Re:search - the Personalised Subject vs. the Anonymous User

Ridgway, Renée

COPENHAGEN BUSINESS SCHOOL
SOLBJERG PLADS 3
DK-2000 FREDERIKSBERG
DANMARK

WWW.CBS.DK

ISSN 0906-6934

Print ISBN: 978-87-7568-018-4
Online ISBN: 978-87-7568-019-1

RE:SEARCH - THE PERSONALISED SUBJECT VS. THE ANONYMOUS USER

PhD Series 21.2021

Renée Ridgway

RE:SEARCH

THE PERSONALISED SUBJECT VS. THE ANONYMOUS USER

CBS PhD School | **PhD Series 21.2021**

CBS COPENHAGEN BUSINESS SCHOOL
HANDELSHØJSKOLEN

Re:search - the Personalised Subject vs. the Anonymous User

Renée Ridgway

Primary Supervisor:

Timon Beyes
Professor
Department of Management, Politics and Philosophy
Copenhagen Business School
and
Professor
Sociology of Organisation and Culture
Centre for Digital Cultures
Leuphana University

Secondary Supervisors:

Christian Borch
Professor
Department of Management, Politics and Philosophy
Copenhagen Business School

Nishant Shah
Chair Professor of Aesthetics and Cultures of Technology
ArtEZ University of the Arts/Radboud University, Faculty of Arts

Proofreading: David Selden
Danish translation: Ditte Vilstrup Holm

CBS Doctoral Studies
Copenhagen Business School
Frederiksberg, Denmark

Renée Ridgway
Re:search - the Personalised Subject vs. the Anonymous User

1st edition 2021
PhD Series 21.2021

ISSN 0906-6934

Print ISBN: 978-87-7568-018-4
Online ISBN: 978-87-7568-019-1

The CBS PhD School is an active and international research environment at Copenhagen Business School for PhD students working on theoretical and empirical research projects, including interdisciplinary ones, related to economics and the organisation and management of private businesses, as well as public and voluntary institutions, at business, industry and country level.

Acknowledgements

> The frontiers of a book are never clear-cut: beyond the title, the first lines, and the last full stop, beyond its internal configuration and its autonomous form, it is caught up in a system of references to other books, other texts, other sentences: it is a node within a network (Foucault 1972:23).

Books are structured through the citations of books before them. As they rearrange and reconstruct knowledge, new knowledge comes forth with every author's text that rests on the preceding citations of others and the labour of those before them. There is not only the organising structure of citation, as this thesis will demonstrate, but all the invisible knowledge, labour and help that went into its realisation.

I would like to thank, first and foremost, my primary supervisor Timon Beyes and my secondary supervisor Christian Borch. I am indebted to both for their ceaseless feedback, mentorship, guidance and patience. Thanks to colleagues for conversations, collegialities and convivialities at both CBS (MPP) and at the Digital Cultures Research Lab (DCRL), Leuphana Universität in Lüneburg, along with all the 'bureaucracy whisperers'— women behind the scenes who helped me negotiate and fathom mountains of red tape and paperwork. I would also like to extend my gratitude to my WIP (Work in Progress) 'opponents' for their critical insights that contributed to the development of this PhD, Anne-Christine Lange, Wendy Hui Kyong Chun, Nanna Bonde Thylstrup and especially Nishant Shah, for his never-ending coaching as well as caring. Faculty and colleagues at my PhD courses deserve thanks as well for their keen comments and critiques, including the publications and presentations organised by Aarhus University faculty (Geoff Cox, Christian Ulrik Andersen, Søren Bro Pold) at the numerous Transmediale festivals the past years. I would also like to extend thanks to my colleagues and participants at the yearly EGOS/LAMOS conferences and at the other conferences I was fortunate enough to be able to attend during my PhD.

Thanks also to the 'anonymous hackers', computer programmers and other people with technical skillsets who helped me understand and resolve my technical problems and especially to Morten Sune Nielsen, IT department CBS. The indispensable contributions of those colleagues at institutions who have also worked on 'search' (Institute of Networked Cultures in Amsterdam, World Institute in Vienna and the Centre for Internet and Society, Bangalore) and members of n.e.w.s: (https://northeastwestsouth.net), have provided a basis for this research.

