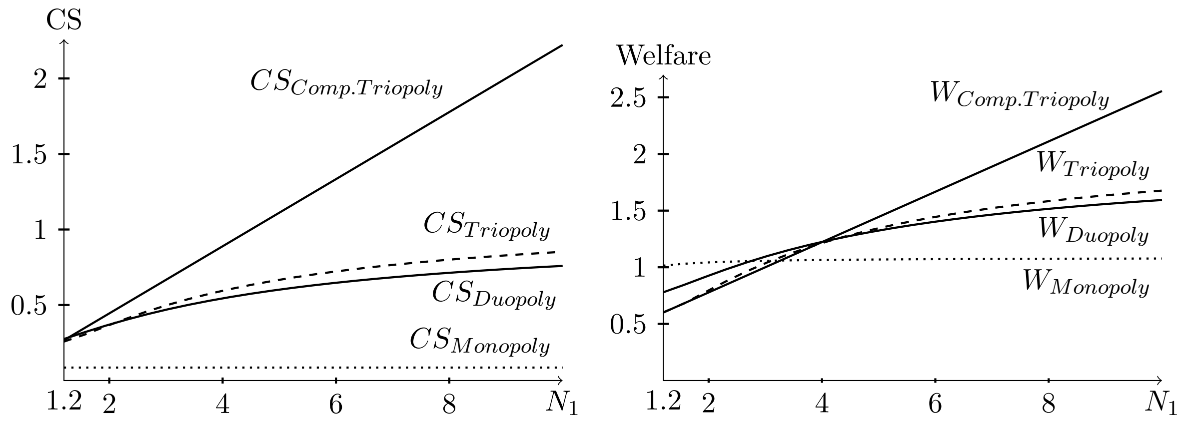

Figure 5: Numerical example for $p = 1, F = 0, N_2 = 1.2, K = 0.5$; LEFT: Consumer surplus, RIGHT: Welfare across market structures.

Figure 5 shows that consumer surplus increases in the level of competition (from monopoly to competitive triopoly) and in the extent of network externalities (apart from the limit quality monopoly case), depicted on the horizontal axis. It also shows that consumer surplus and welfare in market structures without intervention are of the same scale as the one of the intervention case as long as N_1 is sufficiently small. With growing N_1, however, the competitive triopoly case gets more and more attractive and, for high N_1, clearly dominates the other cases. The key trade-off under competitive triopoly is that, on the downside, it multiplies the fixed cost of producing quality because several firms are active in the market and spend on high quality

[29] The results are qualitatively identical whether we consider two, three, or more firms.

[30] As our analysis has shown, quality and welfare under an uncontested monopoly are even lower.

production. On the upside, the competitive triopoly makes sure that all consumers get high search quality, not only the consumers of the dominant firm.

Importantly, the reason for the superiority of the competitive oligopoly over alternative market structures is rooted in the high quality levels that intense competition produces. Figure 6 depicts equilibrium quality of firm 1 under several market structures.

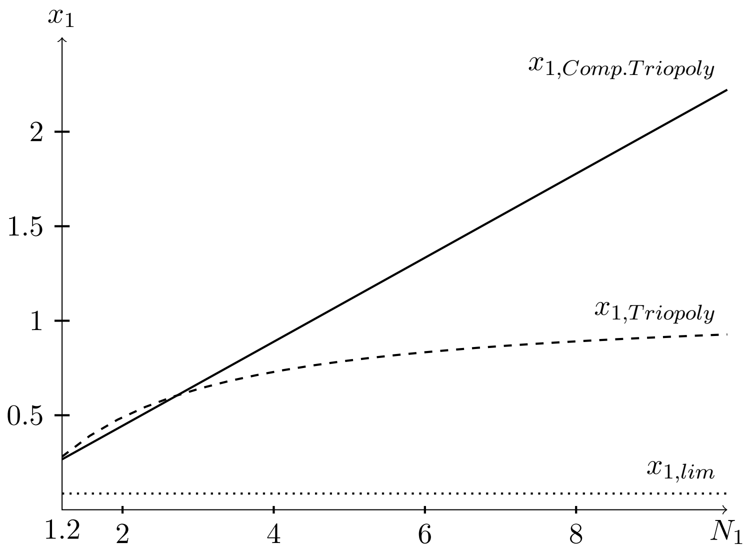

Figure 6: Numerical example for $p = 1, F = 0, N_2 = 1.2, K = 0.5$. Quality of firm 1 under different market structures.

The comparison of the current triopoly with the competitive triopoly is interesting. In both cases, quality increases in N_1—or over time, using our earlier interpretation. In the current triopoly, however, the dominant firm reduces its "rate of innovation" (proxied by the slope of its quality curve) when the market is more and more tipping in its favor. This effect does not exist under the competitive triopoly. There, no search engine can rest on its merits because the only way to sustain its market share and profits is to invest all efficiency gains that come from the exploitation of network externalities into better quality. As a result, producer surplus is unaffected by N, whereas consumer surplus monotonically increases in N.[31]

On the other hand, given that even under the competitive triopoly all three firms make positive profits (assuming that the fixed cost of operation is not prohibitive), it would be false to claim that equal access of all search engines to the complete pool of previous search data wipes out incentives to do business in this market or to enter it in the first place.

