

Deleterious consequences: How Google's original sociotechnical affordances ultimately shaped 'trusted users' in surveillance capitalism

Big Data & Society
January–June: 1–17
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20539517231171058
journals.sagepub.com/home/bds
 SAGE

Renée Ridgway

Abstract

Google dominates around 92% of the search market worldwide (as of November 2022), with most of its revenue derived from search advertising. However, Google's hegemony over search and the resulting implications are not necessarily accidental, arbitrary or (un)intentional. This article revisits Brin and Page's original paper, drawing on six of their key innovations, concerns and design choices (*counting citations or backlinks, trusted user, advertising, personalization, usage data, smart algorithms*) to explain the evolution of Google's hypertext search engine technologies through 'moments of contingency', which led to corporate lock-ins. Underpinned by analyses of patents, statements and secondary sources, it elucidates how early Google considerations and certain affordances not only came to shape the web (*backlinks, trusted user, advertising*) but subsequently facilitated contemporary surveillance capitalism. Building upon Zuboff's 'Big Other', it describes the ways in which Google as an infrastructure is intertwined with Big Data's platformization and the ad infinitum collection of *usage data*, beyond just *personalization*. This extraction and refinement of *usage data* as 'behavioural surplus' results in 'deleterious consequences': a 'habit of automaticity,' which shapes the *trusted user* through 'ubiquitous googling' and *smart algorithms*, whilst simultaneously generating prediction products for surveillance capitalism. Advancing Latour's 'predicting the path' of technological innovation, this cause-and-effect story contributes a new taxonomy of Google sociotechnical affordances to critical STS, media history and web search literature.

Keywords

Google, sociotechnical affordance, critical STS, trusted user, surveillance capitalism, deleterious consequences

This article is a part of special theme on The State of Google Critique and Intervention. To see a full list of all articles in this special theme, please click here: <https://journals.sagepub.com/page/bds/collections/stateofgooglecritiqueandintervention>

Why the politics of search engines (still) matter¹

Facilitated by the hyperlinking of Tim Berners-Lee, during the early days of the World Wide Web, a range of search engines competed for how this 'information superhighway' might be navigated. In the 1990s, Google quickly became a dominant player with its retrieval speed and expansive directory – index – that could be quickly updated (Van Couvering, 2010: 97). Yet it was not only the web's malleability and navigability that became crucial. Written at the dawn of the millennium, Introna and Nissenbaum's article, 'Shaping the Web: Why the Politics of Search Engines Matters', expressed

concern about affording 'prominence to popular, wealthy, and powerful sites at the expense of others' (2000: 181). Their compelling argument was for the promotion of the web where a plurality of voices would be heard, warning:

Department of Digital Design and Information Studies, Aarhus University, Aarhus, Denmark

Corresponding author:

Renée Ridgway, Department of Digital Design and Information Studies, Aarhus University, Helsingørsgade 14, Aarhus, Midtjylland 8200, Denmark.

Emails: rridgway@cc.au.dk; reneridgway@protonmail.com



If search mechanisms systematically narrow the scope of what seekers may find and what sites may be found, they will diminish the overall value of the Web as a public forum as well as a broadly inclusive source of information (180).

At that time, the politics of search engines mattered, hinging on whether they would become ‘democratizing forces’ or if they would be ‘colonized by specialized interests at the expense of the public good’ (Introna and Nissenbaum, 2000: 170).

However, commercial models were already infiltrating the web and the interconnection between texts with hyperlinking would come to ‘shape the web’ and embody another type of politics. Google’s algorithm PageRank determined the relevancy of information through authority and advertising, with users expressing interest and attention by clicking on hyperlinks (Halavais, 2009). This dominance over information exchange and web commerce altered the media and technology landscape, first dubbed in 2003 by Battelle and Salkever as the ‘phenomenon of googlization’ (Rogers, 2009: 183). Sourcing certain topics and bolstering its own interests by search result ranking, Google capitalized on the ‘informational rationality of spawning value from advertising and audience labour’ (Bilić, 2017: 8). With the ‘googlization of everything’, it was Google’s ‘creep’ into other major industries, ‘advertising, software applications, geographic services, e-mail, publishing, and Web commerce itself’ (Vaidhyathan, 2011: 20) that heeded consideration. Google would eventually establish new social, cultural and political logics of search-based information societies and its economies (Lovink, 2009, Rogers, 2009).

Google is the sine qua non dominating around 92% of the search market worldwide (as of November 2022), with a majority (86%) of its revenue (209.49 billion dollars in 2021) derived from search advertising (Statista, 2022). Rather than drawing on tacit knowledge, human memory or libraries to find information, ‘ubiquitous googling’ (Ridgway, 2021) with keywords has become a habit of new media (Chun, 2016). With 8.2 billion search requests per day, collected by devices 24/7, consulting ‘the oracle’ has been grafted as a paradigm for the way users find information as they ‘voluntarily provide’ data in exchange for free services (Jiang, 2014: 212). However, the implications – Google’s hegemony over search and the resulting effects on users in an era of surveillance capitalism (Zuboff, 2015) – are not necessarily accidental, arbitrary nor (un)intentional. What were the original design affordances and considerations of founders Sergey Brin and Larry Page and how did these determine Google’s evolution?

Applying a lens of critical STS theory (Bijker and Law, 1992; Feenberg, 2017; Jasanoff, 2020; Latour, 1990; Winner, 1980) of the black-boxed socio-technical, I first show that ‘predicting the path’ of innovation is contingent

and how ‘search artifacts’ are political. I then revisit Brin and Page’s (1998) paper, ‘The Anatomy of a Large-Scale Hypertextual Web Search Engine,’ highlighting what I deem the six key innovations and concerns (*counting citations or backlinks, trusted user, advertising, personalization, usage data, smart algorithms*) which foreshadowed what was yet to come at Google. In order to ‘document the development of search’ as a means to preserve culture (Ørmen, 2016: 189), it is crucial to investigate how contemporary developments are couched historically. Engaging with Google primary sources (papers, patents, statements) and secondary literature, I propose how at ‘moments of contingency’, certain developmental design choices regarding Google’s search engine eventually became corporate lock-ins.

Advancing Bruno Latour’s ‘predicting the path’ of technological innovation, this cause-and-effect story demonstrates how Google’s design choices transitioned from theory to practice during four time periods, contributing to critical STS and media history/search literature by putting forth a new narrative and taxonomy of sociotechnical affordances. Building upon Zuboff’s (2015) article, ‘Big Other: Surveillance Capitalism and the Prospects of an Information Civilization’, I then reveal how these design affordances made the ‘logic of accumulation’ of user (Big) data possible, which in turn facilitated a ‘habit of automaticity’ in today’s surveillance capitalism. Lastly, I elucidate how an academic ‘research tool’ transitioned into an ad agency, which now finances the development of ‘smart algorithms’ (RankBrain) and other AI projects at Alphabet, furthering Sheila Jasanoff’s ‘predictive politics’ of future worlds (2020). Rather than accepting these ‘deleterious consequences’ (Agre, 1997), I close with reimagining how Google search might have been designed otherwise.

Predicting the path of innovation

During WWII, political economist Joseph Schumpeter defined innovation as the first discovery of new products or processes, identified through three stages: invention, innovation and diffusion (Schumpeter, 1942). Invention first demonstrates an idea; innovation is the acknowledgement of improvement in an idea resulting in the ‘first commercial application of an invention in the market’ and diffusion is the ‘spreading of the technology or process throughout the market’ (Greenacre et al., 2012: 5). A technological innovation in the societal sphere is then both a process and an outcome, implementing creative ideas into business environments. As new firms and novel technologies replace old products and methods, ‘entrepreneurial processes create new organisation’ (Hjorth, 2007: 721). For Schumpeter, the ‘mechanics of economic change in capitalist society pivot on entrepreneurial activity’, which became coterminous in business as a ‘creative

response' (1942: 150–151). Later, his mantra 'creative destruction is the essential fact about capitalism' (1943: 83–85) took hold in the rise of Silicon Valley technology start-up and innovation competition culture, which became the context for entrepreneurs Brin and Page and subsequently Google during the late 1990s.

Brin and Page's (1998) paper explains the superior design affordances of their innovation, a hypertexted search engine with quick indexing capability, a novel (Page)ranking mechanism and scalability. Scaling often enables cost reduction (although infrastructure becomes more expensive) and facilitates incremental improvements. However, these innovations can lead to 'lock-ins' of 'incumbent technologies and systems' where previous decisions constrain future ones (path dependencies) because adopting another standard would be costly, although a 'lock-out' of an innovation 'may be more optimal' (Greenacre et al., 2012: 3). In the new economy of information and technology communications, network effects increase the value for individual users as more users join in/adapt instead of choosing an alternative (search engine). However, technological affordances and lock-ins leave room for criticism regarding how society decides to use them, and why users adapt to these technologies with certain trade-offs such as user data, which I will address in the following sections.

Already in the late twentieth century, the complexity of these innovation processes evolved to incorporate a systems perspective, which entailed not only recognizing the structural components within the system and the multiple entities involved but also the knowledge flows resulting from interactions between actors, both human and non-human (Greenacre et al., 2012: 5). In his article 'Technology is Society Made Durable', sociologist and STS scholar Bruno Latour describes the innovation process of a hotel key with the outcome to weight the key, making it heavy and cumbersome, resulting in the guest leaving the key at the desk and the owner retaining the key (1990). Latour articulates that this:

minor innovation clearly illustrates the fundamental principle underlying all studies of science and technology: the *force* with which a speaker makes a statement is never enough *in the beginning*, to predict the path that the statement will follow (104).

Here a statement could refer to a word in a sentence, object, apparatus or institution: that is, 'anything thrown, sent, or delegated by an enunciator' with its meaning changing over time (Latour, 1990: 106).