This PhD would not have been possible without three-year financial support from the Digital Cultures Research Lab, funded by Volkswagen Foundation (Nieders. Vorab), along with a grant from the Mondriaan Foundation to produce my data visualisations with Richard Vijgen. I also received two fellowships from the Centre for Advanced Internet Studies (CAIS). I am thankful for their support along with all participants at the CAIS workshops and events I organised, which enabled me to refine and develop my research as well as to Michael Christen, for helping me understand the technicalities, alternatives and potential futures of search. I am eternally grateful to all of my family, friends, former students and colleagues throughout the world for their generosity, kindness and understanding the past years. Special thanks to Kirsten Dufour, Finn Andersen, Daniel Spikol and Lene Dalsjö Bull for *hjem og hygge* in Copenhagen.

'You can do this, but always remember to be kind to your reader' David Graeber (1961-2020).

Abstract

This thesis investigates how Google Search as a 'media a priori' organises (us)ers by first delving into how search worked in the past, engaging former European 'address offices' and human endeavours that attempted to 'organise the world's information'. It then explains how Google search developed during the last two decades, advancing an understanding that *Re:search* fuses two concepts: the Scientific Citation Index (SCI) for *research*, which in turn served as an inspiration for the PageRank of Google *Search*. Using my office at CBS as a site of data collection, I designed and carried out an 'experiment in living', searching with Google as the 'Personalised Subject' and with Tor as the 'Anonymous User', with the same set of chosen keywords. Whilst conducting 'interviews' with algorithms—invisible interlocutors—I collected data on myself and produced *Re:search - Terms of Art*. These 'data visualisations as transcription' reflect my search results based on 'locative data' (Google) or 'off the map' (Tor), and these 'critical cartographies' as practices of representation seek to intervene and give shape to the world by making invisible infrastructures more tangible.

Drawing on my methods I demonstrate how advertisement affects the ranking of search results and question the marketing of 'personalisation' as authenticity, along with showing how unique results are determined by signals that comprise its proprietary algorithm—the machine-learning RankBrain, which enables its authorship. The study then 'reimagines search' by exploring the boundaries of anonymity online through ethnographic studies and the search engines of the Dark Net, along with the p2p technologies (encryption) that enable it, such as Tor. Applying the IP (internet protocol) address as an organisational hinge and by way of a comparative analysis and a diagram, the effects of search engines on (us)ers are structured into 'collaborative collectives'––'subjectivities of search' and 'agencies of anonymity'—according to degrees of human-algorithmic interaction. After revealing data profiling and collaborative filtering technologies, I then elucidate how Google Search organises (us)ers, facilitated by the social constellation of 'surveillance capitalism', with its extraction of behavioural data and selling of prediction products.

The thesis builds upon findings of how digital media are habitual, enacting behaviours in (us)ers with 'ubiquitous googling' of omnipotent platforms, which advances recent research on the epistemological and political challenges of 'mediality'. The analysis and discussion additionally contribute to the technological condition of the 'media arcane'—how human algorithmic interaction, or 'cyberorganization' is an invisible and 'intransparent' process. Furthermore, it expands the debate on reimagining search, merging media theory with the work of privacy and anonymity scholars as well as encryption techniques and practices of intervention through human agency. Lastly, I introduce an interdisciplinary methodological framework that contributes to the project of understanding (Post)Digital Cultures through prescriptive, inscriptive and transcriptive technologies, situated within three disciplines: organisation studies, media theory and artistic research.

Abstract (på dansk)

Denne afhandling undersøger, hvordan Google Search som et 'media a priori' organiserer os/brugere ved først at dykke ned i, hvordan søgninger plejede at fungere i tidligere europæiske 'adressekontorer' og i kraft af menneskelige anstrengelser, som forsøgte at 'organisere verdens information'. Den forklarer dernæst, hvordan Google-søgninger udvikledes igennem de sidste to årtier og udvikler den forståelse, at *Re:search* (*Forskning/Søgning*) er en organisationsmåde og en autoritet, der forbinder to begreber: det videnskabelige Scientific Citation Index (SCI) for *research* (*forskning*), som i sin tur tjente som inspiration for PageRank i Google *Search* (*søgning*). Jeg brugte mit kontor på CBS som sted for dataindsamling og designede og udførte et 'experiment in living', hvor jeg søgte på de samme udvalgte søgeord med Google som det 'Personalised Subject' og med Tor som den 'Anonymous User'. Mens jeg således udførte 'interviews' med algoritmer – usynlige samtalepartnere – indsamlede jeg data om mig selv og producerede *Re:search - Terms of Art*. Disse 'data visualisations as transcription' afspejler mine søgeresultater baseret på 'lokativ data' (Google) eller 'uden for kortet' (Tor), og som repræsentationspraksisser søger disse 'critical cartographies' ikke bare at intervenere i verden, men også at give form til den ved at gøre usynlige infrastrukturer mere håndgribelige.

