

BUSINESS MODELS AND BIG DATA: HOW GOOGLE USES YOUR PERSONAL INFORMATION

Marcela Mattiuzzo¹

Introduction

In March 2016, the German competition authority (Bundeskartellamt) released a statement revealing it had initiated proceedings against Facebook under suspicion that the company was infringing data protection regulation to abuse its dominant position in social networks. The president of the German authority, Andreas Mundt, claimed that “[f]or advertising-financed internet services such as Facebook, user data are hugely important. For this reason it is essential to also examine under the aspect of abuse of market power whether the consumers are sufficiently informed about the type and extent of data collected².²”

Digital businesses have been under scrutiny for some time, and antitrust authorities are not the only ones concerned with their handling of users' privacy. The field of data protection gained such prominence that an agreement between the United States and the European Union was crafted solely to address transatlantic data transfer. The Safe Harbor Framework³, as it was known, was later struck down by the European Court of Justice (ECJ)⁴, creating a “data crisis” that led to the approval of the new Privacy Shield agreement⁵.

¹ Visiting Researcher at Yale Law School (2016-2017) and Master of Laws Candidate at the University of São Paulo. Partner at VMCA Advogados. Former Chief of Staff and Advisor for the President at the Brazilian Administrative Council for Economic Defense (2015-2016).

² Bundeskartellamt Press Release - Bundeskartellamt initiates proceeding against Facebook on suspicion of having abused its market power by infringing data protection rules. Available at: <http://www.bundeskartellamt.de/SharedDocs/Meldung/EN/Pressemittelungen/2016/02_03_2016_Facebook.html> Access 19 September 2016.

³ The Safe Harbor Framework between the United States and the European Union was a mechanism by which organizations could transfer data across the Atlantic. It was established for the EU, concerned with US handling of personal information, wished to better protect its citizens' data. The Framework operated owing to seven principles: notice, choice, onward transfer, access, security, data integrity, and enforcement. On October 6, 2015, the European Court of Justice issued a judgment declaring the framework inadequate.

⁴ PADOVA, Yann. The Safe Harbour is invalid: what tools remain for data transfers and what comes next? **International Data Privacy Law**, Vol. 0, N. 0.

⁵ Prompted by revelations of mass surveillance by the American government, the new framework aims to protect “the fundamental rights of anyone in the EU whose personal data is transferred to the United States as well as bringing legal clarity for businesses relying on transatlantic data transfers.” It was drafted owing to the requirements set forth by the ECJ in its October 2015 ruling. Available at: <http://europa.eu/rapid/press-release_IP-16-2461_en.htm>. Access 3 November 2016.

Simultaneously, several jurisdictions have created agencies and authorities whose goal is to deal with privacy regulation. The most noteworthy example, but by no means only, is the European Union. Since 1995, the European Commission (EC) established the region-wide Directive 95/46/EC, which puts forward parameters for the processing and circulation of personal data⁶. The Directive also created the Working Party on the Protection of Individuals with regard to the Processing of Personal Data, known as the Article 29 Working Party. In 2012, the EC proposed a comprehensive review of personal data regulation by means of Regulation 2016/679 (the General Data Protection Regulation – GDPR), reinforcing the role of data protection⁷.

Owing to this body of legislation, several Member States have developed national authorities to enforce European-wide rules⁸. The GDPR, which shall come into force in 2018, will establish a new framework for cooperation between such national bodies⁹.

Concern over privacy regulation in the online world is therefore not a novelty, nor have governments ignored its growing impact in the past decades. But because technologies evolve at a pace with which legislation cannot keep up, there has been much debate on whether the current frameworks adequately protect users' privacy and whether changes are necessary – even if those changes would come at the expense of economic efficiency.

Such concerns also gained prominence due to the process of economic concentration and consolidation in the online environment. Today, a small number of large companies control much of the world's personal information, and the bulk of information grows at an exponential

⁶ The Directive is not immediately binding on companies, but rather on Member States, who had to instate legislation and authorities to oversee the flow and handling of personal data. The three principles of the Directive are transparency, legitimate purpose, and proportionality.

⁷ The European Union allows for issuance of both directives and regulations, and these instruments serve different purposes. While a directive has to be transposed into national law by means of other legal instruments – and the Member State is thus free to establish any framework within the general principles of the directive to enforce it – a regulation is immediately enforceable in all of the EU. That is why the vacatio for regulations is usually longer, since several national authorities must simultaneously adapt to the new rules. The GDPR is no exception to this tendency; approved in 2016, it will come into force in 2018.

⁸ The most prominent of which is the Irish Data Protection Agency. The growth of Irish regulation over privacy is a natural consequence of the geographical distribution of many Silicon Valley companies within the European Union. When expanding across the Atlantic, many of them chose Ireland as their principal non-US headquarters (mostly due to tax exemptions).

⁹ In the United States, the Federal Trade Commission is one of the agency's responsible for privacy regulation. It should be noted, however, that American regulation on the topic is not centralized. The protection extended varies according to the type of transaction, which is why some authors usually refer to the American regulation of privacy as "limited". For more on the regulatory models put forward by the European Union and the United States, see Guilherme Guidi's article.

rate. The size and relevance of such companies shifted their role from objects of regulation to architects of some markets, in a movement some authors refer to as “code is law”¹⁰. The algorithms that shape these online platforms have effectively become the norm for much of the web.

Personal data has been an asset for business for many years. It arguably has been so ever since the concept of advertising was invented, as businessmen have long concluded they have higher success-rates when they target ads according to their audience¹¹. The very definition of personal data helps clarify this point. The Irish Data Protection Commissioner¹², whose Data Protection Act dates back to 1988 – a time before commercial internet existed – defines personal data as “data relating to a living individual who is or can be identified either from the data or from the data in conjunction with other information that is in, or is likely to come into, the possession of the data controller”¹³.

Why then would this topic be particularly relevant for online markets? The answer lies in the fact that the expansion of the internet catapulted data to a much more significant position, creating what is now referred to as Big Data. Big Data is a concept originally coined by Doug Laney¹⁴ to describe the new ways by which data is collected, profiled and utilized by businesses in the internet age. He claimed there are three dimensions to that concept, which were later expanded to the ‘Four Vs’ of Big Data: volume, variety, veracity, and velocity¹⁵.