In this section, we have not taken into account the cost of our policy proposal, yet, which was assumed to be implementable "without frictions" above. We will discuss potential frictions below. Nevertheless, at this stage, it is important to note that our main theoretical result, the dominance of the competitive oligopoly over the alternative market structures analyzed, from a consumer surplus and total welfare perspective, qualitatively holds even if our policy proposal came at substantial transaction costs. In our model, such costs could be reflected by

[31]See (A.29) and (A.30) in the Appendix.

higher F. In the long-run, however, that is if N grows sufficiently large, which is inevitable as long as search engines operate and consistently collect data on user behavior, consumer surplus and total welfare increase to levels that are unrivaled by alternative market structures.[32] The reason is that intense competition between search engines based alone on the merits of the search algorithm provides better incentives to the firms to produce high quality products than the rent enjoyed by a dominant firm that exploits a competitive advantage created by network externalities.

4.2 Technical Feasibility and Practical Implementation of our Policy Proposal

Above, we have analyzed the economic effects of our proposal. But even if the positive economic effects are pronounced in theory, how realistic is our proposal?

Surprisingly, support for the technical feasibility of granting other parties access to large amounts of private data has come from an unexpected direction: Google. On October 13, 2010, Facebook, the world's largest online social network with more than 500 million users, and Bing, Microsoft's search engine, announced an alliance. At the heart of it lies the idea that Bing delivers search results on the basis of the recommendations of the searcher's Facebook friends, in addition to impersonal search recommendations. One day later, on October 14, Google's then-CEO Eric Schmidt commented on the deal with the quote put at the beginning of this paper: "There's always a concern that large private collections of data are not available to search engines... . We've taken a position, both in a religious and in a business perspective, that the world is better off if you take the information that you're assembling and make it accessible."[33]

This "position" is in line with the welfare superiority of the competitive oligopoly case analyzed above. Moreover, the statement indicates that it is technically feasible to "take the information that you're assembling and make it accessible" even if the "private collection of data" is "large".

Furthermore, it is straightforward to understand both the Bing-Facebook alliance and Google's reaction to it by drawing on our analysis: Bing has just increased N_2 (or decreased the distance $|N_1 - N_2|$). Google replied that they want higher N_1 (or to increase the distance $|N_1 - N_2|$). The motivation for both actions is clear if one inspects how the equilibrium profits of firms 1 and 2 under the current triopoly depend on N_1 and N_2.[34]

Next, apart from technical feasibility, how could our policy proposal be implemented effectively? Without going into too much detail, we perceive two avenues for implementation. First, if the number of major competing search engines is limited, as is the case today, it might work that data on previous searches is directly and bilaterally exchanged between search engines, under a monitoring regime. This would allow for a relatively timely delivery of the data

[32] All or a fraction of the old query data may stop being relevant after some time. Thus, it is possible that exponential decay or truncation impacts the evolution of N over time. That would not change the qualitative predictions of the model: a firm that has been dominant in the recent past would presently benefit from a competitive advantage.

[33] See `http://www.fastcompany.com/1695187/google-on-social-search-who-us-worried`.

[34] See equations (A.10) and (A.11) in the Appendix or Figure 2.

as no technical or administrative detours would have to be taken. What would be essentially needed is a body achieving standardization of the way search engine query logs are reported (and anonymized) along with a compliance regime.[35]

The second major way the data exchange could be organized is to create a new hub to which all search engines would have to deliver their private data. In exchange, they would get access to all other data saved in the hub. This is potentially costly as a new structure (and perhaps several of them for technical or legal reasons) would have to be set up and permanently operated, while all search engines may not have the means to collect and mine all data.

Which of these two ways is technically efficient would have to be determined by information and computer scientists, who understand the hardware, software, and network problems involved better than we do. In both cases, though, the enforcement and monitoring of the policy proposal would have to be given to a public agency—or several agencies, which could be located in several jurisdictions.[36] We leave the details of these issues, which combine computer science and public choice theory, for future research.

Note that, in absolute terms, our proposal implies that all search engines, including the market leader, gain extra information on users' clicking behavior. Hence, if, say, the picture search of one search engine is better than its search for travel-related information but another search engine has a comparative advantage in travel-related search, both search engines would obtain higher perceived quality in their weaker search domain after our proposal was implemented. In relative terms, of course, the market leader loses the advantage created by having access to more query log data—which is the very purpose of our proposal.

4.3 Sketching the Legal Issues

The possibility for public intervention in the market for search engines must be studied from a legal point of view. Here we only make a very modest start and sketch the major legal issues: the applicability of competition law and the protection of privacy.

The first issue of interest is whether competition law can be used to implement our proposal. Suppose that rivals of dominant search engines asked the latter to share their data regarding previous queries (possibly against payment). They might receive a negative reply which could be construed as a refusal to deal. Could such a refusal constitute a monopolization practice unlawful under Section 2 of the Sherman Act, for instance?