Latour spelled out how a series of events and actors affect the translation of narratives (such as the hotel key) and that the addition of 'durability' is due to non-human actors. In his section 'Weaving together a story of technology', Latour constructs a full description of innovation, exposing how the

landscape of human actions flow without constraints, as 'powerless engineers to domination' and with the process so complete that it has become invisible (Latour, 1990: 111). Yet if the 'most productive way to create new narratives' is tracing the development of an innovation (Latour, 1990: 111), it is also to make 'predicting the path' visible. Opening and closing the 'black box at certain intervals' engenders entry points that help illuminate and understand the innovation process as 'steps in the story', which are, however, 'highly flexible, negotiable, at the mercy of the contingent event' (Latour, 1990: 109; 113).

The politics of black-boxed sociotechnical search artifacts

Proposing new concepts and 'defining trajectories by actants association and substitution', Latour displays what he calls 'translation operations' as a socio-technical network theory (1990: 113). As Andrew Feenberg points out, critical STS considers the role of social actors – a collaborator, a worker, a participant – who all engage in 'interpreting the meaning of technological artifacts', which can also determine how technology becomes the 'legitimate object of political struggle' (2017: 5). In 1980, political scientist Langdon Winner asked, 'do artifacts have politics?', addressing how the design of black-boxed technologies, in regard to their opacity and transparency, serves a political purpose. Black-boxing alleviates citizenry engagement and stymies 'opportunities to take part in managing it' because choices are made on how to do things while passing over other design decisions (Jasanoff, 2020: 31–32). Transcending the intended and unintended consequences involved in sociotechnical systems, Winner emphasized that the [innovation] process of technologies and political relationships are inherent to artifacts (1980). However, STS scholar Sheila Jasanoff points out that Winner's critique falls short in acknowledging the dynamic aspects of evolving and changing sociotechnological systems through interaction (2020: 31–32).

Introducing their anthology, *Shaping Technologies/Building Society-Studies in Sociotechnical Change*, Wiebke Bijker and John Law emphasize that it is critical to think about how users themselves reshape their technologies and that they influence design choices (1992: 3). Politics are inherent to technologies such as search engines that shape not only the Web but are also the site of 'social relativity of design' (Bijker and Law, 1992: 3), with users reciprocally influencing future social, economic and technical decisions with their interactions and data. Thus, within the innovation process, technology and society are inseparable and 'making markets blurs the distinction between economics and politics and between consumers and citizens' (Jasanoff, 2020: 39). Although forces of the market shape 'what people want', such as preferences (Jasanoff, 2020: 39), cultivating things

includes delegating values to technologies, which are ‘heterogeneous artifacts that embody trade-offs and compromises’ (Bijker and Law, 1992: 7). Imbued with political values that are ‘in some sense the (possible) facts of the future’ (Feenberg, 2017: 6), technological artifacts ‘might have been [designed] otherwise’ or another world might have been attainable (Bijker and Law, 1992: 8). Weaving numerous threads of scholarship, critical STS theory offers a framework to illuminate specific design choices regarding Google’s innovative search engine technologies as well as the values, trade-offs and politics of their effects and outcomes.

Debunking the myth by rooting out the causes

Whilst still PhD students at Stanford, Sergey Brin and Larry Page published ‘The Anatomy of a Large-Scale Hypertextual Search Engine’ (1998), in which they describe (in its various versions) the ‘anatomy’ of their innovative search engine, how PageRank works, and what was yet to come. This is one of the few, if not the only, technical paper that was ‘already a crafted rendition of its mathematical workings’, explaining how Google stood out ‘above its then competitors, and as a fundamentally democratic computational logic’, which eventually became mythologized (Gillespie, 2014: 180). Other primary sources include Google patents and interviews, along with Shoshana Zuboff’s ‘Big Other: Surveillance Capitalism and the Prospects of an Information Civilization’ (2015), a close reading of what she deems two ‘extraordinary documents’ by Google’s Chief Economist Hal Varian that disclose the origins of surveillance capitalism through the ‘logic of accumulation’ of user data. Interweaving ‘moments of contingency’ – technological developments as ‘path dependencies’ – I construct a cause-and-effect story through an interdisciplinary discourse analysis. By revisiting Brin and Page’s original paper and other primary sources, underpinned by analyses of secondary sources, patents and diagrams – notably mostly content *not* enounced by Google’s own Public Relations/marketing department and websites – my aim is to advance a new taxonomy of sociotechnical affordances that shows how the contemporary condition of Google’s surveillance capitalism evolved. Analogous to a black box with inputs and outputs, in the following I attempt to tackle the problem at either end to ‘root out the causes or contain the effects’ (Jasanoff, 2020: 36), beginning with Google’s early design affordances and considerations from Brin and Page’s (1998) paper.

The anatomy of a large-scale hypertextual search engine: Six key innovations and concerns

Counting citations or backlinks

The genealogy of the PageRank algorithm has its antecedents in bibliometrics. One of its founders, Eugene

Garfield, invented the Scientific Citation Index (SCI) that was the basis for an evaluative metrics to measure the ‘impact factor’, which is how academia values a researcher’s published articles. With SCI, the number of citations is divided by the number of pages in the entire book and these citations provide a type of ‘peer review’. Larry Page references Garfield in his US patent ‘Method for Node Ranking in a Linked Database’ from January 9, 1998 (Figure 1); Brin and Page also acknowledge how SCI is applied to PageRank, stating that ‘[a]cademic citation literature has been applied to the web, largely by counting citations or backlinks to a given page’ (1998). Whereas most ‘search engines associate the text of a link with the page that the link is on’ or ‘content scores’, Brin and Page state that ‘[i]ntuitively, pages that are well cited from many places around the web are worth looking at’ (1998). In other words, PageRank calculates each page in the index, yet pages differed in importance as links from all pages are not counted equally. Additionally, Brin and Page hint at PageRank’s citation bias:

A page can have a high PageRank if there are many pages that point to it, or if there are some pages that point to it and have a high PageRank (1998).

Trusted user


Besides ‘shaping the web’ through hyperlinking, the user needs to decide which links to click on. Brin and Page’s goal was to provide users with ‘the most relevant search results for their query’ (1998); to this end, the algorithm ‘efficiently calculated the combined values of pre-weighted objects in the index database’ (Gillespie, 2016: 19–20). In their Feedback section, Brin and Page explain that parameters such as ‘type-weights’ and ‘type-prox weights’ determine the ranking of search results and further state:

Figuring out the right values for these parameters is something of a black art. In order to do this, we have a user feedback mechanism in the search engine. A trusted user may optionally evaluate all of the results that are returned. This feedback is saved (1998).

Brin and Page also made it known they had not done an ‘extensive user study’ to test the quality; instead, they invited the *trusted user* to test out Google (Brin and Page, 1998).

Advertising

Brin and Page assert that less advertisement would be needed to help the ‘consumer’ [*trusted user*] find what they want and that this is what constitutes a ‘better’ search engine because ‘advertising always wants customers to acquire new products’ (1998). In their section Academic Search Engine Research, Brin and Page explain the shift



US006285999B1

(12) **United States Patent**
Page

(10) **Patent No.: US 6,285,999 B1**
(45) **Date of Patent: Sep. 4, 2001**

(54) **METHOD FOR NODE RANKING IN A LINKED DATABASE**

(75) Inventor: **Lawrence Page**, Stanford, CA (US)

(73) Assignee: **The Board of Trustees of the Leland Stanford Junior University**, Stanford, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/004,827**

(22) Filed: **Jan. 9, 1998**

Related U.S. Application Data

(60) Provisional application No. 60/035,205, filed on Jan. 10, 1997.

(51) Int. Cl.⁷ **G06F 17/30**

(52) U.S. Cl. **707/5; 707/7; 707/501**

(58) Field of Search **707/100, 5, 7, 707/513, 1-3, 10, 104, 501; 345/440; 382/226, 229, 230, 231**

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,953,106	*	8/1990	Gansner et al.	345/440
5,450,535	*	9/1995	North	395/140
5,748,954	*	5/1998	Mauldin	395/610
5,752,241	*	5/1998	Cohen	707/3
5,832,494	*	11/1998	Egger et al.	707/102
5,848,407	*	12/1998	Ishikawa et al.	707/2
6,014,678	*	1/2000	Inoue et al.	707/501

OTHER PUBLICATIONS

S. Jeromy Carriere et al, "Web Query: Searching and Visualizing the Web through Connectivity", Computer Networks and ISDN Systems 29 (1997), pp. 1257-1267.*

Wang et al "Prefetching in World Wide Web", IEEE 1996, pp. 28-32.*

Ramer et al "Similarity, Probability and Database Organization: Extended Abstract", 1996, pp. 272.276.*

Craig Boyle "To link or not to link: An empirical comparison of Hypertext linking strategies". ACM 1992, pp. 221-231.*

L. Katz, "A new status index derived from sociometric analysis," 1953, Psychometrika, vol. 18, pp. 39-43.

C.H. Hubbell, "An input-output approach to clique identification sociometry," 1965, pp. 377-399.

Mizruchi et al., "Techniques for disaggregating centrality scores in social networks," 1996, Sociological Methodology, pp. 26-48.

E. Garfield, "Citation analysis as a tool in journal evaluation," 1972, Science, vol. 178, pp. 471-479.

Pinski et al., "Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics," 1976, Inf. Proc. And Management, vol. 12, pp. 297-312.

N. Geller, "On the citation influence methodology of Pinski and Narin," 1978, Inf. Proc. And Management, vol. 14, pp. 93-95.

P. Doreian, "Measuring the relative standing of disciplinary journals," 1988, Inf. Proc. And Management, vol. 24, pp. 45-56.

(List continued on next page.)

Primary Examiner—Thomas Black
Assistant Examiner—Uyen Le
(74) *Attorney, Agent, or Firm*—Harrity & Snyder L.L.P.