¹⁰ Lawrence Lessig, Code, Version 2.0, Chapter 1 - Code is Law. Basic Books, New York, 2006.

¹¹ Privacy and Big Data, Terence Craig; Mary E. Ludloff, 2011, Advertising as the Big Bad Wolf: “In the pre-digital days, there were companies that specialized in analyzing buying behavior, like AC Nielsen, and companies that “rented” out their customer list, segmented by income level, sex, marital status, buying behavior, etc. Chances are your mailbox, like ours, was stuffed with all kinds of offers and you seemed to get phone calls about buying or selling something every hour. Most likely, those offers were the result of information you gave to your bank, credit card company, grocery store, or as a magazine subscription holder. But the information was, to some extent, blind. Your name and address were rented, usually as part of a group, but the renter (the business or organization that bought the advertising) did not have that information until, and unless, you responded. If you did, you then became a part of that company’s mailing list and they would begin to build their own profile about you. So, even then, there were multiple profiles of you in multiple lead or customer databases based on your behavior with a specific company or organization.”

¹² Irish personal data regulation has become particularly relevant over the past years as many Silicon Valley companies established their European subsidiaries in the country – the original reason was mostly related to tax benefits, but it inadvertently rendered Ireland a center of personal data discussions and its law particularly important within the European context.

¹³ As per the Data Protection Act of 1988, revised in 2016.

¹⁴ Available at: <<http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>> Access 2 November 2016.

¹⁵ Some authors talk of five Vs instead of four, including variability (or sometimes value) among the characteristics. I will only consider four Vs as the relevant aspects raised by variability and value are, in my view, already

Data is considered to be “big” because it is produced and collected in impressive **volume**. IBM estimates that the world generates around 2.5 quintillion bytes a day, and commentators point out that most of the current volume of data was created in the last two years¹⁶. Also, data is of an unforeseen **variety**. Businesses have access to basic information about a person such as name, age, and gender, but also to detailed characteristics, running from health condition to daily and hourly location available via GPS.

The **veracity** of data refers to the trustworthiness of the information gathered. That is not to say data is necessarily false, it simply means to convey it can be misleading, either because it is outdated, or because it implies biases. Lastly, data is big when it can be gathered and processed in a speedy fashion. The **velocity** of data analysis is growing, which is essential in enabling its use for business purposes.

This conceptualization of big data is popular, but not uncontroversial. Commentators have argued that there is no real threshold with which to measure data in order to verify when it is “big”¹⁷. True as it may be, the idea of Big Data has gained ground. One of the factors that stimulated its popularization and growth was the massification of smartphones. There were around 7.4 billion mobile subscriptions in the first quarter of 2016, 63 million of which were new subscriptions, and the numbers are expected to reach 9 billion by 2021¹⁸. Smartphones are devices that carry people’s lives around with them. They store an astounding amount of data, including the pictures one took, the contacts that person shares, the places she visited, her emails, SMS messages, the videos she looked up, the apps she downloaded. That information, which was complex and time-consuming to gather, is now available at the reach of one’s hand in smartphones – and for businesses to use in the cloud¹⁹.

encompassed in this definition.

¹⁶ STUCKE, Maurice E. & GRUNES, Allen P. Big Data and Competition Policy. **Oxford University Press**, 2016.

¹⁷ “It is not a precise scientific concept, but a highly contested idea that means different things depending on who is talking about it. There is, and will never be, any consensus on what “big data” means, nor on how its processing differs from the data analytical techniques of the past. There is no clear threshold at which point “data” becomes “big data.” It is a highly fashionable, and therefore inherently suspect, idea that encompasses a complex array of technologies, practices and interests. “Big data” in and of itself means nothing, and signifies nothing, in the absence of a wider understanding of the organizations that are conducting the analysis, and an assessment of those organizations wider interests and motives.” Colin Bennett, **Privacy Protection in the Era of “Big Data”**: Response to Office of Privacy Commissioner’s Discussion Paper on “Consent and Privacy”, p. 2.

¹⁸ **Ericson Mobility Report**. Available at: <<https://www.ericsson.com/res/docs/2016/ericsson-mobility-report-2016.pdf>> Access 5 November 2016.

¹⁹ STUCKE, Maurice E. & GRUNES, Allen P. Big Data and Competition Policy. **Oxford University Press**, 2016. P. 29 – Smartphones as an example of how big data and privacy intersect.

With such an array of data available, it is to be expected that within the realm of Big Data, the variety of business is considerable. This article will concentrate in one specific type of online business that can be referred to as online advertising platforms (OAPs). By that expression I mean every company that carries its business online and relies on advertising to earn the biggest bulk of its profits²⁰. The advertising promoted by OAPs is of a particular nature because it makes use of personal data in an effective and intensive way.

The amount of data available to OAPs due to development of Big Data is tremendously vaster than the amount usually available to other businesses, even those that identified their strategy as “targeted advertising” long before the internet was economically viable²¹. OAPs monetize personal data as their main source of revenue, profiling users and selling targeted advertising as a result.

It comes as no surprise that after reaching such extraordinary levels, the usage of personal data, although profitable and often efficiency-enhancing, has raised privacy concerns. Today’s targeted advertising depends on zettabytes of data collected by OAPs, including everything from a user’s friend list and the place where she lives, to her social security number and most liked books. If accessing such wide array of information allows for terrifically accurate search results, it also brings about some concerns.

Much like in Franz Kafka’s “The Trial”, the handling of personal data by private-owned enterprises is to a great extent nothing short of a black box. It is carried out by a largely inaccessible and somewhat non-accountable authority, which handles information in a rather obscure way, claimed to be beneficial to users, but never revealed to the public. Its gatekeepers are Silicon Valley billionaires and their team of experts.

Even if we assume these people to have the best of intentions, and trust their ability to forego financial interests for “the greater good” – considering they already have enough money for generations of Gates and Zuckerbergs – there still is strong disagreement about what that greater good should be. How much information should companies be allowed to have, or

²⁰ Therefore, companies such as Uber and Amazon, which are usually associated with the expansion of Big Data, are excluded from this definition.