Case law provides for a set of limitative conditions under which a unilateral refusal to supply an essential input can be viewed as anticompetitive and call for antitrust relief.[37] On the basis of US Supreme Court cases *Otter Tail* and *Aspen*, US lower courts have recognized a duty for a firm with monopoly power to deal with its rivals under certain conditions.[38] Those include

[35]Adar (2007) and Menon and Sarkar (2007) explore ways to anonymize large data sets such as query logs.

[36]We are aware of the potential public choice difficulty that handing over enforcement and monitoring to a public agency could create. This is why we suggest to have at least two technical and institutional infrastructures do this job, which should be ideally placed in at least two jurisdictions. Competition between (or at least duplication of) agencies may preempt any attempt to abuse any one agency's power.

[37]We here follow Elhauge and Geradin (2007).

[38]See Otter Tail Power Company v United States, 410 US 366 (1973) and Aspen Skiing Co v Aspen Highlands

(beyond the fact that the monopolist indeed denied use of the essential facility to a rival) (i) the fact that the property to which access is sought is indispensable to the production process, (ii) the fact the facility cannot be easily duplicated and (iii) the fact that sharing the facility is feasible. There is quite some uncertainty about this line of case law, which has always been criticized by scholars and which the US Supreme court has always made a point of never endorsing as such.[39]

EU case law, while different, provides for similar conditions. Those were laid out in the famous *Commercial Solvents*, *Oscar Bronner* and *IMS* cases and their limitativeness apparently got confirmed in the *Microsoft* case.[40] They include (i) the indispensability of the property to which access is sought, (ii) the fact that the refusal to grant access is susceptible to exclude effective competition on the downstream market, and (iii) the fact the refusal has no objective, legitimate justification. An additional condition, in the case of intellectual property, derived from *IMS*, seems to be that rivals use it to introduce a new product to the market but this requirement seems rather formalistic following *Microsoft*.

We take the criteria in turn to determine whether there is any chance to find a legal basis for our proposal in this case law.

On essentiality: It is hard to argue that access to context-dependent data generated by previous queries is *absolutely* indispensable. It makes it easier for engines to return relevant answers by giving them data scale (and immediacy) but it certainly does not preclude them from investing heavily into improvements in the sophistication of the search algorithm and/or in the collection of other kinds of data.[41] At the same time, in *Aspen* and *Kodak*, the US Supreme Court recognized a duty to deal even in the absence of true essentiality.[42]

On duplication: Duplication in US case law plays the same role as the effective competition test in EU law. Both criteria aim at capturing the ability of competitors to exert meaningful competitive pressure on the owner of the essential input in case access is denied. In the context of search engines, duplication is clearly impossible. It is not a matter of money: a search engine can generate voluminous query logs only if end users massively choose to use it to place search queries now, which may well be impossible if the data needed to return relevant results are simply not there.

On feasibility: This US criterion has been interpreted as meaning more than technical

Skiing Corp, 472 US 585 (1985).

[39]See, most recently, Verizon Communications v Law Offices of Curtis V. Trinko, 540 US 398 (2004).

[40]See Commercial Solvents: Cases 6 and 7/73, *Commercial Solvents and Others v. Commision*, 1974 ECR 223; Bronner: Case 7/97, *Oscar Bronner GmbH & Co KG v. Mediaprint Zeitung und Zeitschriftenverlag GmbH & Co*, 1998 ECR I-7791; IMS: Case C-148/01, *IMS Health GmbH & Co OHG v. Health GmbH & Co KG*, 2004 ECR I-5039; Microsoft: Case T-201/14, *Microsoft Corp. v. Commission*, 2007 ECR II-3601.

[41]See, however, Grimes, Tang, and Russell (2007), who argue that, although engines can use field or lab experiments and instrument online panels to acquire data, query logs contain "irreplaceable information for understanding the scope of resources that a search engine needs to provide for the user".

[42]*Aspen* was cited above; see Eastman Kodak v Image Technical Services, 504 US 451 (1992). In *Aspen*, the case revolved around the possibility for one ski domain manager to issue bundled tickets giving access to all mountains in one ski resort, including some managed by another company. In *Kodak*, 18 independent service providers were complaining that Kodak was restricted access to replacement parts needed to service Kodak equipment.

feasibility to include economic feasibility. That is, a duty to deal was denied whenever there seems to be an objective justification for the refusal such as efficiency, customer satisfaction, cost reduction, quality, etc, as the EU allows. Technical complications do not make data exchange impossible in our view, as discussed above. Exchange would also be non-rivalrous in the sense that granting access to competitors would not undermine the ability of dataset owners to use them so as to serve their customers. Thus, there seems to be little room to claim that *ex post* efficiency reasons objectively justify a refusal to deal. Of course, a defense could be mounted with regards to the deleterious effect that any duty to deal would have on *ex ante* investment incentives. That is true of any essential facility case. Notice, however, that the data advantage of a dominant firm on the market for search engines is a *by-product* of its past success: even if it had known that it would later be subjected to a duty to deal, it would not have refrained from returning as relevant results as possible to users' queries.