(57) **ABSTRACT**

A method assigns importance ranks to nodes in a linked database, such as any database of documents containing citations, the world wide web or any other hypermedia database. The rank assigned to a document is calculated from the ranks of documents citing it. In addition, the rank of a document is calculated from a constant representing the probability that a browser through the database will randomly jump to the document. The method is particularly useful in enhancing the performance of search engine results for hypermedia databases, such as the world wide web, whose documents have a large variation in quality.

29 Claims, 3 Drawing Sheets

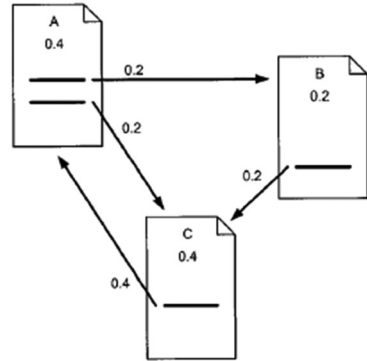


Figure 1. Larry Page's 'method for node ranking in a linked database'(2001) patent 6285999.

from academic to commercial web search development (1998).

Aside from tremendous growth, the Web has also become increasingly commercial over time. In 1993, 1.5% of web servers were on .com domains. This number grew to over 60% in 1997. At the same time, search engines have migrated from the academic domain to the commercial (1998).

This could be considered a small critique of the rise of .coms at that time. Furthermore, Appendix A: Advertising and Mixed Motives explicitly articulates that income derived from ads ‘provide[s] an incentive to provide poor quality search results’ (Brin and Page, 1998).

Personalization

As stated previously, Brin and Page adopted a feedback mechanism—the *trusted user* who evaluates the search results.

Then when we modify the ranking function, we can see the impact of this change on all previous searches which were ranked. Although far from perfect, this gives us some idea of how a change in the ranking function affects the search results (1998).

In their Future Work section, Brin and Page foresee applying *trusted user* search history to PageRank, which ‘can be personalized by increasing the weight of a user’s home page or bookmarks’ (1998). Moreover, they also ‘plan to support user context’ and ‘result summarization’ with these adjustments determining the ranking function, as well as previous searches, with the impact of these changes resulting in personalization (Brin and Page, 1998).

Usage data

Brin and Page note that companies whose business model is co-opting user attention and manipulating the ‘unseen’ metadata contained in search results for profit is a ‘serious problem’, yet they are keenly aware of data generated from usage:

Usage was important to us because we think some of the most interesting research will involve leveraging the vast amount of usage data that is available from modern web systems. For example, there are many tens of millions of searches performed every day. However, it is very difficult to get this data, mainly because it is considered commercially valuable (Brin and Page, 1998).

Additionally, Brin and Page in their Appendix B acknowledge other potentials due to Google’s scalability:

‘Of course there could be an infinite amount of machine generated content, but just indexing huge amounts of human generated content seems tremendously useful’ (1998).

Smart algorithms

Brin and Page state in their ‘Future Work’ section that areas such as ‘updating’ need attention and that much ‘remains to be done’. This includes smart algorithms that would ‘discriminate what should be recrawled and new ones to be crawled’ (1998), presaging artificial intelligence methods (machine learning algorithms) applied to search.

Moments of contingency

In the following, I open and close the black box at certain intervals (Latour, 1990: 109), as ‘moments of contingency’, mapping out the successive considerations, choices, and changes that resulted in the gradual transformation of Google’s design affordances, structured by four time periods.

Shaping the web (1998–2002): PageRank, relevance, trusted user

Taking all of the work in sociometry, citation analysis, and hypertext navigation together, one could argue that all the ‘ingredients’ for PageRank were available from the middle of the 1990s, and that all one had to do was to combine them (Rieder, 2012: 7).

Before Brin and Page wrote their seminal text, Lawrence Page, Sergey Brin, Rajeev Motwani and Terry Winograd published (January 28, 1998) the paper ‘The PageRank Citation Ranking: Bringing Order to the Web’ that first defined ‘PageRank’ and showed how it could be calculated – a page has high rank if the sum of the ranks of its *back-links* is high. As noted above, PageRank is the ‘global ranking of all web pages, regardless of their content, based solely on their location in the Web’s graph structure’ (Page et al., 1998). Google’s innovation was to introduce an importance score that gauges the status of a page separate from a user query and by ‘analysing the topology of the Web graph’ they ‘revolutionised the field of the Web’ (Franceschet, 2010). PageRank is therefore both a link analysis and an IR (Information Retrieval) system (Rieder, 2012: 8), first indexing and then ranking hyperlinks for users, as well as functioning as a visibility engine for certain links and sites, with this ‘authority’ of ranking determining relevance.

Brin and Page’s hyperlinking brought the *trusted user* to another destination to solve the ‘lost in hyperspace’ problem by deciding which links receive visibility and

directing the attention flow on the ‘information highway’ of the web, simultaneously exuding citation bias. In this way, Google is a ‘navigational media’: a ‘type of technically-based media actor that organizes and directs audiences or users to various types of content’ (Van Couvering, 2010: 225). Economically speaking, search results are an expenditure similar to commercial TV programming in that the service is given away for free in order to attract an audience to sell ads, but PageRank also converted attention into action. Brin and Page’s *trusted user* instantiated the feedback mechanism by interacting with the algorithmic interface through keyword searches and by clicking on links, which would become central to reinforcing ‘preferential attachment’ (1998). This interaction quantified user satisfaction ‘in terms of percent clicks on the top results’ (Gillespie, 2016: 19–20), mostly found on the first page of Google, whilst actuating the ‘relevance’ and ‘quality’ of the search results and supplying feedback as data. Moreover, search engines reconfigure the free labor of users to ‘draw attention to themselves. That attention is then sold to producers who wish to advertise their products’ (Halavais, 2009: 83).

Advertising (2003–2007): Traffic, database of intentions, IPO

Frédéric Kaplan sums up the success of Google’s innovation as the story of two algorithms:

The first—pioneering a new way of associating web pages to queries based on keywords—has made Google popular. The second—assigning a commercial value to those keywords—has made Google rich (2014: 57).

Not only hyperlinking and PageRank became important ranking criteria. The monetization of certain keywords (Kaplan, 2014) reflected the users’ thoughts as queries, imparting the power of a certain ‘semantic governmentality’ (Feuz et al., 2011). In 1998, seven months before Google, one of the key developments in the search industry in response to spam was the pay-for-placement search engine GoTo.com, which implemented the purchase of search terms (keywords) by advertisers. As Van Couvering notes, these ‘bid-placements’ operated as an auction, where advertisers competed and GoTo’s model was not to make advertisers pay ‘per impression’ but that the ‘advertiser was only liable for the fee when someone actually clicked the ad – unclicked impressions were given away for free’ (2010: 113–115). Although Brin and Page did not originally wish to mix organic search with paid ads, as stated above, it is noteworthy to show their resolve in solving the problem at hand – making Google economically viable whilst locking themselves into auction-based *advertising* revenue models.

On October 23, 2000, Brin and Page eventually launched AdWords (now called Google Ads), connecting users’ queries to advertisements, which enabled businesses to buy ‘text ads on search-results pages’ that appeared on the margins of the webpage (Oremus, 2013). Google’s key innovation over GoTo (later Overture) was to sell off all slots at once in a second-price auction: the winner pays the price of the second bid plus 1 cent, the second pays the price of the third bid plus 1 cent and so forth. The success of AdWords then became the combination of auction results as well as a ‘quality score’ – the relevance of the ad to the keyword along with the quality of the page the ad links to and most importantly, the amount of clicks that shows users’ interest: what came to be known as ‘Googlenomics’ or, in business jargon, microeconomics (Levy, 2009).² In June 2003, Google launched AdSense, also an auction-based model which syndicated CPC (cost per click) banner ads to partners automatically, placed with JavaScript code on non-Google websites relative to the content of the page.

Already ‘large scale’ by March 2003, Google had over 100,000 advertisers and with the monetization of keywords, a highly intricate series of communication networks and commercial platforms ‘connecting users to advertisers, within a burgeoning media economy’ (Van Couvering, 2010: 124), shaped Google search results. However, through this platform intermediation, Google was able to make ‘connections’ (Langley and Leyshon, 2016: 3; Van Dijck, 2013) and, with its ‘quasi-monopoly position’, gain a competitive advantage from network effects aggregating users and regulating the ‘flow of capital’ (Srnicek, 2016: 31). Revenue and viewer clicks increased with Google controlling what the ‘eyeballs’ saw, as not all advertisements were equally attractive. This battle for attention expedited that ‘search engines were becoming one of the most visited kinds of sites on the web’ (Halavais, 2009: 78). Thus, PageRank’s keyword attention economy facilitated ‘the creation and exploitation of a new commodity for media: traffic’ (Van Couvering, 2010: 92). As Rieder and Sire explain: Internet users query, Google indexes content and advertisers ‘attract visitors *beyond* the traffic received from “organic” results, with Google subsidizing the former and charg[ing] for the latter’ (2013: 5).

On August 19, 2004, Google arranged its IPO (Initial Public Offering), valued at 27 billion dollars. This second lock-in forced Google to answer to shareholders instead of users.³ Search engines were once considered portals yet by 2005, Google’s syndicated *advertising* comprised 44% of its ad revenue (Van Couvering, 2010: 116). By the mid-2000s, users entered keywords and interacted with the search interface that delivered results, both organic and advertisements, and by clicking on mostly the top search results, supplied feedback to Google. This resulted in a third lock-in – the construction of a ‘massive click stream database of desires, needs, wants and preferences that can be discovered, subpoenaed, archived, tracked and exploited for all sorts of ends’ (Battelle,

2006: 7). Although not structured a priori or by specific intent, this ‘database of intentions’ collects ‘our curiosities and thoughts’ as queries and thereby ‘functions as one of the most ubiquitous and powerful record keepers of digital engagement’ (Noble, 2018: 126). Above all else, this ‘vast archive’ became the ‘backbone of digital media economics’ and Google’s ‘key monetizable resource’ (Jarrett, 2014: 17–19).

