

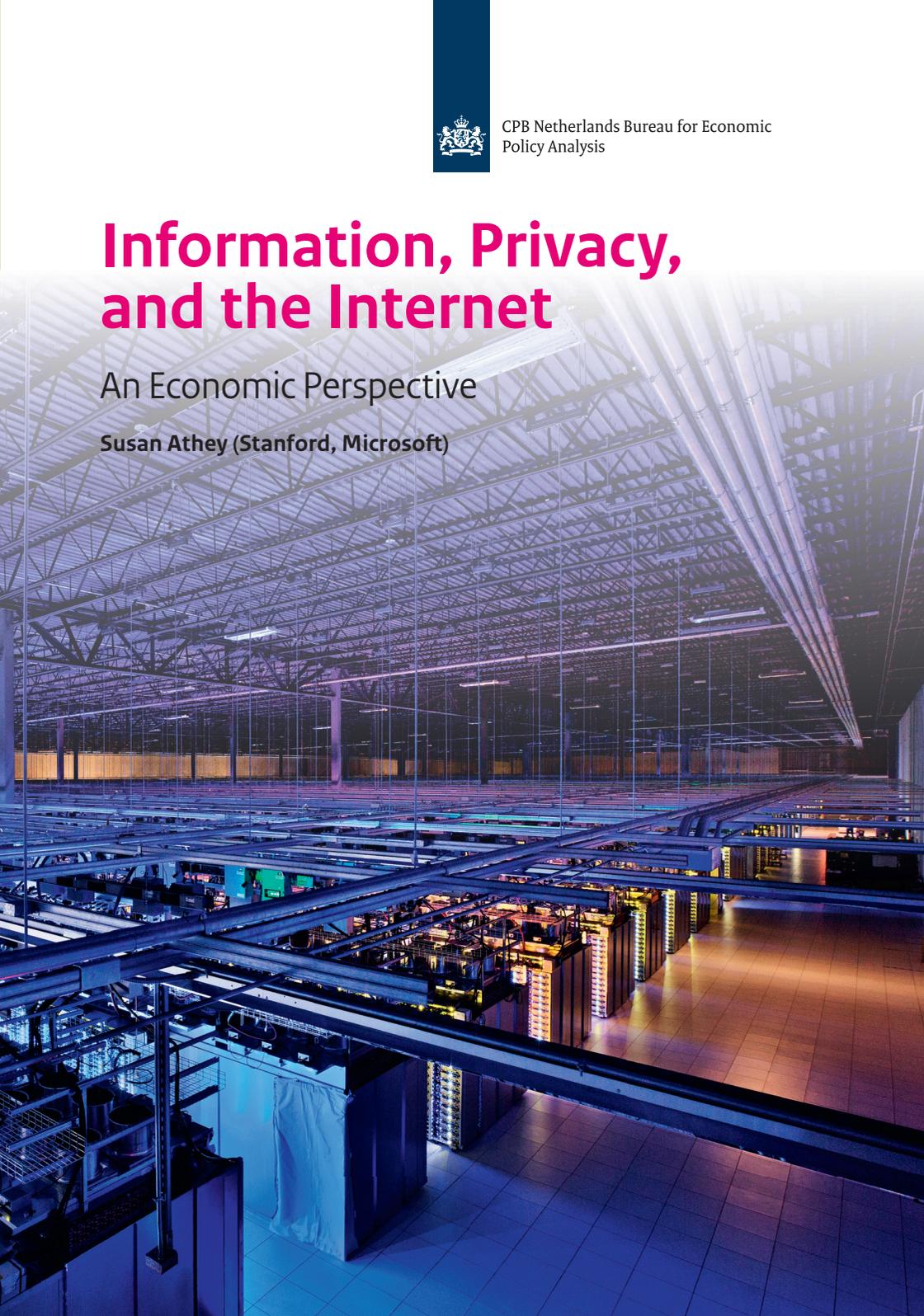


CPB Netherlands Bureau for Economic  
Policy Analysis

# Information, Privacy, and the Internet

An Economic Perspective

Susan Athey (Stanford, Microsoft)



June, 2014

**Information, Privacy, and the Internet:  
An Economic Perspective**

**Susan Athey**

*Stanford Graduate School of Business<sup>1</sup>*

## Contents

Introduction	5
1. Innovation, Value Creation and Data-Driven Advertising	11
2. Competition and Welfare in Search and Online Advertising	19
3. Information Gatekeepers, Data and Competition on the Internet	31
4. The Impact of User Multi-Homing on Publishers	43
5. An Economic View on Privacy	46
Conclusions	61



# Introduction

The arrival of the “era of big data” has been heralded as transformative for industry, economic growth, and efficiency (Mayer-Schönberger and Cukier, 2014). Every day brings new headlines about technological advances that have the potential to greatly improve our lives. Computers learn to understand our speech, e-commerce and mobile application shopping platforms present us with prescient recommendations, news about our friends and their babies is prioritized and ranked by social media sites, our travel reservations are organized into itineraries, and we listen to music curated by an algorithm. We can find inspirational photos that provide design ideas for remodeling our homes that previously would have been available only through professionals, and we can shop for homes online using maps, dozens of photos, and detailed information about homes, all neatly organized for us. We have come to expect these services for free or at very low cost, and we are not at all surprised to see new websites, fresh news coverage, innovative mobile applications, and improvements in speech recognition every day.

We also notice advertisements that are increasingly tailored to our interests, with slideshows of products appearing in display ad windows. After browsing for autos, we begin to see more ads for cars. And advertisements begin to appear in places that were previously ad-free. Services that we have become attached to begin to replace free and unbiased content with paid content, and to promote their own complementary services. Others are acquired

by the largest tech firms, or disappear in the face of an acquisition of a competitor by a large firm.

Although we always knew that our data was out there being used, it has become more salient as we see companies use it in new innovative ways. Yet, many of us are happy to hand over personal data in exchange for valuable services. For example, we might use an application that scans our email looking for travel plans and creates travel itineraries, saving valuable time. We might be delighted when the “personal assistant” function on our mobile phone reminds us of appointments. On the other hand, news of widespread government surveillance has made many nervous, as they had not expected that this surveillance was occurring. Surveillance has captured the imagination of the public, causing us to focus in a new way on the kinds of data companies have about us and how the data may be used in ways we did not originally intend. As Microsoft’s Craig Mundie commented, “People are now being observed in increasingly intimate ways by all the technology in their lives. Too much data is being collected in too many ways” (Burt, 2013).

In the face of all of these trends, it is difficult for policy-makers to know how to respond. Antitrust and privacy regulators both face a classic tradeoff between allowing technological innovation to proceed disciplined only by the market and consumer choice, or intervening and risking doing more harm than good. Despite the fact that policy-makers profess an acute awareness of these tradeoffs, much of the public discourse around these issues misses fundamental facts as well as economic principles that provide some

guidance as to when market solutions are likely to well and poorly, as well as when regulatory solutions might do well and poorly.

For example, conventional wisdom says that when markets are more competitive, social welfare generally increases, as competition induces firms to increase quality, lower prices, or both. However, this wisdom only applies when the firms' primary strategic alternatives are to increase quality or lower prices. Technology platforms, as complex entities, often have a wider array of strategic choices. Many of the most damaging antitrust accusations against technology firms have occurred in a context where the firms in question were dominant but perceived the new arrival competitive threats, indeed threats that could dislodge their core business. For example, the Microsoft antitrust case in the 1990s revolved around Microsoft's response to what it perceived as the threat of the internet browser—Microsoft was accused of promoting its own browser at the expense of others, using its dominant position in operating systems to gain leverage. In the 1980s, when American Airlines owned the Sabre reservation system for ranking flights, it was accused of responding to price competition and entry by low-cost carriers by ranking its own connecting flights above competitors' nonstop flights in the flight reservation rankings, hoping to drive out new entrants by depriving them of customers. The current EU investigation of Google has focused on its manipulation of the search results page to promote its own specialized search sites at the expense of specialized competitors who threatened to compete away users in the most profitable segments of internet search.<sup>1</sup> Despite these

<sup>1</sup> See Edelman (2014) for more examples of potential competition policy violations by Google.

prominent examples, it is tempting for regulators to simply argue that the technology sector is competitive and ignore the fact that not all competitive responses are welfare-enhancing.

In contrast, privacy policy has evolved more in response to governments and consumer organizations than to direct market pressures or consumer purchase and utilization behavior. It has proved challenging for both firms and researchers to detect significant changes in consumer utilization of their products in response to privacy policy changes, and further most consumers do not even read or understand privacy policies to begin with. Even experts can be surprised about new ways that data can be used, or about the implications of long-term retention of personal data. This suggests that relying solely on market forces to achieve outcomes that benefit consumers may be short-sighted. At the same time, the challenges of creating regulation and policy that does more good than harm remain. Particularly difficult challenges emerge when creating regulatory policy for an industry with super-dominant firms, because consumers may have to give up a fair bit in terms of their satisfaction with a product in order to choose a niche competitor with more attractive privacy policies.

In this paper, I will provide a selective review of some of the economic issues surrounding innovation on the internet, the role of data, and privacy. A deeper understanding of these issues is crucial to creating intelligent policy that achieves an appropriate balance between costs and benefits in a complex and highly innovative industry.