All in all, it appears difficult to invoke the essential facility doctrine unless one stretches the (disputed) interpretation of the essentiality requirement. In addition, antitrust remedies can apply only to firms which have been found to be in a dominant position on the relevant market in the EU, or to have monopoly power for monopolization to be actionable under Section 2 of the Sherman Act in the US. Hence, only dominant search engines could face an obligation to share information about previous queries under antitrust laws; those are therefore unfit to allow public authorities to run a *universal* data-sharing scheme.

For this reason, we believe that the only proper legal basis is for public authorities to go for a regulatory intervention. We are aware of the costs usually associated with such interventions. However, in this case, we believe that the limited nature of our proposal is susceptible to keep those costs under control. Indeed, what would be needed is a body achieving standardization of the way search engine query logs are reported (and anonymized) and monitoring compliance. The question how to price access to data is thornier. Successful search engines could indeed argue that the proposal amounts to regulatory takings of proprietary datasets. The issue of access pricing is notoriously involved (see Armstrong (2002) or Vogelsang (2003) for introductory surveys) but it is parts and parcel of regulatory practice. We note, in addition, that the alternative to intervention is not the perpetuation of the current situation but, bar any breakthrough innovation, straight monopolization of the industry. Therefore, the economic cost of nonintervention is given by the depressed incentives to innovate for the dominant search engine.

A second issue of interest concerns the respect of the right to privacy of search engines' users. Let us mention upfront that, as alluded to in Footnote 9, search engines routinely obtain (with consent) more private data about users than we advocate to share. If a search engine offers additional services, such as an e-mail program, other productivity tools or a web browser, and a user stays logged in to those services while using the search engine, it is possible to match data on clicking-behavior with much more detailed information at the individual user level. Under our policy proposal, these data would not have to be shared, which implies that search engines could further exploit some competitive advantage generated by their use.

Whether query logs *per se* constituted 'personal data' in the sense of the European Union's *Data Protection Directive* (*95/46/EC*) was debated for a while. It is now clear that the ability to trace back the physical address of users on the basis of their IP address (or, worse, to clearly identify them as persons on the basis of a unique identifier derived from a permanent cookie) implies that search engines' logs contain 'personal data'.[43] Moreover, the obligations derived from the Directive apply to all search engines doing business in the European Economic Area, independently of the location of their headquarters (or servers). Hence, all search engines potentially fall within the scope of the Directive. Retention periods put objective limits to the amount of data that can be exchanged under our proposal but irreversible anonymization of logs is generally seen as one way search engines can comply with the Directive by eliminating the personal dimension of data.[44] Indeed, to pre-empt any privacy concerns, our policy proposal only refers to anonymized data on search sessions.

5 Discussion and Conclusion

Given our model, it is an open question how Google managed to become market leader and to outcompete both Yahoo and Microsoft by 2003. Likely explanations are that Google's search algorithm was in fact of drastically higher quality than its competitors' at this stage or that query data mining was not yet in its mature phase. Apparently, Yahoo and other search engines active before Google's market expansion were not able to exploit the network externalities present in the industry decisively. Since Google has taken over the pole position, however, we do not know whether it has increased its lead in market shares because its algorithm quality has become even better than its competitors' in relative terms or whether its success has mainly happened due to network externalities, in the absence of drastic innovation. Our model suggests that the latter effect is sufficient to explain the market development throughout the last years.

We believe that, in order to reap the full benefits of competition in this highly innovative market, it is necessary to intervene and reinstall merit-based competition. Our results have shown that the rate of innovation, quality, consumer surplus, and total welfare are higher if search engines only compete with their actual algorithm qualities instead of the current situation, where a dominant firm may still innovate but has less incentives to do so at the margin, given the tipping market effect.

This proposal is in line with the US Department of Justice's (2010) decision, which stated that the Yahoo-Microsoft alliance in February 2010 "will enhance Microsoft's competitive performance because it will have access to a larger set of queries, which should accelerate the automated learning of Microsoft's search and paid search algorithms and enhance Microsoft's ability to serve more relevant search results and paid search listings, particularly with respect to rare or 'tail' queries. The increased queries received by the combined operation will further provide Microsoft with a much larger pool of data than it currently has or is likely to obtain

[43] See Opinion 1/2008 on data protection issues related to search engines of the advisory Working Party created by Article 29 of the Directive; `http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2008/wp148_en.pdf`.

[44] See the source mentioned in footnote 43 for this assessment. See `http://searchengineland.com/google-anonymizing-search-records-to-protect-privacy-10736` for details on how anonymization can work in practice.

without this transaction. This larger data pool may enable more effective testing and thus more rapid innovation of potential new search-related products, changes in the presentation of search results and paid search listings, other changes in the user interface, and changes in the search or paid search algorithms. This enhanced performance, if realized, should exert correspondingly greater competitive pressure in the marketplace."