With the ‘googlization of everything’ (Vaidhyanathan, 2011), many other start-ups or successful companies were bought out, which united Google services and data. On April 13, 2007, a year after its procurement of YouTube for 1.65 billion dollars, Google purchased the web advertising company DoubleClick for 3.1 billion dollars in cash. Although the FTC (Federal Trade Commission) required additional information, it eventually approved Google’s acquisition that also ‘gave Google access to DoubleClick’s user metrics and allowed it to track users on any site on which Google advertising appears’ (Hillis et al., 2013: 17). Additionally, the takeover expanded Google’s relationship beyond algorithm-driven ad auctions, incorporating DoubleClick’s network of web publishers, advertising agencies, and software. This fourth lock-in had even more far-reaching implications for Google; not only were auction-arranged advertisements its main source of revenue, but *trusted user* interactions became constantly monitored and collected as data.

User search data (2004–2012): Personalization, logic of accumulation, behavioural data

Unbeknownst to users, over the preceding years and behind the scenes, Google coevally carried out bucket or A/B testing, where two webpages (A and B) are compared to each other to study performance. Google collated not only user locative and interaction data as ‘signals’ but stored all a user’s queries, creating individual records of search histories. *Personalization* was first released in beta on March 29, 2004 and in April 2005 as a non-beta service, which was subsequently applied to Google Search on November 11, 2005, but only for users with Google accounts. On December 4, 2009, Google announced publicly that personalized searches were for everyone. Therefore, in addition to monitoring the relevance of search result web pages (PageRank and other ‘signals’) with a browser cookie record (DoubleClick), the sites already visited would affect subsequent searches. Even if not signed into a Google account, by pinpointing user/device geographic locations and collating IP (Internet Protocol) addresses, as well as maintaining a log of previous queries and visited web pages, *personalization* adapted these into real-time search results. Google claimed that every search was unique and therefore delivered ‘customised results’ for each user that satisfied her interest. Moreover, this fifth lock-in differentiated Google from its competitors,

with its self-promotion as the search engine that delivered personalized results, making the *trusted user* feel unique whilst publicly justifying the collection of user data.

In ‘Big Other: Surveillance Capitalism and the Prospects of an Information Civilization’ (2015), Shoshana Zuboff recounts two articles (2010, 2014) by Google’s Chief Economist Hal Varian that provide insight into Google’s gathering of user data. First, Varian’s ‘Computer Mediated Transactions’ divulges the ‘pervasive and continuous’ recordings of computer-mediated economic transactions, which facilitate ‘rendering an economy transparent and knowable in new ways’ and will continue to impact the economy for the foreseeable future (2010: 2 cited by Zuboff, 2015: 77). Drawing also on her earlier research, Zuboff proposes that besides ‘automating’, corporations learn from how workers work as a result of the ‘informing’ process, where computer-mediated work extends the organizational codification of ‘real-time’ data collection and management systems into what she calls the ‘electronic text’ (Zuboff, 2015: 77). Later, this codification became applied to users’ interactions with Google’s search interface that facilitated the ‘logic of accumulation’ in which it is imbedded and the conflicts inherent to that logic, as it ‘organizes perception and shapes the expression of technological affordances at their roots’ in a ‘largely invisible’ way (Zuboff, 2015: 77).

Varian’s second article ‘Beyond Big Data’, emphasized one of the ‘new uses’ stemming from computer-mediated transactions, ‘personalization and customization’ (2014, cited by Zuboff, 2015: 78). Zuboff explains that with Google’s capturing of search queries, users supplied ‘behavioral data’ that was then recirculated with ‘Interactive Learning’ and ‘Product-Service Improvements’ (Zuboff, 2017). These refined search results were made possible due to Google’s ‘extensive retention of search data’ (Zuboff, 2015: 80), recycled to improve its services that enhanced user search experience. Exemplified by their ‘ubiquitous googling’, users became habitually accustomed to personalized search services, evermore reliant on the results that engendered a sixth lock-in. Baited with ‘hooks that lure’ the Behavioral Value Reinvestment Cycle was a ‘completely closed loop, a self-contained process.... in which the users’ experiences were ends in themselves’: all the value users created was reinvested in that experience (Zuboff, 2017, 2019) (Figure 2).

Data extraction and analysis (2009–2015): Usage data, behavioural surplus, smart algorithms

Google’s improvement of search services continued, with *personalization* and targeted *advertising* affecting *trusted users*’ results during their hyperlinked journeys; simultaneously they left enormous trails of atomized ‘data exhaust’ (Mayer-Schönberger and Cuckier, 2013). With

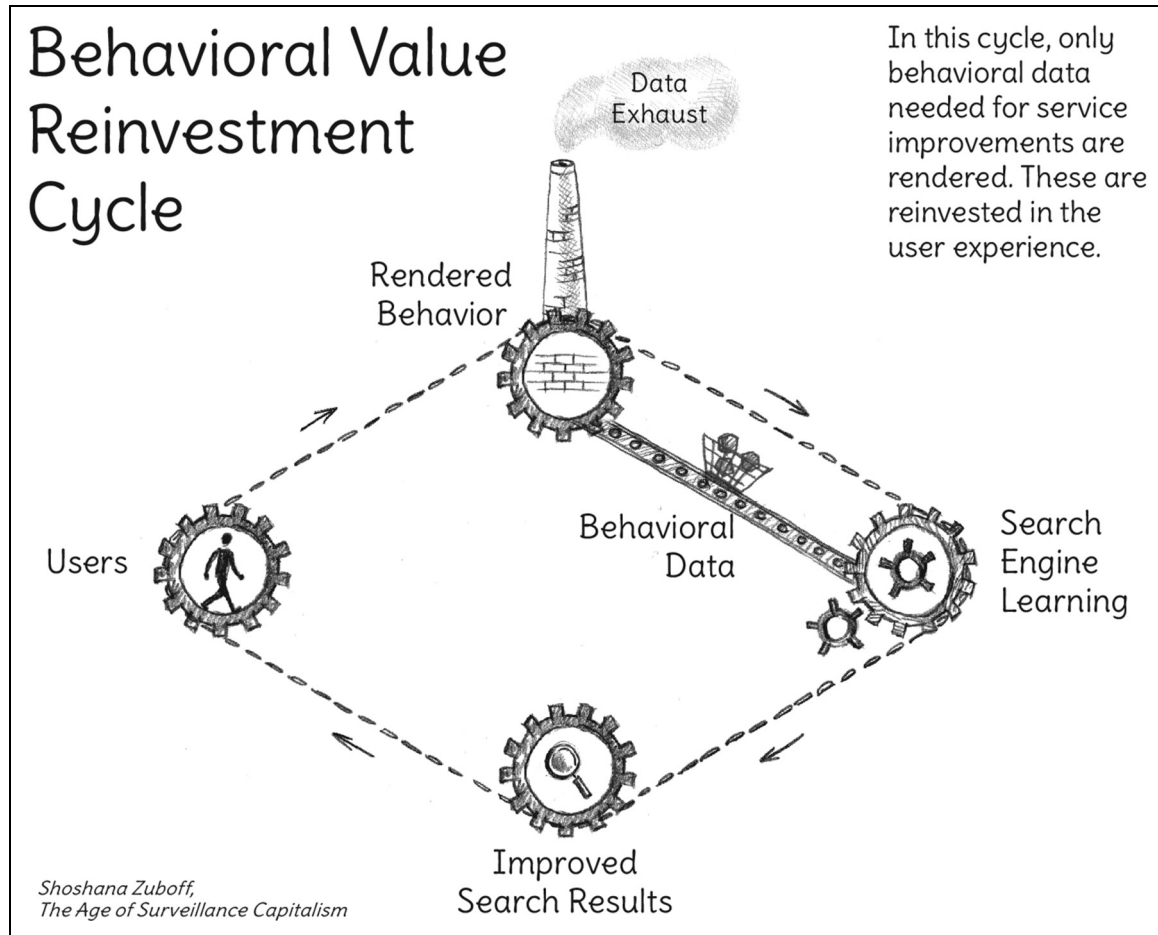


Figure 2. The behavioural value reinvestment cycle (Zuboff, 2019: 70). Artist rendering, Karin Schwandt.

this atomistic capture of search data, Hal Varian realized early on that user data is ‘ubiquitous and cheap’, yet the ‘analytic ability to utilize that data’ was scarce (Levy, 2009). Moreover, despite recycling and refining its treasure trove of *usage data*, Zuboff cogently points out that the Behavioral Value Reinvestment Cycle was not enough considering it used up all of the value users produced (2017). Keyword terms alone cannot measure affect because single-word terminologies of search queries are categorized as ‘revealed preferences’ and do not provide enough ‘information’ as data (Varian, 2014). Due to the pressure mounted from investors at Google as a result of the second lock-in, the decision was made to access the surplus – in this case behavioural data – and to control what users would click on in order to boost ad revenue (Zuboff, 2017).