Some key insights highlighted in the paper include:

- Consumer data creates enormous value for both consumers and advertisers, as measured by user engagement
- Value must be accrued to content creators rather than ad platforms and intermediaries, otherwise innovation will be hampered, decreasing the quantity and quality of future of internet content such as investigative journalism, low-cost or free productivity software, and entertainment
- Competition in advertising platforms is crucial to ensure that the value created with consumer data accrues to internet publishers and advertisers (who are often themselves internet content creators), rather than advertising platforms and intermediaries.
- Data is extremely important for helping consumers find information, and search engines with more data can provide better rankings
- Search engines and news aggregators have enormous power to shift user viewing patterns and thus pick winners on the internet; for example, promoting a link from fifth to first position in algorithmic search rankings leads to a 340% increase in traffic to that site, on average
- Dominant search engines have the incentive and ability to manipulate results to favor their own vertical sites or advertising revenue
- Dominant ad platforms can profit if they have unique access to personal data, or a unique ability to track users across websites. This affects the incentives of dominant ad platforms in their design of privacy policy

- Privacy regulation can play an important role because market forces will not lead to efficient outcomes
- Efficient privacy policy must consider the welfare gains created by efficient advertising as well as strong incentives for entry and innovation
- Overly strong privacy regulation can be regressive, making certain free services unprofitable and disproportionately harming poor people. It can also favor certain types of news businesses (large websites with a mix of commercial and straight news content) over others (specialized websites focusing mainly on political analysis, for example)

# 1 Innovation, Value Creation and Data-Driven Advertising

One of the most important ways in which data has been used in the last 15 years is in online advertising. Advertising is the most important revenue stream for a wide range of internet businesses, from search engines (pulling in more than \$50 billion globally), to social media platforms, to the online news media. In addition, even for many apps and websites that start out as ad-free, advertising is a crucial component of their planned steady state business model, and the businesses include projections of future advertising revenue as a key component of business plans provided to investors. Although historically, advertising provided the business incentive for the creation of news content and radio and television entertainment, the range of services funded by advertising today is far beyond news and entertainment, including office productivity software, travel management, reviews for products and services, and thousands of niche applications that provide productivity and information.

Advertising also plays another crucial role for new ventures. It is one of the only ways for businesses to be discovered. For example, a new application for a mobile device may have a difficult time rising to the top of the rankings without obtaining some initial users (and their positive rankings, as well as feedback) first. As internet search results pages for commercial queries leave less and less space for “algorithmically determined” links, new websites may rely heavily on advertising to be discovered. Only after some initial users are gained can the websites hope to rise to the top of the rankings to get new users organically. Targeted advertising can

allow application developers to find an audience of users that may be interested in a niche application; without targeting, advertising may be prohibitively expensive for the new business, as much of the advertising expenditure may be wasted.

Thus, even though a naïve view of advertising would consider it a negative for consumers, it is hard to imagine life today without all of the consumer benefits that are directly enabled by advertising. This has never been more salient than in the internet era. For this reason, the efficiency and effectiveness of advertising is of first-order importance to determining consumer welfare from digital products and information.

In addition, advertising platforms and exchanges play a crucial role. These institutions match publisher inventory to advertisers, and keep a share of the advertising revenue generated as profit for the ad platform. It is crucial for publishers, app developers, and content creators that ad platforms and intermediaries do not extract too high a share of the surplus created by advertising. That is, ad prices should not be artificially high, or new websites will not be able to afford the acquisition of new users; and the share of ad revenue given to publishers should not be too low, otherwise the advertising revenue will not be sufficient to incentivize publishers to enter and create content. Of course, one way to discipline advertising platforms is to have competition in ad platforms.

Overall, then, the vibrant web and publishing ecosystem consumers enjoy today is crucially dependent on efficient and effective advertising, as well as competitive behavior by ad platforms.

## **How is Data Used to Create Value in Online Advertising?**

Online advertising has created a variety of new and innovative forms of advertising, some of which is highly specialized to the context of the user. By far the most successful to date has been search advertising. In its pure form, a user enters a query and the search engine returns a set of “algorithmically ranked” links as well as a set of links to sponsored content, sold using a real-time auction. One reason that search advertising is so successful is that the advertisements are very similar to the non-paid content that the user is receiving. The user entered the search query in order to receive links to websites that had relevant information or sold relevant products. The advertisements also fill the same need, but are selected according to different criteria, where relevance is still a crucial component of how ads are chosen. Paid classified ads are another historical example where the advertisements are the content: classifieds make users aware of products or services that they might want to buy, and users go to the classified section to find that information.

Thus, search advertising is not purely wasteful nor entirely disliked by consumers; instead, it creates value. A profit-maximizing search engine will typically display more ads than would be ideal from a pure user experience perspective, but will rather trade off user experience and advertising revenue in choosing how many ads to display. So while the value creation is on average lower from ads than algorithmic results given the choices made by search engines, there is still a lot of value creation from the ads (as measured through, for example, click-through rates).

At the other extreme, classic “display” advertising was less tailored to the user’s context or intent. As a result, prices for display advertising were a small fraction of those in search. This advertising created less value to consumers and advertisers, and less revenue to publishers.

There is a continuum of outcomes as to whether advertising is a value-enhancing part of the user experience, essentially a complementary form of content (search ads, classifieds, sponsored links to news, sponsored posts in Twitter) or whether advertising is purely annoying, distracting, misleading, or even dangerous (e.g. distributing malware; see Edelman (2011) for a number of real-world examples). In some cases, such as product listing services or shopping sites, all content is a form of advertising. Though some business models of advertisers and publishers are based on trickery and misleading consumers (see Edelman (2011), many websites have the incentive to provide ads relevant to the consumer’s intent, as these monetize better and create a better user experience.

The general industry trend has been towards making advertising more relevant to the user’s context as well to the user’s interests (independent of context). Industry reports suggest that such changes often increase the value created for both users and advertisers by an order of magnitude, and an ever-increasing proportion of advertising on the web utilizes some sort of user data or segmentation. There are several important categories for how user and contextual data can be used:

- Getting users information that they will value: both tailoring the ad and targeting the user
- Tailoring the ad to the user's context, where users prefer different types of ads when they are doing research, consuming entertainment, or browsing social media
- Publishers tailoring their own content as well as advertising to the user. For example, news websites may rank articles differently based on user history. They may also use browsing information to target ads out of context: for example, a general news page might serve different ads after you visit the finance or technology page

Observe the differences in outcomes that you might expect if “cookies” can be used only within a website (first-party cookies), or if they can be used across websites (third-party cookies). If the same publisher owns a wide range of content, that publisher can use information from a consumer's web browsing on commercial parts of the site (e.g. the autos or finance or technology pages) to place more efficient advertising on less commercial parts of the site (hard news). On the other hand, if a website focuses only on hard news and does not have access to other information about the consumer's web browsing, it will be at a distinct disadvantage with regard to monetization as well as user experience.

An article in *Business Insider* about the shift by advertisers to behavioral targeting campaigns (Maher, 2010) echoes this point:

If agencies continue to spend more of their budgets on behavioral targeting campaigns smaller niche publishers will have a hard time competing for these dollars with the

portals and networks with much larger audiences. This will cause smaller publishers to join networks that offer behavioral targeting and split revenue with them, cutting into their margins.

Another very important use of data in online advertising is in measuring “attribution.” An advertiser needs to understand the return on the advertising investment in order to make choices about the most cost-effective way to advertise. This can be a challenging problem if a user might see many different ads from the same advertiser on different websites and in different contexts. The more difficult it is for an advertiser to track user views of their ads across websites, the more difficult it is for the advertiser to allocate advertising dollars efficiently. A typical example cited in an industry report is given here:

One Adobe client, a hospitality and entertainment group, realized that their apps were driving sales through other online and offline channels. They only realized this once they stopped obsessing on the last click before a sale, and tracked customers across channels.

Potential inefficiencies deriving from the inability of advertisers to track users’ views of their ads are likely to continue to grow in importance. A Google Research report stated that 90% of people move between devices to accomplish a goal. Firms and services that find a way to keep track of users across devices will be at an advantage in terms of helping advertisers place ads efficiently.

## **Measuring the Benefits of Data**

How can we quantify the value of data for the efficiency of advertising? There are few universally applicable measures, since the value of additional data depends on how much data was used initially. However, a number of studies shed some light on this issue. Beales (2010) reported that the price of behaviorally targeted advertising was almost 3 times the price of untargeted advertising, reflecting the value attributed to reaching the right consumers.

Goldfarb and Tucker (2011) showed that the implementation of the 2002 E-privacy directive in the European Union, which restricted the use of targeting techniques, reduced the effectiveness of ads in the European Union by 64% relative to the rest of the world. The impact was larger for general websites (such as Yahoo.com) relative to more targeted websites (such as cars.com), illustrating in practice the fact that access to data (and thus privacy regulation) can have implications for the nature of businesses and content providers that can be successful. In this case, broad, general interest websites are disadvantaged.

Another way to look at the importance of data in advertising is to examine industry trends in different types of advertising. “Real-time bidding” (RTB) refers to advertisers bidding for individual user advertising impressions on an advertising exchange or platform. The use of cookies and behavioral targeting is prevalent in this setting. Non-RTB “Programmatic advertising” refers to automated, data-driven approaches to purchasing advertising and optimizing advertising spend across different channels and different audience segments within channels. Targeting and data

are also important in this category. Nonprogrammatic includes things like buying a banner ad on the front page of a newspaper, for all viewers, or for all viewers in a geographic region, but where advertising buys are chosen “by hand.” eMarketer (2013) reports that the more data-intensive RTB segment is forecast to grow from 11% of U.S. display advertising in 2011 to 52% by 2017, while Non-RTB programmatic advertising will grow from 13% in 2011 to 31% in 2017, leaving just 17% of display advertising in the traditional non-programmatic bucket. These projections when combined with Goldfarb and Tucker’s (2011) results suggest that the overall cost of privacy regulation in Europe could grow dramatically.