We argue that the authorities should not stop at clearing the Microsoft-Yahoo deal but should push for a level playing field in the search engine market more actively. Such an intervention would present policymakers with the same dilemma as other interventions in high-tech sectors: one can well argue that technical innovation will come soon in the future in such a drastic manner that the current firms and technologies will become obsolete. In that case, it is preferable to simply let the current market leader reap the profits from its past research effort and hope that incentives for breakthrough innovation remain sufficient meanwhile. That is, however, a bet and one may well want to tilt the market towards higher levels of incremental innovation now by restoring firms' ability to compete neck-and-neck. Whatever the final decision, an implicit stance is taken on the kind and level of innovative activity that is desirable in the search engine market.[45]

[45] For a reading of the European *Microsoft* case along similar lines, see Larouche (2009).

Appendix

A Computations

A.1 The Triopoly Case

The model is a static game with complete information. Solving (1) for all three firms gives the following system of reaction functions:

$$x_1 = \sqrt{pN_1(x_2 + x_3)} - x_2 - x_3 \tag{A.1}$$

$$x_2 = \sqrt{pN_2(x_1 + x_3)} - x_1 - x_3 \tag{A.2}$$

$$x_3 = \sqrt{p(x_1 + x_2)} - x_1 - x_2 \tag{A.3}$$

Solving this system for optimal quality levels gives the following unique Nash equilibrium with positive market shares:

$$x_1^* = \sqrt{\frac{p^2 N_1^2 N_2^2}{(N_2 + N_1(1 + N_2))^2} - \frac{4pN_1N_2^2}{(N_2 + N_1(1 + N_2))^2}} \tag{A.4}$$

$$x_2^* = \frac{2pN_1N_2(N_1(N_2 - 1) + N_2)}{(N_2 + N_1(1 + N_2))^2} \tag{A.5}$$

$$x_3^* = \frac{2pN_1N_2(-N_1(N_2 - 1) + N_2)}{(N_2 + N_1(1 + N_2))^2} \tag{A.6}$$

Nash equilibrium quality levels lead to the following market shares:

$$D_1 = 1 - \frac{2N_2}{N_2 + N_1(1 + N_2)} \tag{A.7}$$

$$D_2 = 1 - \frac{2N_1}{N_2 + N_1(1 + N_2)} \tag{A.8}$$

$$D_3 = \frac{N_1(1 - N_2) + N_2}{N_2 + N_1(1 + N_2)} \tag{A.9}$$

Figure 1 displays equilibrium quality levels and market shares as a function of N_1. We proceed by analyzing the welfare effects of network externalities in the triopoly model. Equilibrium profits are easily calculated by substituting equilibrium quality levels into (1). We obtain:

$$\pi_1 = \frac{(N_2 - N_1(1 + N_2))^2}{(N_2 + N_1(1 + N_2))^2} p - F \tag{A.10}$$

$$\pi_2 = \frac{(N_2 + N_1(N_2 - 1))^2}{(N_2 + N_1(1 + N_2))^2} p - F \tag{A.11}$$

$$\pi_3 = \frac{(N_2 + N_1(1 - N_2))^2}{(N_2 + N_1(1 + N_2))^2} p - F \tag{A.12}$$

Producer surplus is the sum of all firms' profits:

$$PS = \pi_1 + \pi_2 + \pi_3 = \frac{(3N_2^2 - 2N_1N_2(1 + N_2) + N_1^2(3 + N_2(3N_2 - 2)))}{(N_2 + N_1(1 + N_2))^2} p - 3F \tag{A.13}$$

We proxy consumer surplus by the equilibrium quality levels weighted with the market shares of the active firms:

$$CS = D_1 x_1^* + D_2 x_2^* + D_3 x_3^* = \frac{2N_1N_2(3N_2^2 - 2N_1N_2(1+N_2) + N_1^2(3+N_2(3N_2-2))}{(N_2+N_1(1+N_2))^3}p \quad (A.14)$$

Summing up producer surplus and consumer surplus, we get total welfare:

$$W = \frac{(N_2+N_1(3N_2+1))(3N_2^2 - 2N_1N_2(1+N_2) + N_1^2(3+N_2(3N_2-2))}{(N_2+N_1(1+N_2))^3}p - 3F \quad (A.15)$$

Figure 2 displays a graphical representation of firms' profits and total welfare as a function of N_1.

A.2 The Duopoly Case

It is straightforward to solve π_1 and π_2 (see (2)) for a unique Nash equilibrium in quality levels with positive market shares:

$$x_1^* = \frac{N_1^2}{(1+N_1)^2}, \quad x_2^* = \frac{N_1}{(1+N_1)^2}, \quad (A.16)$$

which leads to the following market shares:

$$D_1 = \frac{N_1}{1+N_1}, \quad D_2^* = \frac{1}{1+N_1} \quad (A.17)$$

(A.17) shows that the duopoly market is developing in the same way as the triopoly market: $lim_{N_1 \to \infty} D_1 = 1; lim_{N_1 \to \infty} D_2 = 0$. In words, the market is tipping.