Therefore, another of Varian’s ‘new uses’ stemming from computer-mediated transactions, ‘data extraction and analysis’, came to pass (2014). Varian’s use of the word extraction connotes a ‘taking from’ rather than either a ‘giving to’ or a reciprocity of ‘give and take’, and ‘also sheds light on the social relations implied by “formal indifference”’ (Zuboff, 2015: 79). With Google’s extraction

activities as a one-way process that augments the emerging ‘logic of accumulation’, there is no relationship. Deemed valuable, Google explicitly found innovative methods to collect human experiences and interests, which users had intentionally opted to keep private. Applying advanced technical know-how, ‘Google was able to access this behavioural data—irrespective of users’ intentions—even when users’ intentions were made explicit’ (Zuboff, 2017). Yet data is never ‘raw’ (Gitelman and Jackson, 2013) or just ‘given’. Moreover, the ‘data culled from these processes of extraction is not “pure” but rather marked by the traces of the bodies from whence it came’ (Thylstrup, 2019: 1). It is an asymmetrical power relationship where the *trusted user* is kept in the dark about Google’s ‘extraction’ practices, including the ‘full range of personal data that they contribute to Google’s servers, the retention of those data, or how those data are instrumentalized and monetized’ (Zuboff, 2015: 81–83).

Google started to relate behaviour as a by-product of query activity and the Behavior Value Reinvestment Cycle became subordinated to a new kind of market process. This ‘logic of accumulation’ facilitated an

expanding business where ‘users were no longer ends in themselves, instead they became a means to other [commercial] ends’ (Zuboff, 2017). Over time, with the ‘logic of accumulation’ of users’ search data,

Google understood that were it to capture more of these data, store them, and analyze them, they could substantially affect the value of advertising (Zuboff, 2015: 85).

Behavioural data became extended, extracted and commodified as ‘behavioral surplus’ and the seventh lock-in was reached (Figure 3). Furthermore, Google’s extraction of ‘behavioral surplus’ created a zero-cost game changer – a new class of ‘surveillance assets’ that provided a genuine market exchange, yet difficulties and ethical issues were implicit in this business model from the outset (Zuboff, 2017).

Additionally, extraction of ‘behavioral surplus’ generated untamed learning opportunities from *usage data*, reflecting Varian’s ‘continuous experimentation’ terminology of usages (2014), including applying artificial intelligence (AI). Since 2015, PageRank has been shifting to RankBrain: a machine learning algorithm that is being implemented to field unrecognizable queries (around 20%) and now deemed the third ‘signal’ after links and content (Sullivan, 2016). As it attempts to interpret the intention of the *trusted user*, Brin and Page’s *smart algorithms* a.k.a. RankBrain, is learning as well as influencing searches and ‘modifying behavior to produce new varieties of commodification, monetization, and control’ (Zuboff, 2015: 85). These AI-driven technologies are ‘intelligent systems based on tracking, interpreting and predicting intuitive behavior of human actors’ (Van Dijck, 2010: 587). Google realized that as user behaviour is influenced and modified – thereby shaped – it increasingly improves the quality of ‘surveillance assets’ that are refined as prediction products, which are now ‘sold into a new kind of marketplace – a market that trades exclusively in future behavior’ (Zuboff, 2017).

Big Data’s Big Other: Surveillance capitalism’s habit of automaticity

My prediction is that when AI happens, it’s going to be a lot of computation and not so much clever blackboard, whiteboard kind of stuff, clever algorithms [...] My theory is that if you look at your programming, your DNA, it’s about 600 megabytes compressed. So it’s smaller than any modern operating system. Smaller than Linux or Windows or anything like that. Your whole operating system. That includes booting up your brain by definition. And so your program algorithms probably aren’t that complicated [...] We have some people at Google who are really trying to build artificial intelligence, and to do it in a large scale. And in fact, to make search better, [...] to do the perfect job of search, you could ask

any query and it would give you the perfect answer. And that would be artificial intelligence, right? (Larry Page, February 16, 2007).

As shown above, Varian’s ‘computer-mediated economic transactions’, or Zuboff’s ‘electronic text’ and ‘logic of accumulation’ are expressions of technologies that have been imposed on users’ searching activities. With its *advertising* business model generating revenue and profit, Google (*Big Other*) capitalized on users willingly sharing their data through the platform (Srnicek, 2016) in exchange for search results. Even if Google does not technically ‘sell’ and ‘share’ its treasure trove of user data as ‘unified dossiers’,

large online platforms mostly let other companies *utilize* their data without fully transferring it, and they let them use their infrastructure to collect more data, to the benefit of both the client companies and the platforms themselves (Christl, 2017: 16).⁴

‘Behavioral surplus’ marketed as ‘surveillance assets’ (Zuboff, 2017) interests data brokers, who seek to combine behavioural data with diverse user groups. These secondary usages of data thereby facilitate the exchange of relevant information for the profiling and surveillance of citizens (Mayer-Schönberger and Cukier, 2013: 103; Christl, 2017: 16). This connectivity and exchange between users, advertisers and now data brokers, is part of the ‘service/dataprofile/ advertising complex’ (Lovink and Tkacz, 2015: 15) and it has become clear that Google search ‘is a personal information economy where the standard exchange is service for profile’ (Elmer, 2003; Rogers, 2009).

The past two decades, by *counting backlinks* with PageRank, Google contemporaneously extracted, organized, and stored ‘the real-time history of post-Web culture’ as data and, by building ‘possibly the most lasting, ponderous, and significant cultural artifacts in the history of humankind’, its ‘database of intentions’ enabled ‘a new culture to emerge’ (Battelle, 2006: 7 cited by Noble, 2018: 148). As each individual has been profiled in the database, Google is able to track users and correlate previous collated data, along with new services and interactions, thus operating as a ‘profiling machine’ (Elmer, 2003; Van Dijck, 2010: 583) (Figure 4). Although deemed *personalization*, individual users’ search queries or clicking on results are ‘assessed by comparison to a network of standardized behavioural expressions of others’ (Alaimo and Kallinikos, 2019: 401). These databases organize ‘digital subjects’ (Goriunova, 2019) into categories or groups based on attributes, interests or preferences of others who are similar (Beer, 2009; Chun, 2016; Feuz et al., 2011; Gillespie, 2014; Noble, 2018). By linking and sharing digital profiles with advertisers and other sources, databased

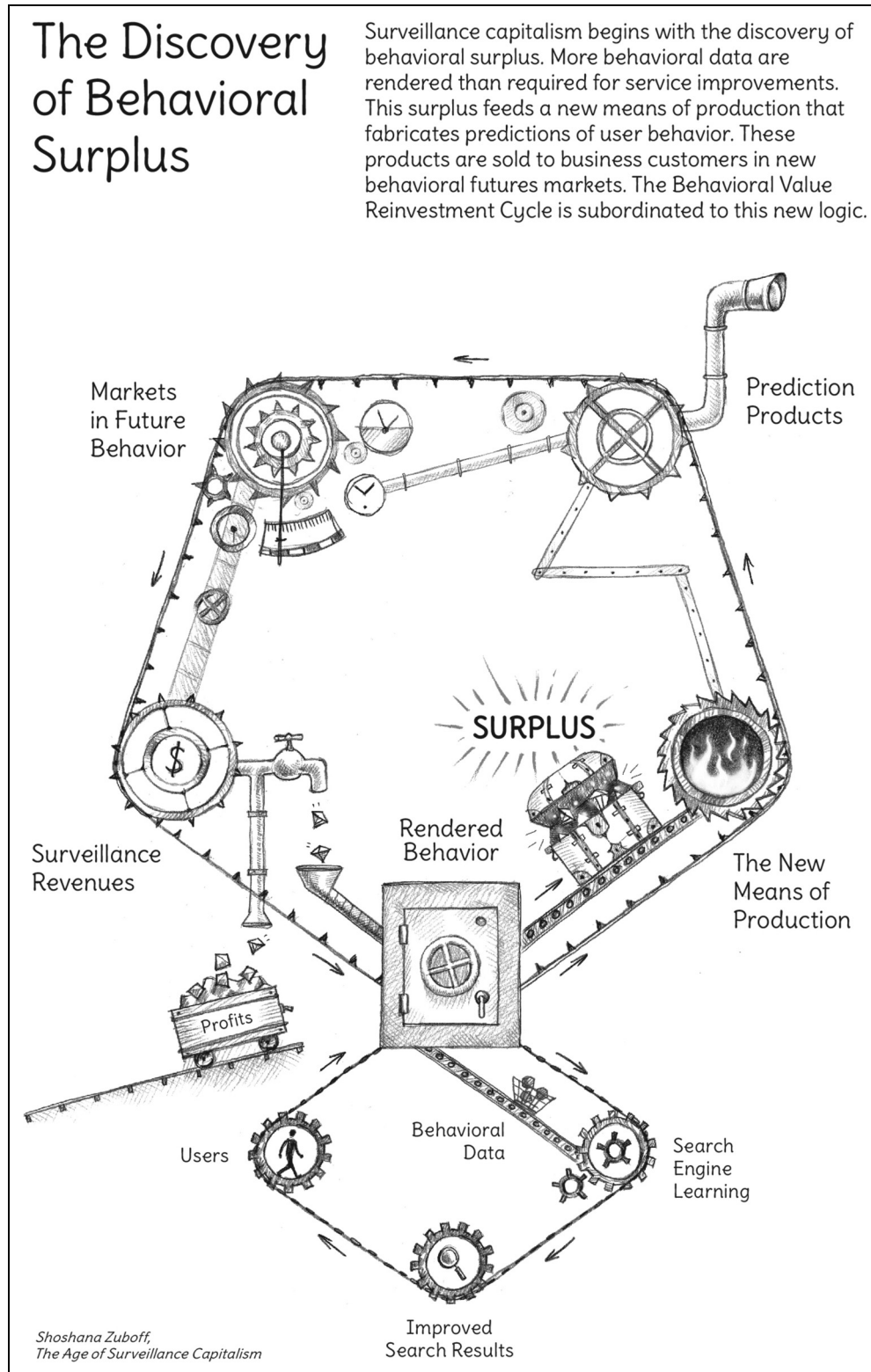


Figure 3. The discovery of behavioural surplus (Zuboff, 2019: 97). Artist rendering, Karin Schwandt.

trusted users are grouped together into ‘customer match audiences’ that become part of ‘Google Properties’ and,

additionally, correlated to ‘similar audiences’ within the YouTube and Gmail databases (Figure 5).