Data is also very important for customizing products and personalizing product offerings. Indeed, the pioneering work on data-driven marketing by credit card firm Capital One in the United States was based around using data and experimentation to get the right credit card offer to the right consumer (Clemons and Thatcher, 1998). This approach helped Capital One grow from a new entrant to a major player in credit cards. Today, from voice recognition to personalized recommendations through sites like Amazon.com, the more data available, the more accurate predictions can be. Overall, it appears that the value created by “big data” to improve recommendations has just scratched the surface, as the quality of the algorithms matures, but also because recommendation engines are bringing in more and more data from a variety of sources, including Twitter and other social media (Booker, 2013).

## 2 Competition and Welfare in Search and Online Advertising

Internet search and online advertising are two examples of “multi-sided markets.” This means that analyzing competition in these markets is quite complex. Since multi-sided markets are so central to competition in internet businesses, it is worthwhile to understand some of the principles of multi-sided markets in greater detail.

A multi-sided market is a business that brings together distinct groups of customers to interact. Here are some common examples:

Example	Platform	Side 1	Side 2
Media	TV, newspaper, internet publisher, Yellow pages, search engine	Advertisers	Readers, Users
Credit Cards	Visa, MC, AmEx	Retailers	Consumers
Video Games, OS	Xbox, Windows	Software Developers	Game Players
Employment	Careerbuilder.com	Employers	Job-Seekers
Dating	Match.com	Men	Women
Online Auctions	eBay	Sellers	Buyers
Real Estate	MLS/Realtors	Sellers	Buyers

All ad-funded businesses are examples of multi-sided markets.

Perhaps surprisingly, the economic theory of multi-sided markets is very recent, as almost all of it was developed after the year 2000. Much of the theory was originally motivated by legal cases, particularly credit cards.

Multi-sided markets differ from multi-product firms because the sides of the markets are typically distinct entities: consumers and advertisers, for example. Each side cares about the behavior of the other side of the market (whether they use the platform and how much, for example), but does not directly care how much the other side is charged. This leads to different incentives for the platform. Economists found that the more realistic and rich theory of platform markets was a much better lens for understanding the behavior of firms in platform markets than trying to simply apply existing theories of multi-product pricing. The new theory is useful even though the line between a traditional market and a multi-sided market is sometimes blurry. For example, Amazon may be considered a traditional firm in its book business; or it can be considered a multi-sided firm, matching publishers to consumers. In the end, whether it is necessary to use the theory of multi-sided markets depends on the question and the context.

Another key feature of many multi-sided markets (and one that helps determine whether this is a useful lens) is the presence of indirect network effects: one side of the market cares about the activity of the other side of the market. For example, advertisers care about how many users are on a platform when they decide

whether to advertise there. These differ from “standard” network effects where users want to directly interact with others on the platform (e.g. telephones and fax machines exhibit “standard” network effects).

Rysman (2009) provides an accessible introduction to multi-sided markets. Evans and Schmalensee (2013) provide a thorough overview of the economics literature on two-sided markets, with an emphasis on antitrust. The latter article provides more detail about theories of competition as well as strategic behavior. One initial result highlighted in both studies is that pricing is often very asymmetric in multi-sided markets. Prices on one side depend on how “elastic” demand is (how price-sensitive that side of the market) as well as the externality that side has on the other side. If one side is relatively price-sensitive, and it is very important to the other side, then prices are likely to be quite low, perhaps even below cost.

Another important point highlighted in the literature is that behavior on one side of the market impacts welfare and competition on all sides of the market. A firm might engage in exclusive behavior on one side of the market, gaining market power there; but the payoff could come from extracting surplus on the other side of the market. For example, a search engine might write a long-term exclusive contract with a publisher to send all of its search traffic to that search engine. This contract might give the search engine access to a large volume of users. This in turn might give the search engine market power in the search advertising market, allowing the search engine to raise prices on advertisers.

The exclusive behavior occurs in the publisher side of the market, but the harm occurs on the advertiser side of the market.

One very interesting set of results about competition in multi-sided markets concerns a particular set of stylized assumptions about the “homing” behavior of the different sides of the market, which refers to whether the individuals use multiple platforms or a single platform. For example, in mobile phones, many consumers purchase only one phone (single homing) while many application developers port their apps to multiple platforms (multi-homing). The literature usually starts from extreme assumptions, for example that all users single home (by assumption) in media markets, and that all advertisers multi-home (by assumption). Of course, the real world is more complex.

The theoretical result about what happens when two platforms compete, when one side (“users”) single-homes and the other fully multi-homes is quite stark. Working backwards, once a platform has attracted a set of users, the fact that the other side uses all platforms (by assumption) means that the platform can charge a monopoly price to the multi-homing side. Anticipating that, the platform is willing to pay up to the per-user monopoly profit to attract the single-homing side. If the good has zero marginal cost (like broadcast media), then the good will be given away for free to the single-homing side, and the firm will make investments to attract the single-homing side. This kind of stark result predicts that competition for single-homers will be intense, where the motivation is to extract revenue from the multi-homing side of the market. The Android mobile phone operating system is free to users, but searches conducted through the Google search engine

raise money from advertisers. (For further reading on multi-sided markets in various industries, see Eisenmann (2008); Eisenmann, Parker, and Van Alstyne (2006); Haigu and Yoffie (2009); or Lee (2011).) Of course, few real-world markets fit this framework exactly, but it is a useful starting point for understanding the kinds of pricing patterns we see in practice.

### **Competition in Online Advertising Markets: A Multi-sided Markets Perspective**

Online advertising markets are complex. There are not many comprehensive articles that cover all of the relevant economic and technical background. For some initial reading, see Evans (2009).

Search advertising is perhaps easier to begin with. Levy (2011) provides a detailed history of the development of Google and Google's online advertising. A description of search advertising auctions is given by Varian (2006), though the market has continued to evolve and become more complex over time. The U.S. Department of Justice (2008) press release on the Google-Yahoo! proposed agreement outlines some facts and assessments of the impact of competition in this market. The U.S. Department of Justice (2010) press release on the Microsoft-Yahoo! search alliance provides further background on the importance of economies of scale for competition.

In search advertising, advertisers place standing bids on keywords (for example, "tennis shoes") that are stored in a database. Bids may be exact match, so that ads are displayed only if the user enters exactly the same term; or "broad match," where the ads may be displayed on terms (for example, "tennis footwear") that search

engine algorithms determine are “related” (a nebulous concept whose definition may change over time). There are several possible “match types” in practice. Advertisers bid and pay “per click,” so that they pay only when a user clicks on the ads. The ads are “text” ads which appear in a similar format to the “algorithmic” or “natural” results that a search engine returns when a user enters a search. Thus, search advertising is very targeted to the user’s intent at the time: the user is expecting to get links they can click on that relate to the search term, and both the “natural” and “sponsored” results meet that criteria. Although in principle the prices for the advertisements are set in what is known as a “generalized second price auction” (see Varian (2006)),<sup>2</sup> in practice prices are determined by the “pure” auction in combination with various “reserve prices” (minimum prices that must be paid in order to be shown). Reserve prices are set at a very granular level, so that in principle the search engine can control prices at the level of the advertiser X search term.

From the perspective of competition, search advertising is a multi-sided market. A key feature is that there are economies of scale and indirect network effects across all sides of the market. Like most markets, competition generally enhances welfare of customers on all sides (though perhaps at the cost of duplicated fixed costs).

<sup>2</sup> For other academic articles describing online advertising auctions see, among others, Jain et al. (2005) and Mahdian, Nazerzadeh, and Saberi (2006) who propose an early system of allocating advertising space, Meek, Chickering, and Wilson (2005) who study incentive-compatible Vickrey auctions, and Aggarwal, Goel, and Motwani (2006), Edelman and Ostrovsky (2007) and Varian (2007) who analyze the generalized second price auction model and its relationship to Vickrey auctions.

The user side of the market is the one most people are most familiar with—most internet users also use search engines frequently to find things on the internet. The user side has economies of scale: the more users, the more data the search engine has to determine what links are most relevant (they record previous clicking behavior on the same or similar searches) and the more users can be harnessed for ongoing experimentation with new and improved algorithms. Search engines run thousands of experiments per year to improve the algorithms used for ranking. Competition among search engines is important because it provides the incentive for ongoing research and development. It also provides the incentive to create the best possible ranking of results.

If having a scale advantage implies that a dominant firm has an advantage at providing accurate rankings, it also gives the firm wiggle room to manipulate results. If a firm advantages its own affiliated websites, or delay releasing innovations that might be good for users but distract users from clicking on the ads that make you money, it doesn't suffer too much in terms of lost users if it already has a quality advantage. Edelman and Lai (2013) provide some empirical evidence about manipulation in search, which has been one of the main focuses of the European Commission's ongoing investigation into Google. See also Edelman and Wright (2012). As discussed below, Athey (2013) provides some evidence from a real-world field experiment about the possibility of manipulation.

On the advertiser side of the market, advertisers are willing to pay the fixed costs of joining a platform and maintaining campaigns if there is enough traffic. Why would a small business bother joining

a search engine that just brought a few clicks a month? The more traffic a search engine has, the more advertisers it attracts. In Europe, only a small fraction of mid-to-small sized advertisers advertise on Bing in addition to Google. A dominant search engine realizes that it is a “must buy” for online advertisers, and that it can exercise market power, raising prices without losing advertising dollars. Search engines can engage in very sophisticated pricing schemes, where each advertiser gets a personalized price. The algorithms are not transparent, and so it is difficult for an advertiser to know why their prices might suddenly one day increase. It could just be that they got a bad draw from the algorithm, or it could be that they have been deliberately penalized. New businesses are particularly vulnerable to excessive prices charged by search engines since they rely more heavily on search advertising to be discovered initially.