Jeg trækker på mine metoder for at demonstrere, hvordan reklame påvirker rangeringen af søgeresultater og anfægter hermed markedsføringen af 'personalisering' som autenticitet, samtidig med at jeg viser, hvordan unikke resultater er bestemt af signaler, der indeholder Googles beskyttede algoritmer—den maskinlærende RankBrain, som understøtter dens forfatterskab. Derefter 'nyfortolkes søgning' ved at undersøge grænserne for online-anonymitet gennem etnografiske studier og søgemaskinerne for det Mørke Net, sammen med p2p-teknologier (kryptering), som understøtter det, som for eksempel Tor. Ved at bruge IP (internet protokol) adresse som en organisatorisk hængsel og gennem en komparativ analyse og et diagram, struktureres effekterne af søgemaskinerne på os/brugere i 'collaborative collectives'—'subjectivities of search' og 'agencies of anonymity'—ifølge graderne af menneske-algoritme interaktion. Efter at have afsløret dataprofilering og kollaborative filtreringsteknologier forklarer jeg, hvordan Google Search organiserer os/brugere, faciliteret af 'overvågningskapitalismens' sociale konstellationer med dens ekstrahering af adfærdsdata og salg af forudsigelsesprodukter.

Afhandlingen bygger på resultater for, hvordan digitale medier er vanebundne og bestemmende for adfærd i os/brugere med 'ubiquitous googling' på almægtige platforme, som avancerer nyere forskning om de epistemologiske og politiske udfordringer ved 'mediality'. Analysen og diskussionen bidrager også til den teknologiske betingelse for 'media arcane' – hvordan menneske-algoritmisk interaktion, eller 'cyberorganisation' er en usynlig og 'intransparent' proces. Dertil udvider den debatten om at 'nytænke søgning' ved at forbinde medieteori med privatheds- og anonymitetsforskere, såvel som krypteringsteknik og interventionspraksisser gennem menneskelig agens. Endelig introducerer jeg et interdisciplinært metodologisk rammeværk, som bidrager til udfordringen med at forstå (Post)Digital Cultures gennem 'præskriptive', 'inskriptive' og 'transkriptive' teknologier, der er situeret imellem tre discipliner: organisationsstudier, medieteori og kunstnerisk forskning.

Table of Contents

Prologue

Can you comment on your relationship with the search engine Pollyhop?

Is it true that it is providing you with voter data?

'Imagine a duel. Conway has a powerful gun, a search engine'.

'He can tell what you think, what you want, where you are and who you are. He can turn all of those searches into votes and that is enough bullets to kill my chances of winning. But I have an even bigger gun. It is called the NSA and one of the perks of being president. That is if the courts allow my surveillance request'.

'Imagine the men hanging on these walls wish they had a gun like that available to them. Your phone, the phone of the person sitting next to you, your neighbors phone and everyone you know and the 300 million Americans you don't know. I can see you. And I can use what I see to rig this election. Now of course a weapon like that, well, you can imagine how risky it is'.[1]