²¹ One example is the broadcast television market. Esther Gal-Or, Mordechai Gal-Or, Jerrold H. May, William E. Spangler, (2006) Targeted Advertising Strategies on Television. *Management Science* 52(5):713-725. It should be noted that some offline businesses do have access to immense amount of data, the primary example being the credit-card industry, but usually the data collected by such companies is protected by contractual clauses that to some extent prevent it from being used for advertising.

differently put, how much should users be allowed to give away? Should we protect users' privacy even when those users would gladly give their data away in exchange for free services? Are users even aware of precisely what they are giving away, and to what will be the commercial use of such data? Should we somewhat limit the way data is transferred among companies?

These and many other questions remain unanswered. It is not the goal of this article to present final answers for any of them, but it is my goal to provide some insight into how precisely one OAP has made use of personal data to monetize its business, and with that hopefully shed some light on how privacy considerations are to be inserted into future policy undertakings that aim to regulate the online advertising industry.

The Google Case

Back when Larry Page and Sergei Brin were Ph.D. students at Stanford University, they designed Google to be a revolutionary search engine. In Page's own words, their goal was, and still is, "to organize the world's information and to make it universally accessible and useful". Search remains the most famous of Google's interfaces (a proof of that is the mere fact that the word "google" is now a synonym for search, officially included in the Oxford English Dictionary as such). But the company grew tremendously and expanded to new areas, which prompted its creators to reorganize it in the form of a holding in the year 2015. This holding is called Alphabet and its business include everything from e-mail servers to mapping the Earth. Nevertheless, the bulk of Alphabet's revenue still comes from advertising. Out of the more than 90 billion dollars the company generated in 2016, approximately 87% came from the platform's ad services²².

The jump from being a search engine to becoming an advertising platform was neither automatic nor uncontroversial. Page and Brin developed Google in order to offer a good product, but they had no predetermined business model that would allow the company to generate income. At one point, venture capitalists realized the company's potential and funded its initial undertakings, but investors soon started to wonder how the two former Stanford Ph.D.s would be able to turn a profit. Ideas on how to make money ran from licensing the algorithm and selling it

²² Data for 2015 is available at: <<http://www.investopedia.com/articles/investing/020515/business-google.asp>> Access 8 September 2016. Alphabet's Fiscal Annual Result for 2016 show that advertising is the primary source of revenue for the company: <https://abc.xyz/investor/pdf/20161231_alphabet_10K.pdf> Access 5 March 2017.

to other internet businesses to going down the road already traveled by Yahoo! and placing advertising on Google's page, just as any other website would do. The problem for Google was the lack of stickiness of search – back then online advertising was sold based on a webpage's ability to retain a given user for a long period of time, since ads were in display format and aimed at calling users' attention. Search, and particularly Google's search, was the exact opposite of that, the company offered users a fast result for a query and directed them to other webpages, which seemed to dismantle the advertising endeavor from the get-go.

It was largely the founders' obsession with the cleanliness of Google's search results, and also their distaste for annoying ads, that lead the company to a different path: the creation of AdWords. One of the key minds responsible for coming up with the functioning of AdWords was Eric Veach. Salar Kamangar, Google employee #9, was also a key participant in the process. Kamangar focused more heavily on the business model, while Veach was the mathematician behind the project²³. The distinctive feature of AdWords is its ability to generate revenue while also allowing for results connected to the search query – and later to the page's content. It moves away from the display ad model towards a textual approach, and it has been laying Google's golden eggs for over a decade.

AdWord

AdWords is Google's platform for advertisers. It is responsible for offering ad space in exchange for payments and it functions as a fairly elaborated auction.

It should never be underemphasized that the specific type of auction developed by Google for AdWords has been so immensely successful that commentators have gone so far as to name the underlying process the “Googlenomics”²⁴. The development of this auction process has taken years to perfect and involved dozens of people²⁵. It goes way beyond the scope of my analysis to fully scrutinize its functioning. Nonetheless, my goal is to give a fairly comprehensive outline of AdWords, focusing on how it depends on users' personal data in order

²³ A brief overview of the story is available at: <<http://www.bloomberg.com/news/articles/2006-03-05/the-secret-to-googles-success>> Access 14 October 2016.

²⁴ LEVY, S. The Secret of Googlenomics: Data-Fueled Recipe Brews Profitability. **The Wired**. September 22, 2009.

²⁵ A complete account of the building and functioning of AdWords is provided by Stephen Levy in his book **In the Plex: How Google Thinks, Works, and Shapes Our Lives**, 2011.

to properly function²⁶.

What advertisers pay for at AdWords is ad space. Such space is distributed so that ads may be placed: (i) in a search results page (which is what happens when a user types a search query into Google Search), (ii) in a page pertaining to Google's search partner network, a network of websites that incorporates Google Search onto their webpages, or (iii) in any random website which chooses to be a part of the Display Network, a system that pays the website for showing AdWords results²⁷.

Since the system functions as an auction, there is competition among advertisers to determine how the ads will be selected. Such choice will take place owing to several factors, but it is largely related to the keywords typed by the user – in case of searches – and to the content of the webpage – in case of the Display Network (and of several Google-owned websites such as Gmail and YouTube), as well as to the specific kind of user those advertisers aim at.

AdWords was not always as it is today. It started out as a two-tiered system, composed of the Premium and Select versions. As Steven Levy puts it:

Google's ads were always plain blocks of text relevant to the search query. But at first, there were two kinds. Ads at the top of the page were sold the old-fashioned way, by a crew of human beings headquartered largely in New York City. Salespeople wooed big customers over dinner, explaining what keywords meant and what the prices were. Advertisers were then billed by the number of user views, or impressions, regardless of whether anyone clicked on the ad. Down the right side were other ads that smaller businesses could buy directly online. The first of these, for live mail-order lobsters, was sold in 2000, just minutes after Google deployed a link reading see your ad here²⁸.