Now consider a less obvious side of the market, web publishers—the innovators who create original content on the internet, with advertising revenue as their primary revenue source in many cases. Many web pages have a little search box in the corner. Those boxes can be very efficient for publishers, since they don’t take up much room on the page, and they are directly useful for consumers. Search engines compete to provide the results when people search there, because that generates more revenue and gives them more user data. Search engines share some of the revenue back with the publishers. How much? That is determined by competition. If the second place search engine is small and doesn’t attract many advertisers, it doesn’t have as much revenue to share. But then, the dominant firm doesn’t have to share much either, as the competition is not very strong. As a result, the dominant search

engine keeps more of the revenue generated by searches. Since the publisher then gets less revenue for every user, there are lower incentives to create web sites and innovate. On the other hand, if there are two evenly matched search engines, most of the revenue goes to the publisher. In large deals, sometimes a smaller search engine will offer to share with publishers an amount in excess of 100% of the revenue they expect to generate, as a way to increase scale. Anticipating that, a dominant search engine might offer to share even more, as a way to prevent the smaller search engine from gaining scale and becoming a more effective competitor. Keeping the competitor small allows the dominant firm to spend less on R&D to attract users, and to share less revenue with all other publishers in the future on the publisher side of the market. Linn (2005) describes the bidding between Microsoft and Google for a deal to serve searches for AOL, in which both sides bid aggressively in order to capture the large chunk of consumer traffic. Google eventually paid more to AOL than what it reportedly earned in advertising revenue.

Thus, competition in search advertising affects welfare on all sides of the market: users, advertisers and publishers. Behavior on one side of the market can affect all other sides. Even though search is “free” to users, search engines make decisions that affect consumer welfare, such as investments in R&D and decisions about whether to demote links to competitor sites. Competition creates incentives for firms to make decisions that are more beneficial to the constituents on all sides of the market.

Stutz (2011) provides a detailed discussion of antitrust issues involved in one of the many examples of vertical acquisitions and

mergers in search from the last few years, Google's acquisition of ITA, the online travel search company.

Now consider display advertising. Most internet publishers, particularly online news, relies on advertising as a major or primary source of revenue. Athey, Calvano, and Gans (2013) discuss some impacts of the internet on markets for online advertising in news media, taking the perspective of the theory of multi-sided markets.

Evans (2009) provides an overview of the online advertising industry. Internet publishers may choose to either sell their own "inventory" (spaces on their web pages) directly to advertisers, or they may rely on a third party display advertising platform to provide the ads. In the latter case, the advertising platform selects the ads and charges the advertisers, and shares the revenue with the publishers, less the platform's fee for providing the matchmaking service.

In some ways, the economics are similar to search advertising, with some important differences. First, in search, a lot of user traffic is "Owned and Operated," meaning that, for example, users go to Google.com to conduct a search, and Google shows search ads on its own site. Above, we discussed the publisher side of the market, where search engines get traffic from third parties, and share some of the advertising revenue with them. In display advertising platforms, most of the "inventory" of advertising space comes from firms other than the owner of the advertising platform. In addition, the advertisements are usually either only loosely targeted at users (e.g. all users on a particular news website) or targeted by user

demographics or past behavior (as inferred by past behavior and recorded in “cookies”). Unlike search, the user’s intent on a page is often only loosely related to the advertising, and the user is often not in the mindset to click on advertising links when, for example, reading online news. Advertisers typically view the display and search markets as distinct.

Levin and Milgrom (2010) provide a discussion of some market design issues in online advertising markets. There are a number of subtleties related to the fact that advertisers may be differentially informed about the value of individual users, due to differential access to cookies. See also Abraham, Athey, Babaioff, and Grubb (2014) for a discussion of the challenges created for auction design in this context.

Competition in display advertising markets has large effects on the internet’s content creators, such as the news media, but also innovative websites that provide consumers information about things like travel. If two display advertising marketplaces are similarly efficient, then they will compete strongly for “inventory” from publishers, charging low fees. On the other hand, if one display advertising platform has a large advantage (e.g. a larger advertiser pool, or unique data about users, their interests, or the set of advertisements the users have seen), they can attract publishers without sharing as high a fraction of revenue with the publishers. That is, they can charge higher fees.

Higher fees charged by display advertising platforms translate into lower revenues for online publishers, and reduced incentives to create content.

The prospect of collecting higher fees gives ad platforms the incentive to amass as much data as possible about users, as well as to ensure that competing ad platforms do not have access to this data. Clearly, this incentive will figure prominently when considering privacy policies.

### **Overall Importance of Competition in Online Advertising**

Summarizing, competition in internet search and in online advertising broadly is very important for welfare. In principle, data creates enormous value for advertisers and publishers, but those parties only benefit from the data if the gains are shared with them rather than paid to ad platforms as fees.

Competition policy must deal with the complexities of the multi-sided nature of the markets. It must also confront the fact that when a business is an information gatekeeper, increased competition may not lead to better behavior. Instead, it can lead to worse behavior from a welfare perspective: the information gatekeeper may choose to decrease rather than increase its quality as a gatekeeper in order to divert customers away from its competitors.

### 3 Information Gatekeepers, Data, and Competition on the Internet

In the early days of the internet, many observers predicted that search costs for consumers would be dramatically lowered by the internet, and thus e-commerce would soon resemble a stylized model of perfect competition. Prices for products would quickly converge to the marginal cost of the products, and consumers would benefit greatly.

However, this view was resoundingly rejected by empirical evidence that showed that price dispersion remained alive and well on the internet. Although search cost did go down, one thing that early observers missed was that the internet would also bring an enormous amount of new content, and so search might still be challenging. (Of course, there are other economic forces behind price dispersion as well). Brynjolfsson, E., and M. Smith (2000), Baye and Morgan (2001), Baye, Morgan, and Scholten (2004), and Ellison and Ellison (2009) all document price dispersion on the internet; see Ghose and Yao (2010) for a survey of some of the findings.

#### **Search Engine Rankings and Consumer Choice**

Perhaps even more unexpected was the fact that the internet didn't just fail to make search *across websites* costless, but that consumers still behave as if search costs are important even when a set of results is presented in an ordered list on a *single page*, such as on search engines, which help roughly two billion people find information on the internet. An extremely robust finding is that the order of search results matters, not just a little bit, but a lot; and

more broadly, the design and layout of web pages, the prominence of various elements of content, and subtle factors like color and font make a large difference in what links consumers click and thus what information they ultimately discover. A key component to how links (both ads and algorithmic results) are ranked in internet search engines is how well these links have performed in the past in getting clicks. Thus, the amount of data available to a website to rank content is crucial to its quality.

As search technology has evolved, search engines have evolved far beyond the original “10 blue links” from the early days of Google. Today, search engines have become “commerce platforms,” pointing people to the products and services they seek by connecting potential customers to a growing number of online business.

A key question for efficiency in internet services concerns what happens if a search engine decides to place links to its own products and services at the top of the search results page, crowding out results identified by its algorithm as the best. This can have a huge effect, if search rankings are impactful.

Even though the basic fact that the prominence of results is important is widely known throughout the Internet ecosystem—anyone who designs a Web page and compares alternatives will quickly discover it—it is still worthwhile to quantify just how important it is. Thus, I conducted an experiment (Athey, 2013) to evaluate the impact of ranking. Prior to my experiment, most of the evidence that was available in the public domain was based on non-experimental evidence, which therefore did not give the *causal*

effect of position. Did the top link get clicked because it was the best link for the search query, or because it was in the top position? What would have happened if the links in the top position and a lower position were reversed?

With a randomized experiment, where different users see different rankings of links or different layouts on the screen, it is possible to address this question more definitively, and to avoid trying to generalize from specific examples that may not be representative. Search engines regularly run experiments to test out the performance of new algorithms. In these experiments, user searches are randomly assigned to either receive the “control” treatment—the baseline search experience—or one of a number of experimental “treatments,” where results are ranked or presented differently.

In order to answer the question about the potential effects of manipulation, I worked with the Bing team at Microsoft to design a special experiment, analyzing the impact of several “treatments” in which we moved the best search result—the one that our algorithms would otherwise place first—to various lower positions on the search results page. The test ran for a few weeks, in the United States and overseas.

The data spoke very clearly about the impact of the treatments: A search engine can divert traffic from one website to another by manipulating the order of search results. In particular, moving the best result down just two positions (from first to third) reduced traffic to that site by half. The diversion effect becomes much more pronounced as a site is moved further down the page. A site that is

moved from the first position to the tenth position typically will lose about 85 percent of its traffic. A site that is moved from the second position to the ninth loses about 75 percent of its traffic. And the results were similar for all users, regardless of the amount of time they spent searching on the site.

If you look at the same results from the perspective of the site that gets promoted from a lower position to first in the rankings, the effects are even more pronounced. (This is because the site appearing further down the page has so few clicks to start with). A site promoted from fifth to first gets a 340 percent increase in visitors from search, and the results are similar when you focus only on users who go to the site and stick around for a period of time. Imagine telling a business that they can more than quadruple their customer base overnight! That is a very tempting thing to do for a search engine if the site it is promoting is its own affiliated website.

Joaquin Almunia, the European Commission vice president responsible for competition policy, stated that this is not just an academic concern: “Google displays links to its own vertical search services differently than it does for links to competitors.” Vice President Almunia explained that “[w]e are concerned that this may result in preferential treatment compared to those of competing services, which may be hurt as a consequence.”

The impartiality of search results will become all the more important in the years to come given that screen sizes on smartphones and tablets are smaller than on traditional PCs.

Smaller screens mean there is even less room for competing services to appear in mobile search results.