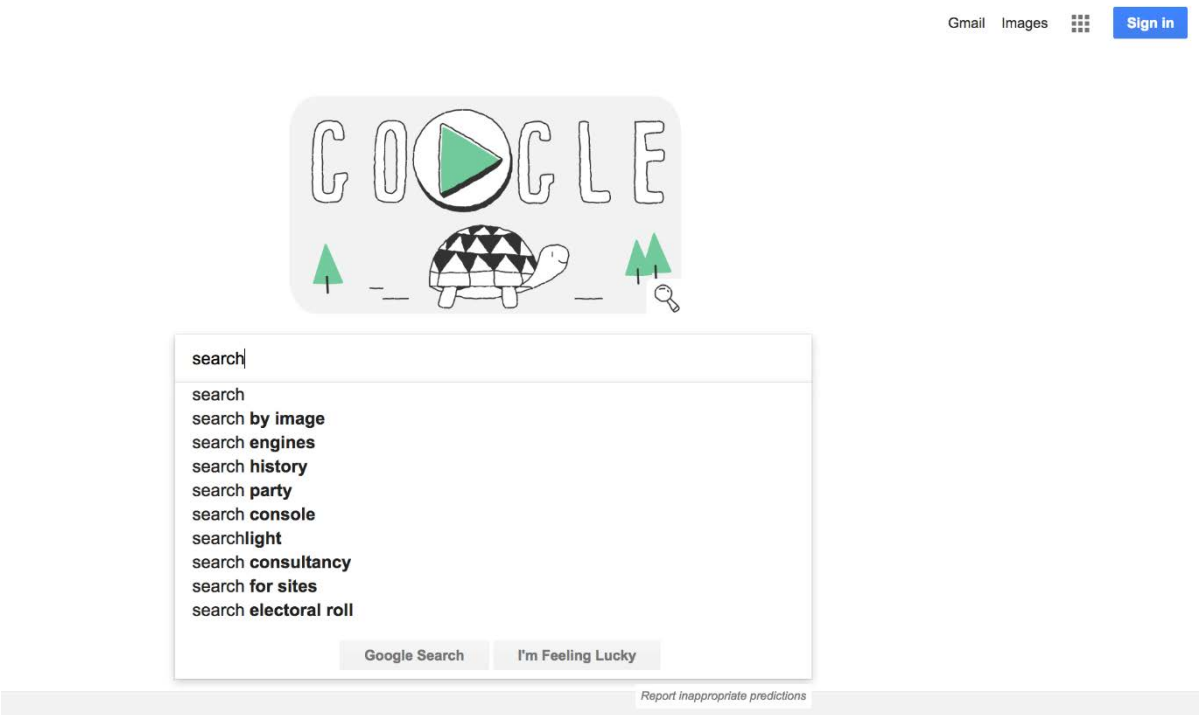


Figure 1: Google search

[1] *House of Cards*. Season 4: Episode 7. Timecode: 16:00
Spoken by character President Frank Underwood. Air date: March 4, 2016.

In recent years massive amounts of data has been indiscriminately collected on individuals worldwide. In June 2013 the 'Snowden revelations' exposed the governmental and military surveillance complex that spies on US citizens, unbeknownst to them and without search warrants. The NSA's secret PRISM programme included the 'monitoring of emails, file transfers, photos, videos and chats,' along with its ability to 'watch your ideas form as you type' with the 'live surveillance of search terms' (Gallagher 2013). (Figure 1) Because of the evidence released by Snowden, public awareness about global surveillance has changed not only 'people's perceptions of technologies' which in turn have an effect on their concerns regarding privacy, 'but also their perceptions of the organizations that deploy the technologies' (Forte et al. 2017:3). With the 'Snowden Effect' (Rosen 2015), the general public has not only been made aware that their privacy is being compromised but they have become more technically savvy, exemplified by the increased use of encryption technologies and anonymising browsers for searching the web such as Tor (The Onion Router) (Gehl 2014:5 citing Borland 2013).

The revelations also divulged that internet traffic is directly siphoned from underwater 'international cables, routers and switches' by governments with the US admitting that it 'collects foreign intelligence—just as many other governments do—to enhance the security of our citizens and protect our interests and those of our allies around the world' (Gallagher 2013). Partially based on Snowden's revelations regarding PRISM's mass surveillance programme that were not in compliance with EU data protection laws, in 2013 privacy activist Max Schrems lodged a complaint concerning EU data privacy restrictions. On October 6, 2015 he won a landmark decision at the European Court of Justice with his lawsuit *Schrem vs. Data Protection Authority*. The decision invalidated the much-used Safe Harbour agreement with which Silicon Valley companies were able to receive transfers of personal data from European citizens for data processing, such as online searches and social media usage.