Firms' profits, producer surplus, consumer surplus, and total welfare can be computed as in the triopoly model, which yields:

$$\pi_1 = \frac{N_1^2}{(1+N_1)^2} - F; \pi_2 = \frac{1}{(1+N_1)^2} - F; PS = \frac{N_1^2+1}{(1+N_1)^2} - 2F \quad (A.18)$$

$$CS = \frac{N_1+N_1^3}{(1+N_1)^3}; W = \frac{1+2N_1+N_1^2+2N_1^3}{(1+N_1)^3} - 2F \quad (A.19)$$

Given that the graphs of the duopoly case qualitatively resemble the graphs of the triopoly case without firm 3, we do not display them here.

A.3 The Monopoly Case

In the case of a pure (uncontested) monopoly, monopolistic firm 1 does not face any competition. Hence, it rationally sets quality $x_1^* = \epsilon$, where ϵ is the smallest positive number that induces consumers to use its services. Nevertheless, it serves the entire market: $D_1 = 1$, which leads to profits of $\pi_1 = p - \frac{\epsilon}{N_1} \approx p$.

The welfare analysis of this case is equally straightforward. Consumer get a surplus of $CS = \epsilon \approx 0$, whereas $PS = \pi_1$. Hence, total welfare sums to $W = p + \frac{N_1 - 1}{N_1}\epsilon \approx p$.

In the case of a contested monopoly, we solve the three-stage sequential game by backward induction, where we set $N_2 = 1$. At stage 3, given that firm 2 entered the market, it sets its quality, x_2, according to its duopoly reaction function:

$$x_2 = \sqrt{px_1} - x_1 \tag{A.20}$$

Note that this only holds for $p \geq x_1$. If $x_1 > p$, firm 2's best response is to set $x_2 = 0$.

At stage 2, firm 2 enters the market if, and only if its expected profit from entry is positive. If we substitute (A.20) into firm 2's duopoly objective function, (2), assuming $N_2 = 1$, this requires:

$$\pi_2(x_1, F, K, p) = p - 2\sqrt{px_1} + x_1 - F - K > 0 \tag{A.21}$$

Solving (A.21) with the equality sign for zero shows that $\pi_2(x_1, F, K, p) \leq 0 \quad \forall \quad x_1 \in \mathcal{I}$, where the interval $\mathcal{I}$ is defined as $\mathcal{I} : [K + F - 2\sqrt{(K+F)p} + p, K + F + 2\sqrt{(K+F)p} + p]$. The upper bound of this interval, however, is larger than p, which hurts the condition below (A.20). It follows that the only condition that firm 1 has to meet such that even firm 2's best response quality would leave it with a loss is $x_1 \geq K + F - 2\sqrt{(K+F)p} + p$. Given that firm 1 has no intrinsic value in providing a higher quality than necessary, it follows that, if it wants to keep firm 2 out of the market, firm 1 sets x_1 equal to the lower bound of $\mathcal{I}$. In other words, firm 1's "limit quality" is:

$$x_1^{lim} = K + F - 2\sqrt{(K+F)p} + p \tag{A.22}$$

At stage 1, when determining its quality level, firm 1 effectively also chooses its competition because its quality determines whether firm 2 enters the market, or not. If it produces the limit quality, it keeps firm 2 out of the market.

If instead firm 1 sets x_1 such that firm 2 could possibly enter the market profitably, firm 2 will set its quality according to (A.20). Substituting this in firm 1's unconstrained duopoly profit function, (2), shows that firm 1 solves:

$$max_{x_1} \quad \pi_1 = \sqrt{px_1} - \frac{x_1}{N_1} - F, \tag{A.23}$$

which is solved by firm 1's Stackelberg quality:[46]

$$x_1^{Stackelberg} = \frac{N_1^2 p}{4} \tag{A.24}$$

If firm 2 enters, it will react to $x_1^{Stackelberg}$ by setting x_2 according to (A.20), which gives:

$$x_2^{Stackelberg} = \frac{p}{4}(2N_1 - N_1^2) \tag{A.25}$$

Resulting market shares are:

$$D_1^{Stackelberg} = \frac{N_1}{2}; \quad D_2^{Stackelberg} = 1 - \frac{N_1}{2}, \tag{A.26}$$

[46] We are aware that classical Stackelberg competition uses quantity levels, not quality levels, as strategic variable. Still, we feel that our model is sufficiently close to the original Stackelberg model to warrant the use of its name.

which implies that firm 2 gives up and sets $x_2 = 0$ as soon as $N_1 \geq 2$. We obtain the following welfare results:

$$\begin{aligned}\pi_1^{Stackelberg} &= \frac{N_1 p}{4} - F; \pi_2^{Stackelberg} = \frac{p}{4}(N_1 - 2)^2 - F - K; \\ PS = \frac{p}{4}(4 + N_1(N_1 - 3)) - 2F - K; CS &= \frac{pN_1}{4}(2 + N_1(N_1 - 2)); \\ W &= \frac{p}{4}(4 + N_1(N_1(N_1 - 1) - 1)) - 2F - K \end{aligned} \tag{A.27}$$

Firm 1 excludes firm 2 by setting x_1^{lim} if, and only if, $\pi_1^{lim} \geq \pi_1^{Stackelberg}$. This holds with the equality sign for $K = (1 + N_1 - \frac{N_1^2}{4})p \pm \sqrt{(4 - N_1)N_1 p^2} - F$. Moreover, we have seen above that firm 2 can only get a positive market share if $N_1 \in [1, 2]$. The combined effect of both constraints is depicted in Figure 3.