### **Search Engine Manipulation**

Manipulation is an important issue for competition policy. Imagine that you have created an amazing new travel website. But, the website needs to attract consumers, and it is based in a country where a single search engine has 90% or more market share. How will you attract users who are interested in travel? At the moment the user does a search query about travel, the user is ready to do travel research. If the website is terrific, perhaps the search engine will send the user to the site! Or perhaps, the user will click on the search advertisement you have purchased. Imagine your dismay if you wake up one morning and see that the search engine has taken the most prominent part of the screen and embedded its own travel content, and your link falls off the page, and traffic plummets. The years of hard work and investment to make an innovative and brilliant product do not matter, nor does it matter whether your website is better than the content that replaced your link. Simultaneously, the prices for your search advertisements skyrocket, ending up above the profit you can make from a click on the links. Scenarios like this have been described by a variety of websites in different industries throughout Europe, and the European Commission filed a complaint against Google for engaging in this type of behavior.

In response to the EC's concerns about manipulation, Google proposed a set of commitments in October 2013, which included the addition of a Rival Links box that would purportedly restore traffic to rivals. The EC rejected these commitments for their

failure to end the preferential treatment of, and traffic diversion to, Google's own specialized results. In January 2014, Google proposed a revised set of commitments. The EC is expected to make a final decision as to the merit of these commitments by the end of summer 2014.

In May 2014, the Open Internet Project (representing 400 companies including major German and French publishers) announced they were suing Google for anti-trust violations. The group demanded a "ban of Google's manipulative favouring of its own services and content." In addition, issues of search manipulation are often brought up by regulatory bodies who investigate monopolistic behaviors and the appropriateness of vertical and horizontal mergers.<sup>3</sup>

There are not very many academic studies of the relationship between internet search and consumer welfare. Much of the existing theoretical work focuses on the interaction between search technology and the prices charged to consumers by advertisers. Examples of such papers include Chen and He (2006), Armstrong, Vickers and Zhou (2009), and White (2008). White (2008) analyzes the tradeoff between high quality search results and paid search profit for the firm, where paid search profit depends on the profits that advertisers generate from consumers.

<sup>3</sup> E.g., the UK Office of Fair Trading's investigation into Google's acquisition of BeatThatQuote, a provider of consumer finance comparison services

Taylor (2010) develops a model where high quality algorithmic results divert clicks away from advertisements, creating an incentive for the search engine to degrade algorithmic search quality. He additionally finds that when consumers exhibit search engine loyalty, the incentive to manipulate leads to a ceiling on equilibrium search engine quality.

Athey and Ellison (2011) build a model of consumer search that analyzes a consumer's decision to click on each successive link, formalizing the feedback effect between the quality of ads and the propensity of consumers to click on ads. Consumers are rational, understanding that higher quality firms bid more aggressively to attain the top positions. They use this model to analyze how market design decisions such as the level of reserve prices affect welfare for advertisers, consumers and the search engine, showing that a profit-maximizing search engine selects reserve prices in a way that sacrifices welfare in favor of extracting revenue from advertisers.

Building on Athey and Ellison's (2011) model, Athey, Kuribko, and Richards (2014) analyze the incentives of a search engine to manipulate the rankings of algorithmic links. The model features rational consumers who recognize the possibility of manipulation and respond optimally in their search patterns. Despite the rational response of consumers, the paper shows it is still profitable for a search engine to manipulate. Consumers respond to manipulation by clicking on more links, which lowers their welfare but avoids harming the search engine, so long as not too many consumers give up. The incentive to manipulate is enhanced

if the firm enjoys a quality advantage over other options, and thus can manipulate without losing too many consumers.

### **News Aggregators and News Consumption**

Another context in which the way information is displayed matters a great deal is in news. Internet search is one of the most important sources of referrals to news. Imagine for a moment that there was only one search engine, and it had a political bias. As Timberg (2013) reported:

Google's motto is "Don't be evil." But what would it mean for democracy if it was?

That's the question psychologist Robert Epstein has been asking in a series of experiments testing the impact of a fictitious search engine — he called it "Kadoodle" — that manipulated search rankings, giving an edge to a favored political candidate by pushing up flattering links and pushing down unflattering ones. Not only could Kadoodle sway the outcome of close elections, he says, it could do so in a way most voters would never notice.

"Elections are won among low-information voters," said Eli Pariser, former president of MoveOn.org and the author of "The Filter Bubble: What the Internet Is Hiding From You." "The ability to raise a negative story about a candidate to a voter . . . could be quite powerful."

With my coauthor Markus Mobius, I conducted another study (Athey and Mobius, 2012) to analyze the impact of news aggregators on the content and diversity of content users read.

The research was inspired by the popular debate about the role of news aggregators, which we will define here to include sites that do not produce much original content, but rather curate content created by others using a combination of human editorial judgement and computer algorithms. The results are presented with a few sentences and perhaps photos from the original article; to read the full article, users can click through and go to the web site of the original content creator. Pure “aggregators,” such as Google News, generally do not make any payments or have any formal relationship with the original authors of the news content; rather, they create their page by “crawling” the web and then using statistical algorithms together with editorial judgements to organize and rank the content. (Observe, then, that this is another context where having more data leads to better rankings, all else equal.)

Only in a few cases does Google News have a direct relationship with the outlets (e.g. Google News had a relationship with the Associated Press, as analyzed by Goldfarb and Tucker (2011)). In contrast, sites like Yahoo! News and MSN primarily show content from contractual partners. Sites like the Huffington Post may use a hybrid strategy of curating blogs and aggregating news from other sources.

Why are aggregators so controversial? Only about half of page views on the Google News home page result in visits to any online newspapers; thus, users may read their news from Google News without ever generating any page views or revenues for any of the content creators. Clearly, this undermines the incentive of newspapers to invest in journalism. In addition, news aggregators

can substitute for the home page of an online news outlet like the New York Times. The aggregator can index not just the content of the New York Times but all other news outlets, giving it an advantage in coverage. It may then replace the “curation” function that gives the New York Times its reputation.

The question of how aggregators impact consumption was first studied by Chiou and Tucker (2011). They study a “natural experiment” where Google News had a dispute with the Associated Press, and as a result, did not show Associated Press content for about seven weeks. The paper has aggregate data about page views to Google News as well as the sites visited immediately after Google news. They use views to Yahoo! News as a control. The paper finds that Google News is a complement to news outlets: taking the Associated Press content away from Google News lead to fewer visits to news outlets (where Associated Press articles are featured).

Athey and Mobius (2012) consider a different application, when Google News added local content in France in late 2009. The paper uses internet browsing data from a subset of internet users to analyze how the content of user consumption changes. They find, similar to Goldfarb and Tucker (2011), that introducing local content increases the consumption of news. The interpretation is that by exposing users to local news that they might not have otherwise read, they increase users’ interest in news and thus their news consumption. However, the paper also finds that users become much less loyal in their local news consumption, and that the role of the newspapers’ home pages (and thus their editorial contribution to news curation) decreases.

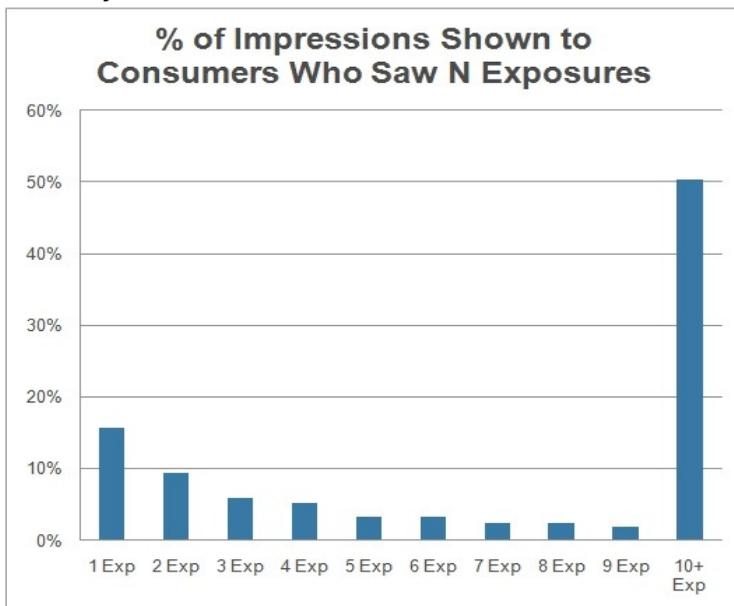
Taken together, these studies reinforce the finding that the way news is presented, and whether it is presented at all, matters a great deal. The information intermediaries, such as search engines and aggregators, are enormously powerful in picking winners and losers on the internet, and in determining the informedness of the population, broadly defined.

I am currently working with Markus Mobius on follow-on research to understand the role of social media in determining the kinds of news people read. In preliminary findings, we show that people read different types of news through social media. For example, we find that within a user, articles read through social media tend to take a more caring tone, and to show more individual perspectives on the news. We also find that readers who read liberal political news tend to skew much more liberal in their reading from social media than in their other reading, and similarly conservative readers read more conservative articles in social media than their other reading. Overall, these preliminary findings support the premise that users are influenced by their context and the information that is presented to them, and also that we should expect to see greater impact of data and personalization on news consumption in the future, especially as social media becomes more important as a source of news.



## 4 The Impact of User Multi-Homing on Publishers

We have presented evidence that internet search and news aggregators can greatly affect the way consumers view information on the internet. News aggregators lead to more diverse browsing patterns. When an audience views content from a variety of different websites, it becomes more difficult for advertisers to control how many times the show ads to the same user. In 2010, ComScore, a U.S. firm that tracks web browsing for a panel of users, provided me with a custom analysis of 30 very large, cross-media online ad campaigns. Most advertisers attempt to reach a given user between 3 and 7-8 times with the same advertisement. Advertising is considered “wasted” if a user sees the ad too few or too many times.



The chart above illustrates the percentage of impressions shown to consumers where there was waste. Clearly, the technology used by the advertisers to attempt to manage wasted impressions did not fully accomplish the objective.