With time, Google shifted strategy and decided to expand the auction system to encompass AdWords Select. The auction was initially only available for search results. The reason is straightforward: search results are an “easy” way to effectively select your audience and they do not require much from the platform. The user types keywords, the platform sells ad space based on those keywords, and that is the end of the story²⁹. Google’s job is to capture that

²⁶ It should also be said I will build my analysis based primarily on materials provided by Google itself, to avoid second-hand reading as much as possible. Google has a platform where it provides users (and specially advertisers) information on AdWords: <www.support.google.com/adwords> Access 5 March 2017. Whenever I use quotation marks and do not specify where they came from, the source is this platform.

²⁷ That is called AdSense. The functioning of AdSense will be described in a separate section.

²⁸ LEVY, S. The Secret of Googlenomics: Data-Fueled Recipe Brews Profitability. **The Wired**. September 22, 2009.

²⁹ Google was not the first to figure this out. GoTo, later renamed Overture, had done so before, by introducing pay

information already provided by the users and monetize it. The idea sounds terribly simple, but the ability to effectively match users and keywords is one of the reasons why Google was the undisputed leader in online advertising for quite some time.

In its efforts to gain more ground, Google introduced the Display Network (DN) as a part of AdWords. In the company's own words, “[t]he Adwords Search Network reaches people when they're already searching for specific goods or services. The Display Network helps you capture someone's attention earlier in the buying cycle. For example, if you run an art supply store, you can catch a mom's eye when she's reading reviews about the best brands of washable paints, but before she puts her toddler in the car seat and heads out to buy.”

Leaving aside the claim that people in the search network are already searching for goods and services, with which I disagree for reasons exposed elsewhere³⁰, it is worth turning to why Google decided to expand its business in such a way. The expansion was only possible because Google dramatically increased the amount of data collected from users and the places within the internet where it collects such data from.

The DN does not offer traditional brand advertising³¹, it goes much further than that. It matches users' online behavior to advertising opportunities outside of the typical search result page. For that reason, the DN functions rather differently than the traditional Search Network and they will be described separately.

The Search Network

Advertisers on the Search Network buy ad space from Google choosing among keywords

per click and auction advertisement. But GoTo's CEO, Bill Gross, never patented any of those inventions. Moreover, the company's business model relied on ranking ads without taking quality into consideration, mixing organic search results and advertisement, and requiring bidders to pay the price they submitted in the auction, incentivizing bid shading. See footnote 21 for more on optimal bidding strategies.

³⁰ MATTIUZZO, M. **Propaganda Online e Privacidade – o varejo de dados pessoais na perspectiva antitruste.** SEAE 2014. See also GOLDFARB, A. & TUCKER, C. **Substitution Between Offline and Online Advertising Markets,** 2011.

³¹ Traditional brand advertising involves raising brand awareness. It entails bombarding potential consumers with information about the brand, in several different situations, not necessarily connected to a specific selling opportunity. A brand that heavily and consistently invests on brand advertising is Coca-Cola. Coca's strategy of placing billboards in highways, full-page ads in the newspaper, etc., is not intended to make the potential consumer immediately stop what she is doing in order to purchase the product, but to make the consumer aware of the product, so that when she is faced with the decision of buying a beverage, she chooses a Coca-Cola instead of something else. Brand advertising can focus on several strategies by associating a brand with a number of different scenarios or moods. For my purposes here, what should be noted is that it is different from what Google offers to most of its advertisers.

they believe to have some relevance for the substance of their ad. For example, a shoe store will probably select keywords such as “tennis shoes” and “sneakers” for its ad, whereas a beer retailer might prefer words such as “bbq drinks” and “cheap beer”.

Among the advertisers that select the same keywords, Google determines who gets to be ranked higher and more prominently in its pages using Ad Rank. According to Google, “Your Ad Rank is a score that's based on your bid, auction-time measurements of expected CTR, ad relevance, landing page experience, and the expected impact of extensions and other ad formats.” Let me break down each of these variables.

The bid price is the maximum price an advertiser is willing to pay for the ad. Google offers two different bid strategies for the Search Network: CPC and CPA³². The Cost-Per-Click (CPC) strategy is the one by which advertisers pay only when their ads are effectively clicked on by the users. Under this strategy, each advertiser determines a “max CPC” (or maximum Cost-Per-Click), which is the maximum price it is willing to spend on an ad. The max CPC is not necessarily the amount the advertiser will effectively pay, since the final value will depend on other potential buyers’ bids. The model is a variation of the second-price sealed-bid auction, also known as Vickery auction: the price paid is not the winner’s bid price, but however much is necessary for him to maintain his position as the winner. If the only consideration to be taken into account were the bid price, then Google would charge the winner the second-best price. (If User P bids \$10, User L bids \$6, and User M bids \$2, User P will win and pay \$6.)³³ This is however *not* the real price paid by the advertiser, as the final price is calculated depending on the final Ad Rank, which will be further explained below.

The Cost-Per-Acquisition (CPA) is a strategy focused on conversions. A conversion is an

³² Google also allows for Cost-Per-Thousand Viable Impressions, which will be covered in the next section.

³³ Google certainly designed this strategy based on auction theory. The company soon realized it makes more money by adopting a Vickery-type auction than by charging the max CPC from the winner (which would represent a first-price sealed auction). The strategy seems senseless at first glance, as the company would make more money if it simply charged User P in my example the \$10 it is willing to pay. But by charging the second-best price the company prevents advertisers from bidding too low (or bid shading) and bringing overall prices down. Essentially, the Vickery model creates an equilibrium in which the optimal strategy is to “tell the truth” and provide each bidder’s real valuation for the good being auctioned. An overview of auction theory that explains these equilibria in detail is KLEMPERER, Auction Theory – A Guide to the Literature. A study focused on Google’s practices is provided by VARIAN, Position Auction.

It should be noted, however, that some scholars have raised questions as to precisely how adherent to Vickery auctions Google’s practices truly are. Edelman, Ostrovsky and Schwarz argue that the Search Network runs a “generalized second-price auction”, in which equilibrium is not reached by bidding true valuation. Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords.

action other than a click made by a user after viewing an ad. It can be anything from the subscription to a mailing list to a purchase. Unlike in CPC, Google does not charge users for each conversion, it rather finds what the AdWords algorithm believes to be an optimal CPC bid whenever the ad is eligible to appear. The advertiser chooses the “target CPA” and Google’s job will be to get as many conversions as possible with that given amount. To do so, Google needs two things: first, historical information about the ad. Second, a way to track conversions.