However, the data continues to flow. Querying with keywords and clicking on hyperlinks, online search is a process of human interaction with computers and hidden protocols that enable connectivity and the navigation of the web. Already in 2012, over 90 percent of American internet users in all age groups up to 65 relied on search engines to retrieve information online (Purcell et al. 2012a cited by Trevisan 2014). 'Survey results from 2012 indicate that 54% of American adults use a search engine every day' (Mulligan and Griffin 2018:573). Like many users who frequently employ search engines for information regarding businesses, medical advice or their own rankings, people use Google Search to find answers to their concerns. The content of queries is captured by Google Trends, a so-called 'public web facility' provided by Google, which is based on Google Search results and reflects how often a keyword, search term or phrase, is entered in the search box.

What 'trends' offers is a view into *how* people are thinking at a given moment in time with a sudden surge of interest in a topic, reflected by a 'spike' that comprises 'understanding relative search interest in the topic compared to itself' (Rogers 2016). For example, although the UK referendum on June 23, 2016 resulted in the Brexit, the question *remains* whether a well-informed public headed to the polls because the search phrases 'What does it mean to leave the EU?' and 'What is the EU?' occurred *after* the polls were closed and predictions of the outcome surfaced in the media. (Figure 2) It then became apparent that people were wondering what they actually had just voted for, if they had voted.

Figure 2: Google Trends Twitter June 24, 2016

Moreover, people increasingly consult Google for historical information with expectations of receiving facts, yet sometimes autocomplete proposes racist results instead, causing complaints. Journalist Carolyn Calladawr received 'did the holocaust happen' as a result and 'reportedly discovered this serendipitously through the search prediction, or the suggestion Google offered to complete for her partially formed query "did the hol" into the search box' (Mulligan and Griffin 2018:559). Confronted afterwards in December 2016 by Calladawr among others, Google nowadays acknowledges the problematics of 'predictive search' yet it is not clear what interventions it performs on a daily basis. Consider, too, how the intention to carry out murders of innocent people might be influenced by autocomplete and biased search results as shown by the testimony of Dylann Roof, who was convicted of a federal hate crime—the murder of nine black church members in Charleston, South Carolina, U.S in June 2015. Typing into Google Search the query 'black on White crime' (Hersher 2017), Roof states that he has 'never been the same since that day' and that 'the first website I came to' was the Council of Conservative Citizens, a white supremacist organization (ibid). Although Roof's manifesto was captured by the Internet Archive's *WayBackMachine,* Google does not make its database of past search suggestions or search results public.

Nor does it remove results, unless legally forced to, because according to Google 'search is a reflection of the content and information that is available on the internet' (ibid). With the EU Commission's ruling in 2014 concerning the 'right to be forgotten' of individuals' search results, EU citizens now have the option of requesting that Google delete information about themselves that they deem embarrassing, even if this information is true (Thylstrup 2014:35).

(Figure 3) '[I]n the face of leaking machines that seem to remember forever', users carry responsibility for what shows up about them in search results, needing to prove that 'their privacy needs are greater than the public's right to know' (ibid:36).