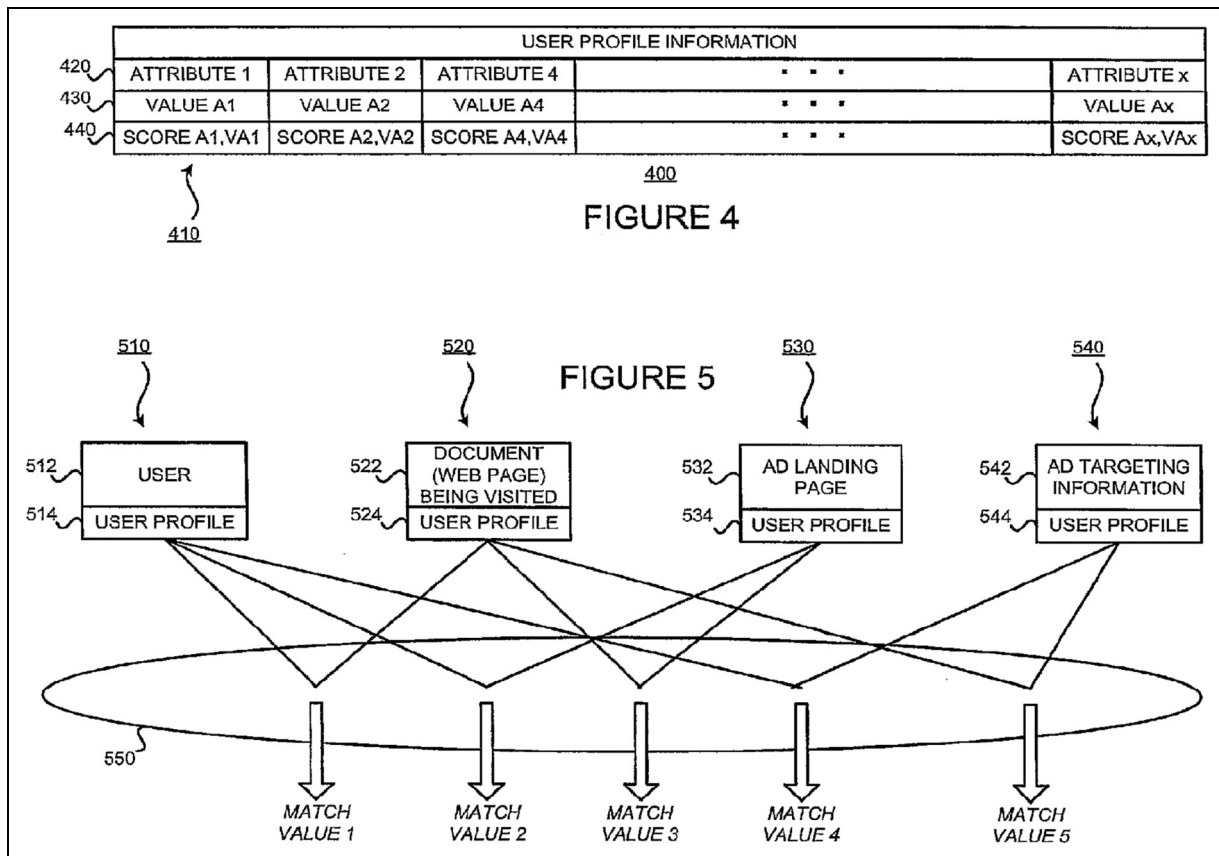


Figure 4. Patent for 'generating user information for use in target advertising'.

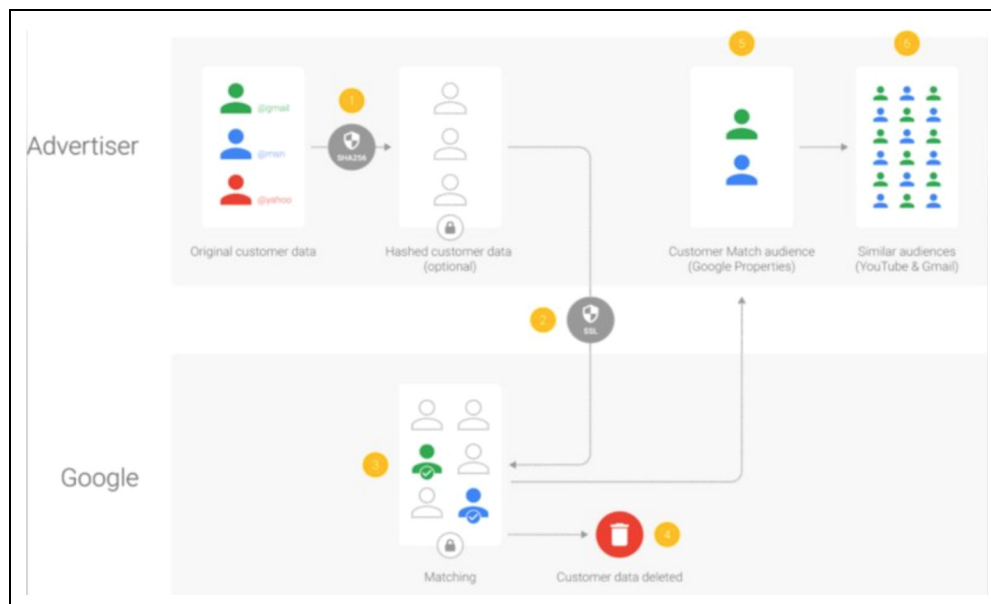


Figure 5. How Google constructs customer match audiences and similar audiences.

Made possible with the proliferation of *Big Data* (Mayer-Schönberger and Cukier, 2013), these 'digital procedures' that collect habitual search data and then 'identify

correlations, analyze patterns, and produce additional information' replace statistical validity (Esposito, 2017: 01). Zuboff stresses – echoing critical STS theorists Bijker and

Law (1992), Latour (1990) and Feenberg (2017) – that *Big Data* is not a technological object, an effect, an autonomous process, nor a thing in itself but rather originates and is ‘embedded in the social’ and is ‘both a condition and an expression’ of surveillance capitalism (2015: 77). As Phil Agre pointed out early on, ‘capture is never purely technical but always sociotechnical in nature’ (1994: 112). This aligns with Zuboff’s ‘logic of accumulation’ that comprises ‘surveillance capitalism’, which decides what is left out, what is accumulated and how it is organized, constructing ‘its own social relations and with that its conceptions and uses of authority and power’ (2015: 77). Moreover, surveillance capitalism’s extraction of human subjectivity discloses the logic and implications of *Big Other* as well as *Big Data*’s ‘foundational role in this new regime’ (Zuboff, 2015: 76). It is not only the classical *Big [Br] Other* activities of the state – surveillance – but the corporative and predatory nature of *Big Other* as it confiscates users’ queries and interactions.

The logic of accumulation and the extraction process of ‘behavioural surplus’ is made possible because users have entered a ‘data contract’ by agreeing to terms of use: Google organizes the world’s information and makes it accessible and useful by delivering relevant search results for free. In return, *trusted users* are now ‘locked-in’, what Zuboff deems as a ‘twenty-first century Faustian pact’, supplying and providing data ad infinitum that perpetuate *Big Other* (2015: 83). Zuboff refers to Varian’s ‘new contractual forms due to better monitoring’ to point out that getting ahead of legislation for users’ privacy, surveillance capitalists ‘skillfully exploited a lag in social evolution’ (2015: 82). Whereas Google can invoke its privacy rights ‘as legitimization for maintaining the obscurity of surveillance operations’, users’ privacy rights are eroded (Zuboff, 2015: 82). Only those in power, surveillance capitalists, exude the capacity to enact ‘a new form of power in which contract and the rule of law are supplanted by the rewards and punishments of a novel kind of invisible hand’ (Zuboff, 2015: 82). Exceeding the Bentham brother’s panopticon of surveilling, monitoring and controlling human subjects and instead of contracts, governance, and legislation, surveillance capitalism is ‘a new kind of sovereign power’:

In a world of *Big Other*, without avenues of escape, the agency implied in the work of anticipation is gradually submersed into a new kind of automaticity— a lived experience of pure stimulus-response (Zuboff, 2015: 82).

Following Zuboff, *Big Other* and *Big Data* break with the past—whereas ‘capitalism once profited from products, then profited from services, it now profits from surveillance’ (2016). Expressed otherwise, whereas power has been associated with ownership and, in Marxist terms, means of production, with surveillance capitalism

ownership is identified with ‘means of behavioral modification’ (Zuboff, 2015: 82). Nowadays, the ‘goal is to program customers’ in order that they ‘act in certain ways (or to predict present conditions or future habits)’ (Chun, 2016: 58). As demonstrated by Brin and Page’s socio-technical design affordances that ‘predicted the path’ of innovation at contingent moments, with ‘ubiquitous googling’ as a ‘habit of automaticity’, there is no escape from the ‘automated ubiquitous architecture of *Big Other*’ (Zuboff, 2015: 86). Whereas once Google shaped the web, it now decides what data is accumulated and how it is organized, shaping the *trusted user* in the process (Ridgway, 2021). Purporting an ‘inevitability doctrine’ that Google knows users better than they know themselves and that they are helpless, its technological programming targets human agency by deleting resistance and creativity from the text of human possibility (Zuboff, 2017).