CPA is significant for my purposes because in order to track conversions an advertiser needs a “tag” on her website. It is “[t]his tag [that] will place a **cookie** on a user's computer or mobile phone when he or she clicks your ad. If the user reaches one of your conversion pages, AdWords looks for the cookie and records a successful conversion for you.” In other words, Google is no longer relying on data that a user directly entered into the search tool, it is rather following the user around the web to see how she behaves after exposure to an ad. That way it can provide a more useful and complete targeting strategy for advertisers³⁴.

The quality of the ad is the bundle of measurements used by Google to determine the relevance of an ad to the user. It encompasses:

1. Expected clickthrough rate	The likelihood that an ad will be clicked
2. Ad relevance	How closely the ad matches the intent behind a user's search
3. Landing page experience	How relevant, transparent and easy-to-navigate the page is for users

The Click-Through Rate (CTR) “measures how likely it is that your ads will get clicked when shown for that keyword, irrespective of your ad's position, extensions, and other ad formats that may affect the visibility of your ads”. In other words, if users click more on Ad A than on Ad B, Ad A will have a higher CTR.

The relevance of an ad to users is a tricky concept, as there is much debate about the usefulness of ads, and also considering the subjectivity of usefulness. But in Google's own words “[t]his status describes how well your keyword matches the message in your ads. For example, if

³⁴ I will refrain from any judgement on the practices carried out by Google, and I urge readers to do the same. My goal now is solely to explain, as best as I can, what the platform does in order to make a profit. A reader should not understand this statement, or any other in this part of the text, as revealing of my approval/disapproval of the company's strategy.

someone searches for your keyword and your ad shows up, would your ad seem directly relevant to their search?”.

Relevance, other than being measured in absolute terms, can also be tailored to specific audiences or individuals. That is to say that Google built into the algorithm a feature that allows ads to have more relevance when they are believed to be particularly suited to a person, given this person’s profile. Google’s privacy policy explains that the platform may show a user who is searching for “vacation” and earlier searched for “bike” ads related to biking while on vacation, which will probably not happen to those who search for “vacation” and had earlier looked for “movie theaters”³⁵. This is but one example of how data can be used to personalize the ad experience, but there are many others. Google can pull data not only from Search, but also from YouTube, Gmail, and partner sites to personalize ad experience. Since changes in the use of personal data for advertising have recently been introduced by the company, they will be explored in more detail in item “My Activity” below.

The landing page is the page where the user is directed to once she clicks the ad. When measuring this variable Google is interested in the experience the page provides to users and in its relevance given the search terms provided. Google’s Chief Economist Hal Varian states that a good landing page is relevant, provides original content, is easy to navigate and transparent.

The expected impact of extensions/formats refers to additional information placed in an ad, such as phone numbers, e-mail, address, etc., which can somewhat reveal more about the business and as such increase the final score in the auction.

The higher a given ad ranks according to these variables, the better it will be positioned in a search result page. Google does not disclose the specific weight of these variables and uses a blind system to determine final positioning – it does not reveal advertisers how much others are bidding, nor other advertisers’ quality scores or extensions/formats. However, it does disclose the advertiser’s Ad Rank – the weighing of bid price, quality, and ad extensions/formats. As mentioned, the final price paid by the advertiser will be however much is necessary for her to maintain her position as the first in line. See Table 1 below for an example provided by Google³⁶:

³⁵ Available at: <https://privacy.google.com/how-ads-work.html?modal_active=how-ads-work-proof-overlay&article_id=c4-p-search-ads-1> Access 27 October 2016.

³⁶ Insights on the AdWords Auction. Available at: <<https://www.youtube.com/watch?v=PjOHTFRaBWA>> Access 24 October 2016.

Table 1

<i>Bidder</i>	<i>Bid</i>	<i>Quality</i>	<i>Format</i>	<i>Ad</i>
				<i>Impact</i>
				<i>Rank</i>
<i>Blue</i>	\$4	Low	N/A	5
<i>Green</i>	\$3	High	Low	15
<i>Yellow</i>	\$2	High	High	20
<i>Red</i>	\$1	Medium	Medium	8

What determines the order in which ads will appear is the result in the last column, Ad Rank. This means that, in this example, Yellow will be placed first, followed by Green and Red – assuming there are three positions available. Blue, despite having the highest bid, will not be displayed. Moreover, Yellow will not necessarily pay \$3 – which is the second-highest bid, it will pay however much it needs in order to maintain its Ad Rank above 15. The same holds true for Green, which needs only to keep its Ad Rank above 8, and for Red, which should maintain Ad Rank above 5.

There is no way of knowing exactly how much these bids will be. Experience has shown that Yellow will likely pay less than its bid price, but only the algorithm can tell the exact numbers. Despite the lack of full transparency, it is undisputable that Google tries to rank its ads owing to relevance, not simply based on how much advertisers are willing to pay for it. That indicates the company is effectively interested in showing users more useful ads – which is not only a symptom of benevolence, rather a consequence of the payment model, as Google's profit to some extent depends on whether the ads are clicked on by users. Nevertheless, this indicates strong concern with how annoyed users become with the ads shown by the platform. In the words of Anastasia Holdren when describing AdWords and its functioning:

AdWords works because it doesn't seem like advertising. To Google's credit, the displayed advertising results are extremely relevant to the searcher's query. Ads are displayed at the moment someone is looking for something and presented as potential solutions to their search. This relevancy is the key to the effectiveness of the system for both searchers and advertisers³⁷.

As mentioned, the concept of relevance in advertising is contestable, but the statement by

³⁷ HOLDREN, A. Google AdWords. O'Reilly Media, Inc. 2011. Introduction to Google AdWords.

Holdren is revealing. She stresses how AdWords is a revolution of sorts because it turns ads into “solutions” to problems or requests by users³⁸. This understanding of “ads as solutions” is entirely dependent on the OAP’s reliance on personal data, and its ability to match users and advertisers.