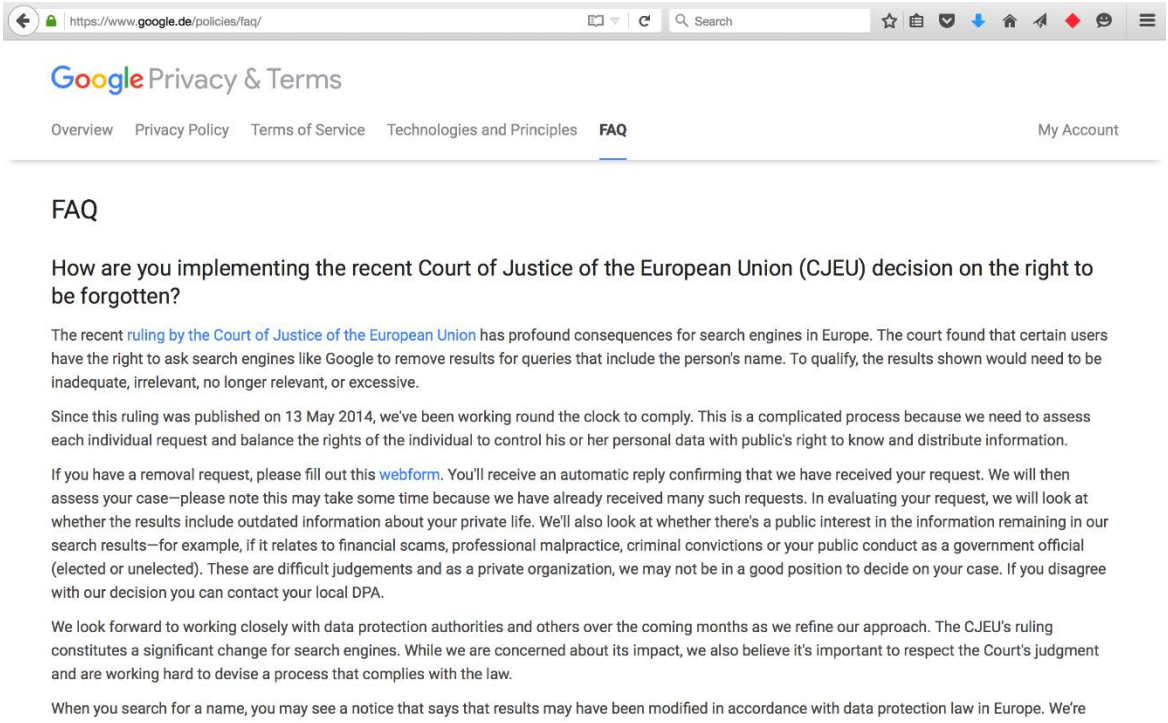


Figure 3: Google's response to *Right to be Forgotten* ruling

The very nature of what constitutes democratic participation in elections (voting) is reproduced not only by the queries of citizens, but by what news they 'like' or what interests them when they click on search results. In the lead up to the US elections in 2016, 'clickbait' (clicking on links to increase traffic and distribution) became prominent with the dissemination of 'fake news'. Search engine queries can also contribute to the outcomes of legal cases by showing the intentions of purported suspects and as evidential timelines. In late 2015 the NSA was able to penetrate high ranking Russian intelligence officers' mobile devices and learned news of a planned hacking operation, confirmed by the fact that the 'Russians search[ed] the internet for any news about the oncoming attack' (Modderkolk 2018). Evidence provided by the Dutch secret service (AIVD) showed that the Russian hacking team Fancy Bear entered the search query phrase 'company's competence'—the same phrase *before* it appeared on the inaugural blog post by 'Guccifer 2.0'—who claimed to be a lone Romanian hacker (Graff 2018). Or in another example, 'after the mass shooting in Las Vegas, the top search result linked Devin Kelly to ISIS and connected users to 4chan' (McKay 2017 cited by Tripodi 2017:33). This shows Google delivering and spreading propaganda and disinformation after acts of terror and mass shootings (ibid).

The *Right to be Forgotten* decision (2014), the Roof federal hate crime (2015), the UK referendum results in Trends (2016), 'predictive search' in regard to historical and criminal events (2016), the 'clickbaiting' tactics of malevolent actors manipulating US elections (2016) and disinformation dissemination (2017) all indicate how Google increasingly 'curates' (Groys 2013) information back to the public. In these contexts, '[s]earch engines no longer merely shape public understanding and access to the content of the World Wide Web' (Mulligan and

Griffin 2018:557). Google is an apparatus that continually shapes public understanding in contemporary society (ibid) and it has the power to remove/censor parts of the index or manipulate the ranking (Hermann et al. 2014:1). Although Snowden's revelations have opened users' eyes to the Five Eyes and state surveillance of citizens, Google's proprietary IP black box remains closed. These search results are opaquely created as Google collects a user's habits, interests and data with every search query. Increasing awareness of this surveillance also inhibits users' exploration of knowledge, as Snowden relates in the documentary *Citizen Four* (2014) by Laura Poitras:

> Many people I have talked to have mentioned that they are careful about what they type into search engines because they know it's being recorded and that limits the boundaries of their intellectual exploration.

The massive collection of (meta)data, specifically search queries, is constant and undertaken without specific permission as people consult 'the oracle' Google, not only for answers to their questions but in search of knowledge or 'truth'. Questions about who controls data—search data—and whether it's an asset (intangible, tangible) are major concerns in the 21st century. Is data an extension of the self, or is it the 'new oil economy' or can it be both? While corporations desire user data, do users need to have agency to be able to decide what to share and to be in control of their data when they search? Besides investigating civic and commercial interactions with search engines, researchers also need to formulate a response to these sweeping societal changes and what this means for democracy. Thus, closer scrutiny is called for in order to understand the workings of Google Search and to reimagine search in a surveillance society, as this PhD attempts to do.