Research tools

In 1998, Brin and Page published their paper as a ‘research project’ while they were still PhD students at Stanford. Although PageRank’s novel recipe is still shrouded by IP rights, ‘The Anatomy of a Large-Scale Hypertextual Web Search Engine’ exposes some of their original intentions that came to shape the Web, *counting citations or (back) links*. Research and search are therefore inextricably interlinked, not only through hyperlinked citations (PageRank) but that most search engines were first developed in academia before transitioning to commercial enterprises (Van Couvering, 2010). Returning to their considerations of design affordances, in Appendix A, Brin and Page shared their belief that the ‘issue of advertising causes enough mixed incentives’ (1998), yet at that time Google did not have a business model to make their hypertext web search engine economically viable. Despite this statement, Google’s intermediation with advertisements affecting search results resulted in it transitioning from an academic research-oriented search engine into a ‘multinational advertising company’ (Noble, 2018: 50).

As shown above, ‘moments of contingency’ yielded technological lock-ins over recent decades based on path dependencies and trade-offs, making visible Brin and Page’s original innovations that incited ‘ubiquitous googling’ as well as advertisers participating in Google Ads auctions. Echoing back to socio-technical actor network theory (Latour, 1990), with more companies joining the Google Ads market, exchanges are no longer just between the *trusted user* and platforms but with other actors – ‘customers who learned how to make money from low-risk bets on a user’s future behavior’ (Zuboff, 2017). Connecting a range of services – *advertising, personalization, predictive analytics, A/B testing* and other measuring devices – ‘these marketers aim to influence behavior at scale’ (Christl, 2017: 5). However, as Kylie Jarrett points

out, it is not only Google's domination of how 'commodified user behavior acts upon us' but, by drawing on Astrid Mager's scholarship (2012) of how the 'capitalist spirit gets embedded in search algorithms', she emphasizes that users instantiate and continue to purport Google and 'its advertising-based metrics' (2012: 19–20). As a platform of capitalism (Srnicek, 2016), Google influences the *trusted user* experience by creating detours in the path to information through *advertising*. Eradicating the boundaries between public and private, Google became a capitalistic 'omnipotent digital infrastructure' (Peters, 2015) instead of a public good (Introna and Nissenbaum, 2000: 177), whilst enacting 'this combined infrastructuralization of platforms and the platformization of infrastructures' (Plantin et al., 2018: 301).

Nowadays, Google's hegemony on finding information in the search engine market dominates the U.S. (two thirds) and well over 90% in most European countries (Lewandowski, 2017: 10), South America, Africa and India. Earning most of its revenue from search *advertising*, it is therefore not just a 'search engine' – Google is just the most lucrative service of its holding company, Alphabet. Ostensibly as an attempt to prevent anti-trust lawsuits and to defer fines, Google's 'metonymia' resulted in it quietly rebranding itself as a research company during August 2015, with the 'googlization of everything' becoming 'Alphabetized' from A-Z (<http://abc.xyz>) (Ridgway, 2021). Although Brin and Page imparted that Google is 'a research tool' (1998), Alphabet spends the surplus generated from *advertising* on research and development, investing in manifold projects at X-The Moonshot Factory ranging from Brain, Taara, Tapestry, Intrinsic to Waymo.

This vision is already apparent in Brin and Page's paper:

Another goal we have is to set up a Spacelab-like environment where researchers or even students can propose and do interesting experiments on our large-scale web data (1998).

With Google's merger of 'search' and 'research', Alphabet is presently focused on maximizing its profit through the creative use of its treasure chest of *usage data*, considered by some as the world's most valuable resource. As demonstrated through *smart algorithms* such as RankBrain, Google incorporates developments within the discipline of artificial intelligence, machine learning LLMs (large language models) (BERT, MUM, PaLM, GLaM, 1000 Languages Initiative, LaMDA), and most recently the chatbot Bard in Google search. This is made possible by massive amounts of *usage data* and energy usage as there is now enough computational power to process the data. Google continuously extracts and refines 'behavioral surplus' as a form of 'data colonialism' – an 'emerging order for the appropriation of human life' (Couldry and Mejias, 2019), with the interactions of [*trusted*] users

subsequently made useful to surveillance capitalism through a never-ending loop of prediction products (Zuboff, 2015).

Reimagining search

Returning to their Appendix A, Brin and Page state that 'it is crucial to have a competitive search engine which is transparent and in the academic realm' (1998). Already in 2000, Introna and Nissenbaum recommended that there should be public funding to develop transparency as well as access to 'more egalitarian and inclusive search mechanisms and for research into search and meta-search technologies' (181). Academics have attempted to reimagine search in response to Google's hegemony, suggesting a 'human knowledge project' (Vaidhyanathan, 2011) or an Imagine Search Engine that attempts to show how information could be presented differently (Noble, 2018: 180–181). Other researchers have put forth incorporating society-oriented design to 'democratise search' (Rieder, 2009) or argued that search engines could be 'good for democracy' (Sison and Sack, 2008), advocating autonomy and access to information without advertisements. Brin and Page closed their 1998 paper with the 'hope Google will be a resource for searchers and researchers all around the world and will spark the next generation of search engine technology'.

This enquiry examined six of Brin and Page's (1998) key innovations and concerns to explain the evolution of Google's hypertext search engine technologies through 'moments of contingency', resulting in subsequent lock-ins. Highlighting these decisions makes visible which intentions were already present in the early design affordances of Google search and how its auction *advertising* model and *usage data* instantiated a 'habit of automaticity' – ubiquitous googling – that provides 'prediction products' for surveillance capitalism: what Phil Agre might have deemed 'deleterious consequences' (1997).⁵ Instead of propagating techno-scientifically driven futures as unsettling 'predictive worlds' that encompass deviations in defining 'objects of governance, the development of instruments of intervention, and the constitution of political subjects', STIs (socio-technological imaginaries) employing 'co-production' are ways of thinking forward (Jasanoff, 2020: 32, 41). Sheila Jasanoff elaborates further to acknowledge STIs are 'a field of political action' and that they represent 'collective political imaginations', where human participation shapes these very imaginaries as an 'authentic politics of the future' (2020: 41). This connects to Safiya Noble's proposal that rather than advancing the neoliberal capitalist project of keeping search mechanisms opaque, the public, workers, librarians and institutions could come together to 'reconceptualize the design of the indexes of the web' to 'radically shift our ability to contextualize information' (2018: 133). Or, echoing Introna and Nissenbaum, a public index of the

web could be made accessible for diverse search engines, thereby creating a more diverse and pluralistic online searching ecology (Lewandowski, 2014, Open Search Foundation, 2019-ongoing). Above all else, it is important to reiterate that Google could have been designed and innovated otherwise and why the politics of search engines [still] matters.

Acknowledgements

The author would like to thank the guest editors (Astrid Mager, Ov Christian Norcel and Richard Rogers) and the three anonymous reviewers for their excellent reflections, insightful comments and constructive feedback, along with Peter Danholt and Matthew Zook for their additional support. Finally, the author thanks Shoshana Zuboff for kindly extending permission to allow the reproduction of her images by Karin Schwandt.


Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

ORCID iD

Renée Ridgway  <https://orcid.org/0000-0002-1483-9996>

Notes

1. Some of the arguments here are drawn from my PhD, specifically the Introduction, Chapter 3, 5, 9,10 and Conclusion.
2. The CTR (Click-Through-Rate) is the most important feature when determining the quality score and emotionally charged ads receive even higher CTRs. Additionally, Diane Tang's Keyword Pricing Index regulates the 'quality scores' of Adwords. For further information on Google's innovative auction models, see Steven Levy's WIRED article (2009).
3. 'In the case of Google (now Alphabet), A shares carry one vote, while B shares each carry 10 votes. Brin and Page between them own 51 percent of those B shares, giving them joint control of the company, even though they own less than 12 percent of its total shares' (Zetlin, 2022). Inc.com. Available here: <https://www.inc.com/minda-zetlin/warren-buffett-google-sergey-brin-larry-page-ipo-public-shares.html>
4. 'Google's chief economist, Hal Varian (2018), notes that only data that has been sold or licensed can be clearly identified and measured. He argues that personal data are not 'sold'; rather, access to personal data is "licensed" through contractual arrangements' (Birch et al., 2021: 4).
5. As Brin and Page were developing Google, coevally the computer scientist and humanities scholar Phil Agre published a seminal essay (1997) defining what he termed 'critical technical practice'. Although Agre addressed the development of AI, he elaborated on how, with each stage of research, critical insight enhances methods including feedback from the community, though not all members evaluate a research project with the same criteria. He speaks of the struggle 'against the patterns

of difficulty that are inherent in its design' and that 'the first step is to figure out what deleterious consequences those mistakes should have in practice' (Agre, 1997: 15). What if, instead, Brin and Page had applied a 'critical technical practice' in designing Google, which 'require[s] a split identity – one foot planted in the craft work of design and the other foot planted in the reflexive work of critique' (Agre, 1997: 15)?