What should be added is that Google’s organic search results – the results that are not determined by the bid process, rather by a different Google algorithm intended to provide the best result for the search query – cannot be bought by advertisers. The algorithm, and the algorithm alone, determines the ranks of pages, and though some companies heavily invest on moving their pages up on the organic search, they do so without paying Google off³⁹.

Display Network and AdSense

The Display Network and AdSense are essentially two sides of the same coin. The DN is the advertiser platform, whereas AdSense is the website-owner interface – or as Google prefers to call it, the publisher. Both are brought together – and combined with DoubleClick Ad Exchange – to create a market for online advertising⁴⁰. Any website that offers advertising space can be a part of AdSense free of charge. On the other hand, to be able to place ads at the DN advertisers use a platform similar to the AdWords Search Network and must pay for Google’s services.

The two main differences between the Search Network and the Display Network regard the format in which ads are presented and the way those ads are selected by the advertiser and by

³⁸ She is not alone in that statement, see JACOBSON, H. Google AdWords for Dummies. Wiley, 2009. How to Think Like a Prospect. “In the Magic Yellow Pages, you don’t have to flip through hundreds of pages. In fact, the book doesn’t *have* any pages — just a blank cover. You write down what you’re looking for on the cover, and then — Poof! — the listings appear. The most relevant listings, according to the Magic Yellow Pages, appear on the cover. Subsequent pages contain more listings, in order of decreasing relevance.”

³⁹ There are antitrust investigations in several jurisdictions that try to prove this is not entirely accurate and that Google may have altered the algorithm to privilege some websites in organic search. The allegations, however, focus on Google’s alleged attempt to leverage the dominant position in search onto other markets. There is no allegation that Google somehow accepts money for better positioning.

⁴⁰ To be perfectly accurate, DN is wider than AdSense. It also covers DoubleClick Ad Exchange and the Google websites. Ad Exchange is a different platform, also owned by Alphabet, that has a separate bidding process and different functionalities. The reason why I will not get into details about Ad Exchange is simple: any advertiser who chooses the DN to display its ads automatically has access to all Ad Exchange websites that comply with AdWords’s guidelines. Moreover, Google controls the auction process so that the comparison between Ad Exchange bids and AdWords bids is normalized. To understand more about Ad Exchange and how it differs from AdSense: <<https://support.google.com/adxseller/answer/4599464?hl=en>>. Access 5 March 2017.

the platform. Display allows for ad formats other than text, including images, videos, and rich text, whereas Search only accepts text. More importantly, ads for the DN are selected based not only on keywords, but also on topics, audiences, and placement.

Google offers three methods for targeting ads in the DN: (i) contextual targeting – focused on the traditional keywords and on topics (which are the themes a given webpage is related to, such as sports, fashion, electronic equipment, etc.); (ii) audiences – focused on the users she intends to reach, the advertiser may select according to interest categories, remarketing and demographics (meaning the advertiser can either choose an user specifically interested in the content of her ad, an user who has already visited her website before, or a user part of a gender or age group); and (iii) placement – which allows the advertiser to either exclude or target specific websites and apps.

Other than being selected by the advertiser owing to different criteria, the platform selects ads to be shown at the DN in a different way. The overall parameters are the same, meaning an ad will appear depending on the calculation of the Ad Rank, but the way in which bids can be cast differs, and so does the effective price advertisers pay for the placement.

First, regarding the bid price. Other than the traditional CPC and CPA, advertisers may also choose the Cost-Per-Thousand-Viewable-Impressions (vCPM). vCPM is a strategy focused on impressions, as the name suggests. From the options provided by Google, this is the one that most closely resembles brand awareness. The advertiser will not pay for the ad when a user clicks on it, rather when the user sees it in a webpage^{41, 42}.

Second, with reference to payment, Google introduces some variations that aim at making the results fairer, since payment on the DN depends heavily on incremental clicks⁴³.

To understand incremental clicks, one must remember Google will not always face a dilemma of either showing the ad or not showing it. The platform has several placements available and it can decide to show both ads, but also to place them so that the winner is more visible.

⁴¹ Google has a way of measuring views, called the Active View. The definition of a “viewable” ad is the following: “An ad is counted as “viewable” when 50 percent of your ad shows on screen for one second or longer for display ads, and two seconds or longer for video ads.”

⁴² The competition between CPC and vCPM bids is corrected by Google to avoid comparing apples and oranges. The platform does so by calculating an expected click-rate for every 1,000 impressions, instead of the usual clickthrough rate.

⁴³ It can also depend on service fees for audience targeting, if that is the chosen method.

Because different placements can render different amounts of clicks, Google does not charge the same price for both advertisers when it shows both ads, it charges the winner more based on the incremental clicks it will enjoy from being better positioned in a page.

Imagine two ads, β and μ . Both ads rank high enough so that the platform decides to show them both. β bids \$10 and μ bids \$5, both are of equal quality, making β the overall winner who will enjoy better positioning. How should Google charge β ?

What the company does is divide payment according to the so-called incremental clicks. Because of the more favorable positioning, β is clicked on ten times for every eight clicks for μ , meaning there are two incremental clicks for β that should reflect the price paid by the advertiser. Assuming the result rested solely on bid prices⁴⁴, β would pay \$5 for those two clicks and the same price as μ for the remaining eight. Imagining the third place on the auction went to Ω , who bid \$4 and was not selected to be shown, β would pay \$4 for the remainder of the clicks.

In this example, considering a total of ten clicks, β would pay $(\$5 \times 2) + (\$4 \times 8) = \$42$.

To highlight the difference between the DN and the Search Network, one can imagine what the final price for β would be if the ad was placed in search. If there were two available placements and the bids followed the exact same pattern, β would win and pay \$5 for each click. μ would also be shown, though in the less visible position, and pay \$4 per click. There would however be no comparison between incremental clicks for β and μ , meaning the price would remain constant and depend solely on clicks or conversions. β would then pay $\$5 \times 10 = \50 for every ten clicks/conversions⁴⁵.