A.4 The Competitive Oligopoly Case

Each of the three firm maximizes its profits as given in (1) but with equal installed bases $N = N_1 + N_2 + N_3$. The resulting Nash equilibrium in quality levels is:

$$x_1^* = x_2^* = x_3^* = \frac{2}{9} N p \equiv x^{comp} \tag{A.28}$$

which leads to market shares of 1/3 for each firm. The associated welfare results are:

$$\pi_1 = \pi_2 = \pi_3 = \frac{p}{9} - F; \quad PS^{comp} = \frac{p}{3} - 3F \tag{A.29}$$

$$CS^{comp} = \frac{2}{9} N p; \quad W^{comp} = \frac{p}{9}(3 + 2N) - 3F \tag{A.30}$$

Quality, consumer surplus and total welfare are depicted in Figures 5 and 6.

References

Adar, Eytan (2007). "User 4XXXXX9: anonymizing query logs", *WWW '07 conference paper*, available at http://www.cond.org/anonlogs.pdf

Adler, Paul S., and Kim B. Clark (1991). "Behind the learning curve: a sketch of the learning process", *Management Science*, 37(3), 267-281.

Anderson, Simon, André de Palma, and Jacques-François Thisse (1992). *Discrete choice theory of product differentiation.* Cambridge: MIT Press.

Armstrong, Mark (2002). "The theory of access pricing and interconnection", in Martin E. Cave, Sumit K. Majumdar, and Ingo Vogelsang (eds.), *Handbook of Telecommunications Economics, Volume 1.* Amsterdam: Elsevier North-Holland, 295-384.

Arrow, Kenneth J. (1962). "The Economic Implications of Learning by Doing", *The Review of Economic Studies*, 29(3), 155-173.

Athey, Susan and Glenn Ellison (forthcoming). "Position auctions with consumer search", *The Quarterly Journal of Economics.*

Baumol, William J., Panzar, John C., and Willig, Robert D (1982). *Contestable Markets and the Theory of Industry Structure.* New York: Harcourt Brace Jovanovich, Inc. 1982.

Bar-Yossef, Ziv, and Maxim Gurevich (2008). "Mining search engine query logs via suggestion sampling", *VLDB '08 conference paper.*

Benkard, C. Lanier (2000). "Learning and forgetting: the dynamics of aircraft production", *American Economic Review* 90(4), 1034-1054.

Besen, Stanley M., and Joseph Farrell (1994). "Choosing how to compete: strategies and tactics in standardization", *Journal of Economic Perspectives* 8(2), 673-699.

Boston Consulting Group (1972). *Perspectives on experience.* Boston: Boston Consulting Group.

Chen, Yongmin, and Chuan He (2006). "Paid Placement: Advertising and Search on the Internet", *NET Institute Working Paper* No. 06-02.

Economides, Nicholas (1996). "The economics of networks", *International Journal of Industrial Organization* 14(6), 117-131.

Economides, Nicholas (2010). "Antitrust issues in network industries", in Ioannis Lianos and Ioannis Kokkoris (eds.), *The reform of EC competition law: new challenges.* Alphen aan den Rijn: Kluwer International Law, 343-376.

Edelmann, Benjamin, Michael Ostrovsky, and Michael Schwarz (2007). "Internet advertising and the generalized second-price auction: selling billions of dollars worth of keywords." *American Economic Review* 97(1), 242-259.

Elhauge, Einer, and Damien Geradin (2007). *Global competition law and economics.* Oxford: Hart Publishing.

Evans, David S. (2008). "The economics of the online advertising industry", *Review of Network Economics* 7(3), 359-391.

Gandal, Neil (2001). "The dynamics of competition in the Internet search engine market", *International Journal of Industrial Organization* 19, 1103-1117.

Gandal, Neil (2002). "Compatibility, standardization and network effects: some policy implications", *Oxford Review of Economic Policy* 18(1), 80-91.

Gans, Joshua S. (2011) "When is static analysis a sufficient proxy for dynamic considerations? Reconsidering antitrust and innovation." In Josh Lerner and Scott Stern (eds.), *Innovation policy and the economy 2010*, vol. 11, Chicago: University of Chicago Press.

Ghose, Anidhya, and Sha Yang (2009). "An empirical analysis of search engine advertising: sponsored search in electronic markets" *Management Science* 55(10), 1605-1622.

Goldfarb, Avi, and Catherine Tucker (2011a). "Search engine advertising: channel substitution when pricing ads to context." *Management Science* 57(3), 458-470.