The inability of advertisers to avoid wasting ad impressions has a depressing effect on advertising prices, both because it affects efficiency and because it affects the way prices are determined. Athey, Calvano, and Gans (2013) analyze a model with partially multi-homing consumers, showing in such a world, high-value advertisers buy ads on multiple outlets in order to achieve maximum “reach,” accepting that some impressions will be wasted. Lower-value advertisers, whose value determine the market prices of ad space, tend to focus on one or a small number of outlets, as they find that once waste from duplicated impressions is accounted for, the price of ad space is greater than the benefit to them. In contrast, when they focus on a single outlet, there is no waste, so the impressions that they buy are worthwhile. The paper further shows that a decrease in consumer loyalty (an increase in switching) generally leads to lower advertising prices, and that outlets that have a broad “reach” can command higher advertising prices, creating a force in favor of consolidation.

Problems of fragmented viewing patterns in online news are exacerbated as consumers shift to doing more and more of their internet activity on mobile devices. Due to the increased use of apps to access the internet instead of web browsing, it becomes more difficult for advertisers to track users across multiple outlets and devices. Indeed, the ecosystem has been described as “fractured, complex, and hugely important” (Baye, 2013).

Further, “In mobile advertising, the rules of the road change with different combinations of device, wireless operator, and operating system. And there are few shared protocols or standards: Mobile lacks the technical consensus that enables ad targeting, delivery, and measurement to work fairly seamlessly across the desktop world.” Consistent with this, mobile so far attracts only 3% of the ad dollars, even though it is responsible for 12% of utilization time (Danova, 2014).

There are a number of possible concerns about online advertising on mobile devices, as well. Mobile devices are much more tightly controlled than, for example, personal computers running Microsoft Windows have been historically, or browsers in the PC environment. Applications must meet guidelines to be allowed into “app stores,” for example. It is conceivable that a mobile platform could, through policies or technology, eliminate the possibility for third party ad platforms to operate on a given mobile device platform. This could eliminate competition and allow the platform’s own advertising platform to charge high fees. Similarly, regulation that prevented all data sharing across apps could disadvantage small, independent app providers, since they would have no way to know enough about consumers to serve relevant ads, and thus no way to monetize their apps. As a result, niche apps might not be invented, or might not get as much funding and investment.



## 5 An Economic View on Privacy

In order to effectively balance the costs and benefits of regulation, privacy policy must be formulated against a backdrop of the way competition and innovation on the internet work. We have reviewed research that suggests that eliminating the ability to track consumers will have large implications for the incentives for firms to create specialized news websites and niche apps. We have also highlighted how important data is as a strategic asset for firms.

Let us now review a simple framework for evaluating the costs and benefits of privacy regulation. See Acquisiti (2010) for a more complete discussion of costs and benefits, as well as Goldfarb and Tucker (2011).

To start, consider the goals and potential benefits of privacy regulation. Any thorough discussion should begin by analyzing the different reasons we care about privacy. There are a few main categories, many of which have been introduced in this paper already:

- Direct economic harm to a consumer from the use of the data, e.g. a user being denied insurance, employment, credit, etc. on the basis of the data, or experiencing reputational harm or legal difficulties.
- Direct economic harm to a consumer from a security breach involving the data, e.g. a stolen identity, public embarrassment, harm to career or reputation
- Direct economic benefits from the use of data, e.g. personalized services, more relevant ads

- Indirect economic benefits from the use of data, e.g. the existence of web sites, political news and investigative journalism, and ad-funded services
- Indirect economic benefits from the security of data against other nations and from surveillance to protect a nation, e.g. national security
- Consumer intrinsic value (“feelings” or “utility”) about personal considerations such as being treated fairly, having their personal rights respected, not being tricked, and sharing in value creation from something they feel is theirs
- Consumer intrinsic value (“feelings” or “utility”) about the values incorporated in the economic and political systems in which they live

There are a number of ways in which markets don't function well when it comes to privacy. First, many consumers are not informed about privacy and do not understand the risks and benefits of alternative privacy policies. This is due to a combination of factors, but one appears to be the complexity of the policies and the lack of standardization in the industry. Another key factor is that it is objectively difficult for even experts to learn how governments and other entities make use of data, as well as how it might be used if a company suffered a security breach. Thus, the true implications of privacy policies are extremely difficult to understand. Indeed, many of the case studies where firms have been forced to change their privacy policies have come about when the news media covers new revelations from experts, insiders, advocacy groups, governments or competitors. As a recent example, Google publicly “shamed” Comcast, a major U.S. cable provider and internet service provider, into encrypting email exchanged Gmail, by publishing a

list of which email providers use encryption (Constantin, 2014). Without this public announcement and subsequent media coverage, it would have been difficult for typical users to ascertain that their provider was not following industry best practices.

This first factor contributes to the second, which is that although privacy policy changes can generate a public outcry from informed individuals, there is only modest evidence that consumers change their purchase and utilization behavior in the short term in response to broad differences in policy, and long term effects would be hard to measure without a long term experiment. Thus, regulation can play a role, if this is something that consumers do indeed value (or should value, according to some objective criteria, if the government is paternalistic). (To be fair, there is some experimental evidence in favor of consumers responding somewhat to treatments such as prominently displaying privacy reminders (Tsai et al, 2011).)

Third, for many technology products, markets are highly concentrated, and consumers do not perceive choices that are different enough on privacy policy to understand and consider. Thus, a policy of “notify and consent” may not seem meaningful if there is no comparable alternative to a company’s product, or if the consumer has already invested in learning and using a company’s product and does not want to switch when privacy policies are changed. And how can they know whether competitors will also change their policies? Thus, the incentive for consumers to “punish” a firm for poor privacy policy is low.

Fourth, it is very difficult for firms to measure the consumer benefits of improved privacy policy in a coherent way. It is hard to measure something that consumers do not understand. In particular, consumers may not understand how to compare privacy alternatives, and they may have never lived in the counterfactual worlds we might ask them to consider. For example, they may not be able to anticipate how they will feel if advertisers serve behaviorally targeted ads on unrelated web sites that rely on information from their email. Consumers might have told researchers that they did not want their email providers to serve targeted advertisements based on the content of their email, but Gmail still gained wide adoption despite using this practice. Acquisti et al (2011) reinforces the finding that privacy preferences are context-dependent. Research by the World Economic Forum (2013) also argued that privacy preferences are highly context-specific and vary across countries. One piece of context they identified as important is whether or not consumers perceive that they get something of value in exchange for their data, preferably something related to the use of the data. This finding has been echoed by other industry observers as well (Smith, 2013). Although policy makers face a similar challenge, the incentives of policy makers differ, and they may be more likely to account for consumer welfare even when it is difficult to measure.

Fifth and related, consumers' preferences are very difficult to measure even when they relate to products and policies in their current environment, because the way that questions are framed and the supplementary information that is provided around choices can make a big difference. Acquisti (2013) establishes using a field experiment that users respond differently when they

have privacy and are offered money to give it up, than when they don't have privacy but are asked to pay to gain it. In this environment, it is difficult to believe that firms will be able to discern "true" preferences for privacy and respond to them; this favors a more direct approach to regulation that either makes policy alternatives simpler and clearer to consumers, or that directly regulates certain aspects of privacy.

Sixth, consumers may also change their feelings about the risks of a large firm retaining their data after news about government subpoenas or U.S. National Security Administration surveillance. Again, the signals firms would get from consumers about their preferences would be misleading in this case, since the revelation occurs after decisions have been made. Goldfarb and Tucker (2012) and Marthews and Tucker (2012) provide some evidence that views do indeed evolve over time in response "shocks" to information. Indeed, Marthews and Tucker (2012) show that users change their search behavior, reducing their queries on politically sensitive terms, after media reports about government surveillance. In such an environment, it is difficult to know how to put a dollar value on benefits to privacy protection to trade off against harm to long-term welfare, innovation, and so on. Government policy may struggle with the same problems faced by the private market. Still, there may be a role for regulation in an environment where the market gives almost no weight to a potentially relevant concern (like the effects of government surveillance on the costs and benefits of long-term data retention about individual users, for example).

Seventh, regulation may correct for the free rider problem faced by consumers—no individual has the incentive (nor the expertise) to

audit major technology firms with which they interact. It may also provide expert opinion about what is important.

Eighth, if one reason consumers do not factor privacy into decision-making is that it is too difficult to understand policies and their consequences, regulation can, in principle, provide organized and expert-designed information to consumers about these choices, focusing on important attributes of a privacy policy, and making it simple for consumers to choose between a clearly defined set of alternatives.

Now consider some potential harms from privacy regulation. First and foremost, as discussed above, privacy regulation that interferes with the effectiveness of online advertising makes it hard for new ventures to attract an initial user base, as well as hard for new ventures to monetize their content by showing ads and sharing the revenue from them. In short, decreases in efficiency in online advertising lead to decreases in innovation and the creation of content. For example, Miller and Tucker (2014) show that state level protections for genetic testing have a mixed impact. Providing users with informed consent leads to the users not purchasing the product.

Second, harming the efficiency of online advertising is typically regressive. Advertising supports free products. Low income people appreciate ad-supported free products more than wealthy people. For example, free productivity software such as Google Docs or Office Web Apps are especially beneficial to students, new or small businesses, and low-income people. (Note that these services do not currently show ads directly, but they tend to

increase user utilization of related services offered by the same firm that are ad-funded.)

In a more striking example, two studies by Miller and Tucker (2009, 2011) together imply that U.S. states adopting especially stringent privacy laws decreased the adoption of electronic medical records systems, which in turn increased infant mortality (which is very high in the U.S. among poor women, where not all women have access to prenatal care). The evidence suggests that disadvantaged women were harmed by the lack of availability of medical information when they came to the hospital. Certainly, those who designed the privacy policies did not account for the fact that these policies would end up leading to the death of economically disadvantaged babies; yet this example, while extreme, is not at all isolated. Data enables services that help the most vulnerable.