As mentioned, the DN relies not only on Google's own platforms, such as YouTube and Gmail, it also includes third-party websites. Those websites are part of AdSense, a vast network reportedly encompassing 90% of the Internet⁴⁶ and governed by a set of policies put forward by the platform^{47,48}. From those policies, the one that is particularly relevant for my purposes

⁴⁴ Again, it does not. The final Ad Rank encompasses ad quality and extensions/formats. Nevertheless, simplifying the analysis in this regard is effective and sufficient to clarify the workings of the platform.

⁴⁵ The problem earlier identified by Edelman, Ostrovsky, and Schwarz apparently does not exist here. Because the Display Network internalizes incremental clicks in its charging mechanism, advertisers' best strategy truly is to bid true valuation.

⁴⁶ Signing up to AdSense is free of charge. The way Google makes money by use of this tool relies on sharing the revenue from the ads placed on a publisher's webpage. The platforms gets 32% and the publisher retains the remaining 68%.

⁴⁷ The DN also includes DoubleClick Ad Exchange. Ad Exchange is a different platform, also owned by Alphabet, that has a separate bidding process and different functionalities. The reason why I will not get into details about Ad Exchange is simple: any advertiser who chooses the DN to display its ads automatically has access to all Ad

regards advertising cookies.

If Google lets advertisers target specific audiences, it must be able to properly profile such audiences, creating well specified groups to attract the advertisers who will extract value from tailored eyeballs. That is why to be part of AdSense a website must abide by a privacy policy that “discloses that third parties may be placing and reading cookies on your users' browsers, or using web beacons to collect information as a result of ad serving on your website⁴⁹. ”

Cookies and beacons are both mechanisms that allow for some form of user identification or tracking. A beacon is an image, usually transparent, placed in a website to track the way such website is navigated. Unlike cookies, beacons do not store any identifiably personal information about a user. A cookie, on the other hand, is a text file placed in a user's browser by a webpage, aimed at tracking traffic through that page⁵⁰. It is assigned to that individual user, meaning the website can “remember” her whenever she comes back, which significantly increases the user experience, helps to improve the site, and is also naturally useful for advertising purposes.

Cookies can be first-party or third-party. First-party cookies are placed on the browser by the website the user is visiting, whereas third-party cookies are placed by other websites – a common use for third-party cookies is precisely advertisement⁵¹. Having a wide set of third-party cookies placed in a range of different websites is what allows Google to create user profiles.

In short, Google requires access to users' behavior on a publisher's website to give access to AdSense. It is worth noting, however, that ever since this system has been in place – and especially after DoubleClick's acquisition in 2007, which significantly broadened the scope of third-party cookies under Google's control – the company has kept the first-party and third-party

Exchange websites that comply with AdWords's guidelines. Moreover, Google controls the auction process so that the comparison between Ad Exchange bids and AdWords bids is fair.

⁴⁸ AdSense program policies include everything from limitations on clicks and impressions – the third-party websites, which Google refers to as publishers, may not click their own ads to inflate impressions or clicks – to copyright law – more specifically, Google reserves the right to pull the page from AdSense if it receives notice that the publisher is in violation of the Digital Millennium Copyright Act.

⁴⁹ As per the AdSense program policies, available at: <<https://support.google.com/adsense/answer/48182?hl=en>>. Access 5 December 2016.

⁵⁰ For more on cookies, see: <<https://www.google.com/policies/technologies/cookies/>> Access 5 December 2016.

⁵¹ As Hoofnagle, Behavioral Advertising: The Offer You Cannot Refuse (2012) point out, “The privacy problem from cookies comes from the aggregation of this tracking across different websites into profiles and through attempts at linking this profile to the user's identity. By tracking these identifiers across websites that users visit, advertisers can infer users' interests, perhaps sensitive ones, such as medical conditions, political opinions, or even sexual fetishes.”

cookies database apart. In other words, though Google had access to a lot of personal information about its users, it never sold advertisement customized to specific individuals, just to specific “groups of data”.

While Google requires from advertisers that they transmit users’ data, it also requests its partners to do so under some conditions. The information provided cannot be accompanied by usernames, passwords, email addresses, and any other personal identifiable information. Failing to comply with this rule is cause for exclusion from AdSense. Google always claimed it intends to sell relevant ads for a user, but it does not need to know who that person really is. The goal is to reach the 25-year-old runner, who lives in the countryside of Brazil and likes to jog in the morning, by providing her with ads for sneakers, but there is no need for Google to know that person’s name, address, or social security number⁵².

This paradox is the reason why Google had such a hard time explaining how it uses Gmail to target advertising. There is no individual in California who reads the messages and decides which user likes (or could potentially enjoy) specific products, the company uses robots and scans the messages for keywords that will enable targeted ads – but it should come as no surprise that users had a hard time coming to terms with the idea⁵³.

On June 2016, the panorama described above changed. When MyActivity was introduced, Google modified its privacy policy and establish a new way to sell advertising, by mixing data collected in its own platforms (Gmail, YouTube, Maps, etc.) and information acquired via third-party cookies.

MyActivity and the Rise of Personalized Web Targeting

As the Ad Rank makes clear, personal data plays a decisive role in monetizing the Search and Display Networks. In Search, personal data used to target ads is primarily the information the user types into the platform. In the DN, that information is also the users’ online behavior and demographics. But the information Google collects on users’ is much vaster than that,

⁵² Arguably, it is also a feature Google implemented in order to protect itself from requests by national security and police authorities.

⁵³ Google’s Privacy Policy states: “Our automated systems analyze your content (including emails) to provide you personally relevant product features, such as customized search results, tailored advertising, and spam and malware detection.”

though the company has long kept much of it out of reach from advertisers.

Because of the services Google provides, such as Gmail and Maps, and the array of websites that provide third-party cookies information to the DN, the company has a database that arguably encompasses sufficient information about a person to identify her name, address, work, political views, friends, hobbies (including the ones she would not care to share with others), where her friends live, at what time she regularly goes to bed, which road she most frequently takes to get to work, and even her health condition.

Google highlights that it does not sell personal data, it only uses that data to target ads. What it does sell is ad space, commercialized owing to personal information. The company also emphasizes that the auction process is entirely automatized and no human being actually reads emails, checks demographic data, or verify users' search history in order to place ads. The algorithms are responsible for running robots through each users' information and coming up with Ad Ranks.