Goldfarb, Avi, and Catherine Tucker (2011b). "Online display advertising: Targeting and obtrusiveness," *Marketing Science*, doi:10.1287/mksc.1100.0583.

Goldfarb, Avi, and Catherine Tucker (2011c). "Substitution between Online and Offline Advertising Markets," *Journal of Competition Law & Economics* 7(1), 37–44.

Grimes, Carrie, Diane Tang, and Daniel M. Russell (2007). "Query logs alone are not enough", *WWW '07 conference paper*, available at `http://dmrussell.googlepages.com/query_logs_are_not_enough_2007_02_12.pdf`

Jansen, Bernard J. (2006). "Search log analysis: what it is, what's been done, how to do it", *Library & Information Sciences Research*, 28(3), 407-432.

Jansen, Bernard J., Mimi Zhang, and Ying Zhang (2007). "Brand Awareness and the Evaluation of Search Results", *WWW 2007 conference paper.*

Larouche, Pierre (2009). "The European *Microsoft* case at the crossroads of competition policy and innovation: comment on Ahlborn and Evans", *Antitrust Law Journal* 75(3), 933-963.

Levy, Steven (2009). "Secret of googlenomics: data-fueled recipe brews profitability", *Wired Magazine* 17.06. Available at `http://www.wired.com/culture/culturereviews/magazine/17-06/nep_googlenomics`

Menon, Syam and Sumit Sarkar (2007). "Minimizing information loss and preserving privacy", *Management Science*, 53(1), 101-116.

Pollock, Rufus (2010). "Is Google the Next Microsoft? Competition, Welfare and Regulation in Online Search", *Review of Network Economics* 9(4), 1-29.

PwC (2010). "IAB Internet Advertising Revenue Report", `http://www.iab.net/media/file/IAB_report_1H_2010_Final.pdf`.

Ratliff, James D., and Daniel L. Rubinfeld (2010). "Online Advertising: Defining Relevant Markets", *Journal of Competition Law & Economics* 6(3), 653–686.

Riggins, Frederick J., Charles H. Kriebel, and Tridas Mukhopadhyay (1994). "The growth of interorganizational systems in the presence of network externalities", *Management Science*, 40(8), 984-998.

Rochet, Jean-Charles, and Jean Tirole (2006). "Two-Sided Markets: A Progress Report", *The RAND Journal of Economics* 37(3), 645-667.

Rochet, Jean-Charles, and Jean Tirole (2007). " Introduction to the Symposium on Two-Sided Markets", *Competition Policy International* 3(1), 146-149.

Shapiro, Carl, and Hal R. Varian (1999). *Information rules: A strategic guide to the network economy.* Cambridge, MA: HBS Press.

Silvestri, Fabrizio (2010). "Mining query logs: turning search usage data into knowledge", *Foundations and Trends in Information Retrieval*, 4(1-2), 1-174.

Spink, Amanda, Wolfram, Dietmar, Jansen, Major B. J., and Saracevic, Tefko (2001). "Searching the Web: The Public and Their Queries", *Journal of the American Society for Information Science and Technology*, 52(3), 226–234.

Spulber, Daniel F. (2009). "The map of commerce: Internet search, competition, and the circular flow of information", *Journal of Competition Law and Economics* 5(4), 633-682.

Taylor, Greg (2010). "Search quality and revenue cannibalization by competing search engines." Mimeo, Oxford University.

Telang, R., T. Mukhopadhyay and U. Rajan (2004). "Market structure for Internet search engines", *Journal of Management Information Systems* 21, 137-160.

US Department of Justice (2010), "Statement of the Department of Justice Antitrust Division on its Decision to Close its Investigation of the Internet Search and Paid Search Advertising Agreement Between Microsoft Corporation and Yahoo! Inc.", February 18.

Varian, Hal R. (2006). "The economics of Internet search", *Rivista di Politica Economica*, 177-191.

Varian, Hal R. (2007). "Position auctions", *International Journal of Industrial Organization* 25, 1163-1178.

Vogelsang, Ingo (2003). "Price Regulation of Access to Telecommunications Networks", *Journal of Economic Literature* 41, 830-62.

Wang, Eric T. G., and Abraham Seidmann (1995). "Electronic data interchange: competitive externalities and strategic implementation policies", *Management Science*, 41(3), 401-418.

Wiersma, Eelke (2007). "Conditions that shape the learning curve: factors that increase the ability and opportunity to learn", *Management Science*, 53(12), 1903-1915.

Wright, Theodore P. (1936). "Factors affecting the cost of airplanes", *Journal of the Aeronautical Sciences* 3(4), 122-128.

Xie, Jinhong, and Marvin Sirbu (1995). "Price competition and compatibility in the presence of positive demand externalities", *Management Science*, 41(5), 909-926.

Yao, Song, and Carl F. Mela (2011). "A dynamic model of sponsored search advertising", *Marketing Science*, 10.1287/mksc.1100.0626.

Zangwill, Willard I., and Paul B. Kantor (1998). "Towards a theory of continuous improvements and the learning curve", *Management Science*, 44(7), 910-920.