Third, past attempts at privacy regulation have resulted in behavior where privacy policies are typically too difficult to read. There is little evidence that the way “notice and consent” has been implemented across a wide range of firms has had much impact on consumer behavior; indeed, only a tiny fraction of users read such notices, and an even smaller fraction understand them.

Fourth, as discussed above, it is difficult to measure welfare benefits provided to consumers, and so it is hard to propose efficient strategies. In addition, views are changing, so it is hard to predict long term benefits. Although expert policy-makers may do better at assessing the factual information about privacy policies and at understanding various risks, they may not have much

advantage at putting a dollar value on the benefits consumers get from feeling like their data is used fairly or that their data is private.

Fifth, in concentrated sectors privacy regulation can be used by incumbents to keep users out. For example, incumbent firms may have more data than entrants for targeted advertising. Privacy policies can make it hard for small, new firms to get a toe-hold (Goldfarb and Tucker, 2011). More established firms may also have better luck getting consumers to read a long disclosure agreement. As discussed above, this may favor some specific business models, as well. Niche firms, tailored “hard news” websites, and entrepreneurs may be particularly disadvantaged by privacy policies. If one believes that competition from a new generation of firms is one way to keep established, large firms disciplined in terms of privacy policy and other areas, then discouraging entry can be counter-productive to the larger policy goals and social welfare. Campbell, Goldfarb, and Tucker (2013) explore a model that develops this point in more detail.

What kinds of policies, then, have some hope of balancing the costs and benefits appropriately? Burt (2013) reported on one kind of proposal, from Craig Mundie:

Microsoft’s Mundie has been promoting changes around both privacy policy and technology to address the new ways data is being collected and used. He said cryptographic wrappers and metadata could be used to give people more control in how their data can be used, and laws could be put in place to ensure that businesses and

government agencies follow rules in the metadata that dictate how the information is used. And how large a legal penalty should there be for companies that violate the rules in the metadata? “Personally, I’d say make it a felony,” Mundie said. “Otherwise, the penalty is too low to deter that behavior.”

Mundie’s proposal has the feature that technology enabling efficiency enhancing innovations such as targeted advertising could be used, but users could control the use of the data even in a complex ecosystem whose details they might not understand. “Metadata” could incorporate user preferences, while users of the data would have freedom to develop new technology so long as they respected user preferences.

This proposal can be thought of in a broader context where regulation helps establish property rights. Property rights are a broad concept that can be applied even in a world of fast-changing technology and across many contexts. There is some hope that consumers can learn to understand what it means to own their data and allocate property rights.

This kind of proposal can be contrasted with an approach of trying to ban particular technologies. Policy aimed primarily at, e.g., cookies can be undermined through the use of other technology that accomplishes a similar goal, and cookies may not even be relevant in new form factors or settings (like the “internet of things,” the “smart home,” “wearables,” or mobile.

Another type of policy is one that attempts to provide broader protection through limits on data retention. Chiow and Tucker (2014) argue that data retention limits may pass the cost-benefit test, providing evidence from recent data that changes in retention policy did not change the quality of search engine results. Their finding is consistent with general industry understanding that recent data is much more important for predicting what consumers want. A potential policy would limit the retention of data, and require it to be anonymized and/or aggregated after a certain time period.

Although there is always some value to having older data, particularly for research and development and for analyzing trends over time, there are also large potential costs to keeping that data. To see why, let us take the perspective that an individual values privacy because of the risk of economic harm or reputational risk due to discovering information about the individual. (Of course, there are many other perspectives on privacy, as outlined above.) Note that there may be many sources of information about an individual's current behavior. One could observe their shopping physically, for example. On the other hand, over time, it is more likely that a user might have changed their preferences and behavior, and thus face some costs if their previous behavior was revealed. At the same time, as time passes, there are fewer and fewer ways for an outsider to find detailed data about a user's past behavior, other than the digital data retained by online firms. Thus, eliminating the digital data has a material impact on the risk that the information is revealed.

A natural alternative is to regulate the use of old data rather than its retention. However, it is very difficult to anticipate or even understand how and why historical data is harmful, and thus difficult to regulate all the different uses that could be harmful. Furthermore, a security breach might occur. If the data does not exist (or exists only in anonymized form), then a single security breach is less likely to expose harmful information.

Limits on retention are also easy for consumers to understand (though it may be more subtle to understand residual risks of the retention of “anonymized” or aggregated data). A consumer can have confidence that something that happened two years ago is more or less “gone” unless they have specifically opted in to retention (e.g. retaining old credit card or bank statements, or historical orders on an e-commerce site, which are easier to remember than website viewing or shopping). These give users a feeling of control, and may create more utility for consumers if part of the value to consumers is not having to worry about the unknown or about technologies they don’t understand.

Limits on retention may seem like a blunt instrument, but such limits also provide blunt protection against a wide range of issues, including security breaches as well as unwanted use of data or government surveillance. Although historical data does have real value, and in some contexts (such as studying health conditions that develop over many years) it may be indispensable, in many online contexts, the benefit of long retention of non-anonymized historical data may not outweigh the privacy costs and risks. If limits on retention help consumers become more comfortable with richer uses of current data, and thus policy permits the use of current data to create more value and efficiency (for example in

online advertising for small websites and apps), such a policy may have substantial welfare benefits.

### **Case Study on Retention and User Awareness**

It may also be important to have independent “auditors” help interpret the actual practices of internet firms. For example, Google “history” is turned on for many or most Gmail users. Depending on the date on which a Gmail account was created, users were automatically opted into this service, and as of June 2014, the Google “About Google Web History” website states “When you create a Google Account, Google Web History is automatically turned on.” Even though Google publicly states that it only retains search logs for a limited period (e.g. 18 months), Gmail users may discover that their entire search log history as well as their entire internet browsing history from the creation of their account is stored by Google (a user’s history can be found by logging into Gmail and then typing <http://www.google.com/history/>). This author’s history, for example, can show the daily search activity in September 2006, or the set of political news articles read on a particular day several years ago in a different tab of a browser where I was reading Gmail.

Google’s “About Google Web History” website also gives the following information:

Google Web History saves information about your activity on the web, as well as details about your browser, including:

- Searches on Google
- Pages you click on from the search results page
- Results that appeared, including private results from Google products like Google+, Gmail, and Google Calendar
- Ads you respond to by clicking the ad itself or completing a transaction on the advertiser's site
- Your IP address
- Your browser type and language
- Your searches and other activity on Google Maps, including maps around the web

In terms of how the information can be used, Google refers users to the general privacy policy, which states: "We use the information we collect from all of our services to provide, maintain, protect and improve them, to develop new ones, and to protect Google and our users. We also use this information to offer you tailored content – like giving you more relevant search results and ads."

This example is interesting that even in the case of a major service provider, most users are not aware that their searches and internet web browsing from many years ago can still be used for advertising. Users may not have even realized that just because they logged into Gmail, they were providing Google with all of their web browsing, on and off of Google sites. It is not clear how much value is created for users or for the efficiency of advertising through the retention of this data, and it seems unlikely that most users desire the retention of all of this data.

Presumably government surveillance of data from such a long time frame could also be problematic. There is no public information about what fraction of Google users has disabled Web History, nor the average number of years of data per user.

# Conclusions

This paper has attempted to survey a large volume of literature to understand the role of data, gatekeepers, information, internet search, and advertising. This paper argues that bringing an economic framework to bear is essential for achieving beneficial policy outcomes.

The paper has also highlighted some subtleties in achieving a broad set of policy objectives. For example, if one values the services and information created by small entrepreneurs and new ventures, and if one believes that these new entrants are a key source of competitive pressure on established incumbents in the technology industry, then it is important to consider the impact of public policy on that ecosystem. Privacy policy can hurt these innovators by decreasing the effectiveness of advertising and thus their ability to monetize their content and attract new users. Competition policy may be necessary to ensure that the gains from advertising accrue to these innovators rather than ad platforms, intermediaries, and information gatekeepers; and it may be needed to protect these innovators from manipulation in search and discovery on the internet by incumbents who fear that the innovators may grow into competitive threats.

More broadly, this paper has argued that privacy policy needs to consider carefully economic costs and benefits, and it must also be sensitive to the mechanisms through which firm behavior is impacted. Relying on uninformed individual consumers to police firms through “notice and consent” policies is unlikely to result in

efficient outcomes. Policy should recognize the limitations of markets in environments where consumers get limited return from the substantial investment they would need to make to understand how privacy practices impact them. Themes in effective policy include simplifying and standardizing information, and making sure that the most important aspects of privacy from a cost/benefit perspective are highlighted to consumers in ways they can understand. In some cases, there may be industry standards that should be enforced by governments, since consumer behavior cannot be relied upon to provide sufficient incentives.

More robust policies may include the establishment of property rights for data, which at least have the potential to allow the efficiency benefits of using data for personalization to be realized, as well as broad measures such as limits on retention that are easy for consumers to understand and also solve a wide range of potential privacy and security concerns simultaneously, without limiting technology. Even retention policies must be carefully considered in each domain, however, because in some domains (such as health), longer retention of data may be justified.