In August 2016, however, the rules for targeting changed. The company had long resisted the trend, set primarily by Facebook, of combining personal information and advertising. But now, according to the new privacy policy, the “[i]nformation we collect when you are signed in to Google, in addition to information we obtain about you from partners, may be associated with your Google Account”⁵⁴.

Users connected and logged into one of Google's services will have their information collected and used for advertising, unless they opt-out by changing their preferences in MyActivity.⁵⁵ The concern, from a fundamental rights standpoint, is most users have the tracking mechanisms activated without understanding what changed and what the consequences for their privacy would be. In June, Google launched MyActivity, but no comprehensive explanation about the modifications and their potential risks was detailed. The move prompted two consumer groups to file complaints with the United States Federal Trade Commission (FTC), claiming it to be “deceptive”⁵⁶.

MyActivity is a central that gathers and controls all activity by a user who is logged into

⁵⁴ Available at: <<https://www.google.com/policies/privacy/#infocollect>>. Access 5 December 2016.

⁵⁵ The use of tools such as VPN or Tor will also affect data collection. Since these mechanisms are not the norm and their use, though relevant, is still very limited when compared to the bulk of people who access one of Google's services every day, I will ignore their effect.

⁵⁶ Available at: <<https://www.propublica.org/article/google-has-quietly-dropped-ban-on-personally-identifiable-web-tracking>> Access: 29 December 2016.

Google's services. The categories of data collected by Google are: (i) Web & App Activity – includes searches and browsing activity, ranging from search history to recent apps⁵⁷; (ii) Location History – a map of everywhere you go with any of your devices⁵⁸; (iii) Device Information – contacts, calendars, music, as well as information about the device itself⁵⁹; (iv) Voice and Audio Activity – commands you give your device, as well the frequency with which you tap the microphone icon⁶⁰; (v) YouTube Search History; and (vi) YouTube Watch History.

MyActivity also allows users to decide to keep these categories of data collection activated or paused. A user who pauses collection can make use of services and not have her data stored. A user who does not pause collection allows Google to use that information to make the services more accurate and possibly more useful, and also to sell advertising.

Final Remarks – The Impact of Business Models on Personal Data Regulation

Personal data is an asset for many companies, and particularly for those which, like Google, center their business model on online advertising. The tension between users' privacy and companies' interests has led to regulatory efforts in many jurisdictions, aimed at preventing abuses and controlling the ways and ranges by which personal information is collected and how it is processed and used for commercial purposes⁶¹.

⁵⁷ According to Google, the information saved as Web & App Activity includes: "Searches and other things you do on Google products and services, like Maps; Your location, language, IP address, and whether you use a browser or an app; Ads you click, or things you buy on an advertiser's site; Information on your device like recent apps or contact names you searched for; Websites and apps you use; Your activity on websites and in apps that use Google services; Your Chrome browsing history." Access 15 January 2017. Available at: <https://support.google.com/websearch/answer/54068?p=web_app_activity&hl=en&authuser=0&visit_id=1-636200982761610875-3253655500&rd=1> Access 5 December 2016.

⁵⁸ Google mentions five categories of information included in Location History: "Quality and length of your connections to cell networks, GPS, Wi-Fi networks, or Bluetooth; State of your location settings; Reboot occurrences and crash reports; Apps used for turning Location History on or off; Battery levels". Available at: <https://support.google.com/accounts/answer/3118687?visit_id=1-636200986127793708-3253655500&p=location_history&hl=en&rd=1> Access 15 January 2017.

⁵⁹ The information about a device can include the activity on your screen (whether or not it is on), the battery level, the quality of Wi-Fi or Bluetooth connection, touchscreen and sensor readings, and crash reports. Available at: <https://support.google.com/accounts/answer/6135999?p=account_device_info&hl=en&authuser=0&visit_id=1-636200987855038708-3253655500&rd=1>. Access 15 January 2017.

⁶⁰ As Google states, audio may be saved even when you are offline. Available at: <https://support.google.com/websearch/answer/6030020?p=account_voice_audio&hl=en&authuser=0&visit_id=1-636200991416533018-3253655500&rd=1>. Access 15 January 2017.

⁶¹ The rise of regulation cannot be solely attributed to the private sector's use of personal information, but that certainly was a factor in the passing of legislation. Other relevant circumstances include the collection of data by the government itself and the complexification of the digital environment.

The new model put forward by Google through MyActivity gives users control over what type of personal information is to be stored and used by the company. It is a solution clearly based on a self-regulatory model, in which users are free to do with personal data as they see fit – under these assumptions, data is to some extent understood as property. Whether or not this method is the best solution for the problem is a discussion for a different article, but what cannot be denied is that tools such as MyActivity require a considerable degree of digital education and engagement from the part of the user, as they assume the individual to have knowledge not only about herself and the data she produces, but also about the platform and how it processes her information.

Another debate regards the fit of this model to personal data regulation around the world. It is true that from the outset a surprising degree of convergence was observed in personal data regulation in many countries, which adopted similar principles in regulating the matter⁶². However, significant differences do exist, especially in terms of implementation. Google apparently has adopted a method that, as mentioned, is well-suited for a self-regulatory approach such as the one put forward by the United States. But not all jurisdictions have followed this trend. Countries with regulation more closely associated with the European model have a different take on privacy matters and particularly on the implementation of personal data protection⁶³. It remains to be seen how they will deal with Google's new privacy policy.

The businesses that rely on Big Data are growing in relevance and expanding their services. Hence, personal data issues are bound to increase. Due to the fluidity of this market and the fast-pace of technology, one should question not only if the current regulation is equipped to deal with today's problems, but also if it is sufficiently flexible to address the challenges that will arise from the development of technology and from new business models.

⁶² BENNETT, C. J. **Regulating Privacy: Data Protection and Public Policy in Europe and the United States**. (1992) Evidence of Policy Convergence, pp. 95-115.

⁶³ NEWMAN, A. L. **Protectors of Privacy: Regulating Personal Data in the Global Economy**. (2008) Privacy regimes: Comprehensive and Limited Approaches, pp. 23-41.