References

- Agre P (1994) Surveillance and capture: two models of privacy. *The Information Society* 10(2): 101–127.
- Agre P (1997) Toward a critical technical practice: lessons learned in trying to reform AI. In: Bowker G, et al. (eds) *Bridging the Great Divide: Social Science, Technical Systems*, Available at: <http://polaris.gseis.ucla.edu/pagre/critical.html> (accessed 20 May 2022).
- Alaimo C and Kallinikos J (2019) Recommender system. In: Beyes T, Holt R and Pias C (eds) *Oxford Handbook of Media, Organisation and Technology*. Oxford: Oxford University Press.
- Battelle J (2006) *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*. New York: Penguin.
- Beer D (2009) Power through the algorithm? Participatory web cultures and the *technological unconscious*. *New Media & Society* 11(6): 985–1002.
- Bijker W and Law J (1992) Introduction. In: *Shaping Technologies/Building Society-Studies in Sociotechnical Change*. Cambridge, MA: MIT Press.
- Bilić P (2017) A critique of the political economy of algorithms: brief history of Google's technological rationality. *Westminster Advanced Studies* 5.
- Birch K, Cochrane D and Ward C (2021) Data as asset? The measurement, governance, and valuation of digital personal data by Big Tech. *Big Data & Society* 8(1): 205395172110173.
- Brin S and Page L (1998) *The Anatomy of a Large-Scale Hypertextual Web Search Engine*. Stanford, CA: Computer Science Department, Stanford University. Available at: <http://infolab.stanford.edu/pub/papers/google.pdf> (accessed 20 May 2022).
- Christl W (2017) Corporate surveillance in everyday life. How companies collect, combine, analyze, trade, and use personal data on billions. A Report by Cracked Labs, Vienna, June 2017.
- Chun WHK (2016) *Habitual New Media: Updating to Remain the Same*. Cambridge: MIT Press.
- Couldry N and Mejias U (2019) *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford: Stanford University Press.
- Elmer G (2003) *Profiling Machines: Mapping the Personal Information Economy*. Cambridge: MIT Press.
- Esposito E (2017) Algorithmic memory and the right to be forgotten on the web. *Big Data & Society* 4(1): 205395171770399.
- Feenberg A (2017) Critical theory of technology and STS. *Thesis Eleven* 138(1): 3–12.
- Feuz M, Fuller M and Stalder F (2011) Personal web searching in the age of semantic capitalism: diagnosing the mechanics of personalisation. *First Monday*, peer-reviewed journal on the

- internet. Volume 16, Number 2-7. Available at: <http://firstmonday.org/article/view/3344/2766> (accessed 20 May 2022).
- Franceschet M (2010) Pagerank: standing on the shoulders of giants. *Communications of the ACM* 54(6): 92–101.
- Gillespie T (2014) The relevance of algorithms. In: Gillespie T, Boczkowski PJ and Foot KA (eds) *Media Technologies. Essays on Communication, Materiality, and Society*. Cambridge, MA: MIT Press, pp.167–193.
- Gillespie T (2016) Algorithm. In: Peters B (ed) *Digital Keywords A Vocabulary of Information Society and Culture*. Princeton: Princeton University Press.
- Gitelman L and Jackson V (2013) Introduction. In: Gitelman L and Jackson V (eds) *Raw Data Is an Oxymoron*. Cambridge: MIT Press.
- Goriunova O (2019) The digital subject: people as data as persons. *Theory, Culture & Society* 36(6): 125–145.
- Greenacre P, Gross R and Speirs J (2012) Innovation theory: a review of the literature. *ICEPT Working Paper*, May 2012. Ref: ICEPT/WP/2012/011. Imperial College Centre for Energy Policy and Technology, London.
- Halavais A (2009) *Search Engine Society*. Cambridge: Polity Press.
- Hillis K, Petit M, et al. (2013) *The Culture of Search*. Abingdon-on-Thames: Routledge.
- Hjorth D (2007) Lessons from Iago: Narrating the event of entrepreneurship. *Journal of Business Venturing* 22(5): 712–732.
- Introna L and Nissenbaum H (2000) Shaping the web: why the politics of search engines matters. *The Information Society* 16(3): 169–185.
- Jarrett K (2014) A database of intention? In: König R and Rasch M (eds) *Society of the Query Reader #9: Reflections on Web Search*. Amsterdam: Institute of Network Cultures.
- Jasanoff S (2020) Imagined worlds: the politics of future-making in the twenty-first century. In: Wenger, et al. (eds) *The Politics and Science of Prevision-Governing and Probing the Future*. Abingdon-on-Thames: Routledge.
- Jiang M (2014) The business and politics of search engines: a comparative study of Baidu and Google's search results of Internet events in China. *New Media & Society* 16(2): 212–233.
- Kaplan F (2014) Linguistic capitalism and algorithmic mediation. *Representations* 127(1): 57–63.
- Langley P and Leyshon A (2016) Platform capitalism: the inter-mediation and capitalisation of digital economic circulation. *Finance and Society* 3(1): 11–31.
- Latour B (1990) Technology is society made durable. *The Sociological Review* 38(1_suppl): 103–131.
- Levy S (2009) Secret of Googlenomics: data-fueled recipe brews profitability. *WIRED*. May 22, 2009. Available at: <https://www.wired.com/2009/05/nep-googlenomics/> (accessed 20 May 2022).
- Lewandowski D (2014) Why we need an independent index of the web. In: König R and Rasch M (eds) *Society of the Query Reader #9: Reflections on Web Search*. Amsterdam: Institute of Network Cultures.
- Lewandowski D (2017) Users' understanding of search engine advertisements. *Journal of Information Science Theory and Practice* 5(4): 6–25.
- Lovink G (2009) Society of the query: the Googlization of our lives. In: Stalder F and Becker K (eds) *Deep Search: The Politics of Search Beyond Google*. Innsbruck: Studien Verlag.
- Lovink G and Tkacz N (2015) MoneyLab: sprouting new digital-economic forms. In: *Moneylab: An Intervention in Digital Economy*. Amsterdam: Institute of Networked Culture.
- Mager A (2012) Algorithmic ideology. *Information, Communication & Society* 15(5): 769–787.
- Mayer-Schönberger V and Cukier K (2013) *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. London: John Murray.
- Noble SU (2018) *Algorithms of Oppression, How Search Engines Reinforce Racism*. New York: New York University Press.
- Open Search Foundation (2019-ongoing) A free and open Internet search infrastructure: for our society and a prospering digital economy in Europe. White Paper. Available at: <https://opensearchfoundation.org/en/how>.
- Oremus W (2013) Google's big break. *Slate*. October 13, 2013. Available at: http://www.slate.com/articles/business/when_big_businesses_were_small/2013/10/google_s_big_break_how_bill_gross_goto_com_inspired_the_adwords_business.html (accessed 22 May 2022).
- Ørmen J (2016) Googling the news: Opportunities and challenges in studying news events through Google search. *Digital Journalism* 4(1): 107–124.
- Page L (2001) *Method for Node Ranking in a Hyperlinked Database*. U.S. Patent 6,285,999, filed January 9, 1998, and issued September 4, 2001
- Page L (2007) *Video of plenary lecture at the American Association for the Advancement of Science (AAAS) Annual Meeting*, February 16, 2007 in San Francisco, CA. Available at: https://www.google.com/press/podium_archive.html (accessed 16 April 2023).
- Page L, Brin S, Motwani R, et al. (1998) The PageRank citation ranking: bringing order to the web. Technical Report. Stanford InfoLab.
- Peters JD (2015) *The Marvelous Clouds: Toward a Philosophy of Elemental Media*. Chicago: University of Chicago Press.
- Plantin JC, Lagoze C, Edwards PN, et al. (2018) Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society* 20(1): 293–310.
- Ridgway R (2021) *Re:search: the personalised subject vs. the anonymous user*. Copenhagen Business School [PhD]. PhD Series No. 21.2021.
- Rieder B (2009) Democratizing search? From critique to society-oriented design. In: Stalder F and Becker K (eds) *Deep Search: The Politics of Search Beyond Google*. Innsbruck: Studien Verlag.
- Rieder B (2012) What is in PageRank? A historical and conceptual investigation of a recursive status index. *Computational Culture* 2. Available at: http://computationalculture.net/what_is_in_pagerank/ (accessed 22 May 2022).
- Rieder B and Sire G (2013) Conflicts of interest and incentives to bias: a microeconomic critique of Google's tangled position on the web. *New Media & Society* 16(2): 195–211.
- Rogers R (2009) The Googlization question and the inculpable engine. In: Stalder F and Becker K (eds) *Deep Search: The Politics of Search Beyond Google*. Innsbruck: StudienVerlag.
- Schumpeter JA (1942) *Capitalism, Socialism and Democracy*. 6 ed. New York: Routledge.

- Sison J and Sack W (2008) What makes a search engine good for democracy? PUBLIC OPINION POLLING AND THE EVALUATION OF SOFTWARE. Proceedings. Computer Science.
- Srnicek N (2016) *Platform Capitalism*. Cambridge: Polity Press.
- Statista (2022) Available at: <https://www.statista.com/statistics/266249/advertising-revenue-of-google/> (accessed 22 November 2022).
- Sullivan D (2016) FAQ: all about the Google RankBrain algorithm. *Search Engine Land*. June 23, 2016. Available at: <http://searchengineland.com/faq-all-about-the-new-google-rankbrain-algorithm-234440> (accessed 20 May 2022).
- Thylstrup N (2019) Data out of place: toxic traces and the politics of recycling. *Big Data & Society* 6(2): 205395171987547.
- Vaidhyanathan S (2011) *Googlization of Everything (And Why We Should Worry)*. Oakland: University of California Press.
- Van Couvering E (2010) *Search engine bias: the structuration of traffic on the World-Wide Web*. PhD Thesis. The London School of Economics and Political Science (LSE).
- Van Dijk J (2010) Search engines and the production of academic knowledge. *International Journal of Cultural Studies* 13: 574–592.
- Varian HR (2010) Computer mediated transactions. *American Economic Review* 100(2): 1–10.
- Varian HR (2014) Beyond big data. *Business Economics* 49(1): 27–31.
- Winner L (1980) Do artifacts have politics? *Daedalus* 109(1).
- Zetlin M (2022) Here's how warren buffett persuaded Google's reluctant founders to go public. *INC*. Available at: <https://www.inc.com/minda-zetlin/warren-buffett-google-sergey-brin-larry-page-ipo-public-shares.html>. (accessed 13 November 2022)
- Zuboff S (2015) Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology* 30: 75–89.
- Zuboff S (2016) Google as a fortune teller: the secrets of surveillance capitalism. *Frankfurter Allgemeine Zeitung*. Available at: <https://www.faz.net/aktuell/feuilleton/debatten/the-digital-debate/shoshana-zuboff-secrets-of-surveillance-capitalism-14103616.html> (accessed 20 May 2022).
- Zuboff S (2017) Secrets of surveillance capitalism. Lecture. Queens University, Canada. November 2017.
- Zuboff S (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: Public Affairs.