## References

- Abraham, I., S. Athey, M. Babaioff, and M. Grubb**, "Peaches, Lemons, and Cookies: Designing Auction Markets with Dispersed Information," Working Paper, Stanford Graduate School of Business, 2014.
- Acquisiti, A.**, 2010, "The Economics of Personal Data and the Economics of Privacy," Working Paper, CMU.
- Aggarwal, G., Ashish G., and R. Motwani.** "Truthful auctions for pricing search keywords," *ACM Conference on Electronic Commerce*, 2006.
- Armstrong, M., J. Vickers, and J. Zhou.** "Prominence and consumer search." *The RAND Journal of Economics* 40, no. 2 (2009): 209-233.
- Athey, S.** (2013). "The Importance of Search Result Location," Microsoft on The Issues Blog, March 25, 2013.
- Athey, S., Kuribko, N. and G. Richards** (2014). "Search Engine Incentives." Working Paper, Stanford GSB.
- Athey, S., and G. Ellison.** "Position Auctions with Consumer Search." *The Quarterly Journal of Economics* 126, no. 3 (2011): 1213-1270.
- Athey, Susan and Markus Mobius**, "The Impact of News Aggregators on Internet News Consumption: The Case of Localization," February, 2012.
- Baye, M., and J. Morgan.** "Information Gatekeepers on the Internet and the Competitiveness of Homogeneous Product Markets." *American Economic Review* 91, no. 3 (June 2001): 454-474.
- Baye, M., J. Morgan, and P. Scholten**, "Price Dispersion in the Small and in the Large: Evidence from an Internet Price Comparison Site," *Journal of Industrial Economics*, Vol. 52, No. 4, December 2004, pp. 463-96.

**Beales, H.** (2010). The value of behavioral targeting. *Network Advertising Initiative*.

**Booker, E.** (2013). "Why Recommendation Engines Are About to Get Much Better." *Information Week*.

**Burt, Jeff** (2013). "Microsoft's Mundie: Data Collection Fuels Need for New Privacy Rules," *eWeek*.

**Brynjolfsson, E., and M. Smith.** "Frictionless Commerce? A Comparison of Internet and Conventional Retailers." *Management Science* 46, no. 4 (April 2000): 563-585.

**Campbell, J., Goldfarb, A. and C. Tucker,** "Privacy Regulation and Market Structure," Working Paper, MIT.

**Chiou, L. and C. Tucker,** 2013, "Copyright, Digitization, and Aggregation," *RAND Journal of Economics*.

**Chiou, L. and C. Tucker,** 2014, "Search Engines and Data Retention: Implications for Privacy and Antitrust," Working Paper, MIT.

**Clemons, E.K., and Thatcher, M.E.** 1998. "Capital One: Exploiting an Information-Based Strategy," *IEEE* 316, pp. 311-320.

**Danova, T.,** "Mobile Ad Spending Still Lags Way Behind Time Spent," *Business Insider*, May 28, 2014.

**Ellison, G. and Ellison S.,** 2009, "Search, Obfuscation, and Price Elasticities on the Internet," *Econometrica*.

**Edelman, B., and M. Ostrovsky.** "Strategic bidder behavior in sponsored search auctions." *Decision Support Systems* 43, no. 1 (2007).

**Edelman, B.,** 2011. "Revisiting Unlawful Advertisements at Google," [www.benedelman.org](http://www.benedelman.org)

**Edelman, B. and Z. Lai.** "Exclusive Preferential Placement as Search Diversion: Evidence from Flight Search," Working Paper, Harvard Business School, 2013.

- Edelman, B. and J. Wright**, 2012. "Debate on Antitrust Scrutiny of Google," *Journal of Law (The Post)*, pp. 445-464.
- Edelman, B.**, 2014. "Leveraging Market Power Through Tying And Bundling: Does Google Behave Anti-Competitively?"
- Eisenmann, T., Parker, G. and Van Alstyne, M.** 2006. "Strategies for Two-sided markets" *Harvard Business Review*, R0610F
- Eisenmann, T.** 2008. "Managing Proprietary and Shared Platforms", *California Management Review*, CMR 402, 50 (4), 30-53
- Evans, D.** 2009. "The Online Advertising Industry: Economics, Evolution, and Privacy", *Journal of Economic Perspectives*.
- Evans, D. and Schmalensee, R.** 2013. "The Antitrust Analysis of Multi-Sided Platform Businesses", *Roger Blair and Daniel Sokol, eds., Oxford Handbook on International Antitrust Economics, Oxford University Press*.
- Goldfarb, A. and D. Tucker.** 2011. "Comments on 'A Preliminary FTC Staff Report on 'Protecting Consumer Privacy in an Era of Rapid Change: A Proposed Framework for Businesses and Policymakers.'" Working Paper.
- Goldfarb, A., and C. Tucker.** 2012. "Shifts in Privacy Concerns." *American Economic Review*, 102(3): 349-53.
- Goldfarb, A. and C. Tucker**, "Privacy Regulation and Online Advertising," *Management Science*, Vol. 57 No. 1, January 2011, pp. 57-71
- Ghose, A. and Y. Yao**, "Using Transaction Prices to Re-Examine Price Dispersion in Electronic Markets," *Information Systems Research*, 2010, 1-23.
- Google Privacy Policy**, accessed June 3, 2014  
<https://www.google.com/policies/privacy/>
- Google About Web History**, accessed June 3, 2014.

[https://support.google.com/accounts/answer/54068?hl=en&ref\\_topic=14148](https://support.google.com/accounts/answer/54068?hl=en&ref_topic=14148)

**Google Research**, 2012. "The New Multi-Screen World: Understanding Cross-Platform Consumer Behavior."

**Hachman, M.** 2014. "Comcast plans to encrypt email exchanged with Google Gmail," *PC World*.

**Hagiu, A. and Yoffie, D.** 2009. "What's Your Google Strategy?" *Harvard Business Review*.

**Jain, K., A. Mehta, K. Talwar, and V. Vazirani.** "A simple characterization for truth-revealing single-item auctions." Proceedings of the First international conference on Internet and Network Economics. Hong Kong, China: Springer-Verlag, 2005. 122-128.

**John, L, A. Acquisti, and Lowenstein,** "Strangers on a Plane: Context-dependent Willingness to Divulge Personal Information," *Journal of Consumer Research*, 37(5), 858-873, 2011.

**Lee, R.** 2011. "Home Videogame Platforms", *The Oxford Handbook of the Digital Economy*, eds. Martin Peitz and Joel Waldfoegel, Chapter 3

**Levin, J. and P. Milgrom.** 2010. "Online Advertising: Heterogeneity and Conflation in Market Design", *American Economic Review: Papers & Proceedings*, 100 (2), 603-607.

**Levy, S.** 2011. "In The Plex: How Google Thinks, Works, and Shapes Our Lives." Simon & Schuster; First Edition edition (April 12, 2011).

Linn, A. 2005. "Google, AOL Deal Leaves Microsoft Spurned," *Mail and Guardian*.

**Mahdian, M., H. Nazerzadeh, and A. Saberi.** "Allocating online advertisement space with unreliable estimates." *ACM Conference on Electronic Commerce*. 2007.

**Maher, R.**, “Display CPMs still down big in Q4 But Sequential Growth Encouraging,” *Business Insider*, -2010.

**Marthews, A. and C. Tucker**, “Government Surveillance and Internet Search,” Working Paper, MIT, 2014.

**Mayer-Schonberger, Viktor, and Kenneth Cukier**, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Dolan/Mariner Books, 2014.

**Meek, C., M. Chickering, and D. Wilson**. “Stochastic and Contingent Payment Auctions.” *ACM Conference on Electronic Commerce*. 2005.

**Miller, A. and C. Tucker**, “Privacy Protection, Personalized Medicine and Genetic Testing,” Working Paper, MIT, 2014.

**Rysman, M.** 2009. “The Economics of Two-Sided Markets”, *Journal of Economic Perspectives*, 23 (3), 125-143:

**Smith, Nicola**, 2013. “Consumers require value in exchange for data,” Marketing Week.

**Smith, Cooper**, 2014. “Facebook Has a Program That Gives Special Access To Elite Marketers — These Are Their Insights,” *Business Insider*.

**Stutz, R.** 2011. “An Examination of the Antitrust Issues Posed by Google’s Acquisition of ITA,” *The American Antitrust Institute*.

**Taylor, G.** “Search Quality and Revenue Cannibalisation by Competing Search Engines.” *Journal of Economics and Management Strategy*, 2012.

**Timberg, C.**, 2013. “Could Google tilt a close election?” *Washington Post*.

**Tsai, J., S. Egelman, L. Cranor, A. Acquisti**, “The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study,” *Information Systems Research*, 22, 254-268, 2011.

**U.S. Department of Justice Press Release**, November 5, 2008, "Yahoo! Inc. and Google Inc. Abandon Their Advertising Agreement."

**U.S. Department of Justice Press Release**, February 18, 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."

**U.S. Department of Justice Press Release**, April 8, 2011, "Justice Department Requires Google Inc. To Develop And License Travel Software In Order To Proceed With Its Acquisition Of Ita Software Inc."

**Varian, H.** "The Economics of Internet Search," *Rivista Di Politica Economica*, November-December 2006, pp. 177-191.

**Varian, H.** "Position auctions." *International Journal of Industrial Organization* 25, no. 6 (2007): 1163-1178.

**White, A.** "Search Engines: Left Side Quality Versus Right Side Profits." Working Paper, 2008.





## Susan Athey

Susan Athey is Professor of Economics at Stanford Graduate School of Business and chief economist at Microsoft Corporation. Professor Athey's current research focuses on the economics of the internet, internet search, online advertising and the news media, and virtual currencies. She advises governments and businesses on the design of auction-based marketplaces

In 2007, Professor Athey received the John Bates Clark Medal, awarded by the American Economic Association to "that American economist under the age of forty who is adjudged to have made the most significant contribution to economic thought and knowledge."

She was elected to the National Academy of Science in 2012 and to the American Academy of Arts and Sciences in 2008. She received her bachelor's degree from Duke University and her Ph.D. from Stanford, and she holds an honorary doctorate from Duke University. She previously taught at the economics departments at MIT, Stanford and Harvard.