Introduction

1.0 Why the politics of search engines (still) matters

In the 1990s early net programmers and users envisioned a 'digital democracy' with their 'bulletin board' postings, chat rooms and networks, instilling a belief that searching the Web wasn't only about information retrieval but knowledge exploration. In 1992 Mark Perry McCahill introduced the term 'surfing' to refer to this act and coevally Jean Armour Polly popularised it with her book *Surfing the Internet* (1992).[2] 'Surfing'—clicking on hyperlinks and not knowing where these would lead—was the modus operandi for many users and part of this process of exploration was finding things one would not have anticipated. Those discoveries of disparate or unrelated phenomena relied on serendipity, which is the ability to come across books, articles, images, information, objects and so forth, by chance.[3] As the web grew so did the amount of information, accruing into millions of documents that necessitated organisation in order to be retrieved. Concomitantly, the question of how to navigate such a space in order to find what users were seeking was answered by the new technologies of search (Stark 2009:1).

During the early days of the web various search engines competed for users' attention and how this 'information superhighway' might be navigated and explored. Two aspects organise searching on the web, one is the preference of those searching and the other are those parties, or websites who desire to be found, with 'enormous inequality' between these two forces (Introna and Nissenbaum 2000:177). Written at the dawn of the millennium, *Shaping the Web:Why the Politics of Search Engines Matters* expressed concern with 'the evident tendency of many of the leading search engines to give prominence to popular, wealthy, and powerful sites at the expense of others' (ibid:181). Introna's and Nissenbaum's compelling argument was for the promotion of the web where a plurality of voices would be heard, warning: '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' (ibid:180).

While Introna and Nissenbaum's portrayal of search engines was as a public good—a fact-finding benevolent apparatus that informs the public with neutral results—commercial models were already infiltrating the web. Search engines regulate 'locating, organizing and distributing information and knowledge' (Jiang 2014:212) and the need for finding things in an overload of information, where attention is an intrinsically scarce resource, became of utmost importance. 'In an attention economy, the search engine is the ultimate aggregator of such wealth, and advertisers are the clearest source of revenue' (Halavais 2009:78). As technology advanced, adverts have become more effective in receiving attention and manipulating the user's response, their click.

---

[2] 'I wanted something that expressed the fun I had using the Internet, as well as hit on the skill, and yes, endurance necessary to use it well. I also needed something that would evoke a sense of randomness, chaos, and even danger' (Polly 1992). Polly's quote must have been in the *House of Cards* scriptwriter's mind. See Prologue.

[3] The accidental, or happenstance of coming across things in libraries seems outdated. If one is carrying out research online, how does one find the 'book next to' what was originally looked for? Would this be on the first page of Google search results or on the 99th page?

Thus, over time, Google transformed itself into an advertising company 'whose products are free of charge to users' (Jiang 2014:212), delivering search results and capitalising on the 'informational rationality of generating value from advertising and audience labour' (Smythe 1981; Jhally and Livant 1986 cited by Bilic 2017:8). Battelle (2003) and Salkever (2003) first dubbed this dominance over information exchange and web commerce the 'phenomenon of googlization' as it altered the media and technology landscape. Later on, it was Google's 'creep' into other major industries, 'advertising, software applications, geographic services, e-mail, publishing, and Web commerce itself' that Siva Vaidhyanathan characterized in his book *The Googlization of Everything (and why we should worry)* (2011: 20).[4] This googlization established new social, cultural and political logics of search-based information societies and its economies (Lovink 2013). As of December 2020, there are around 1,8 billion websites connected to the internet globally, with around 200 million active and Google.com is the 'world's most popular website' (Zuboff 2015:77).

Yet the implications and consequences of this hegemony in regard to what information is returned and the effects of receiving 'relevant' search results on users begs further investigation. Over the past twenty years Google has become the sine qua non that organises and enables access to information by providing users with a range of services, yet the ways in which 'media organise' (Martin 2019) define the practices of human organisational life. Google's 'search engine' is part of a larger 'media ecosystem' comprised of various actors (human/non-human) that control what the user queries and receives in return, however these are neither neutral nor unbiased. This human/algorithmic interaction, specifically querying and clicking on hyperlinks, has life-changing effects. Google prioritises information that sources certain topics and bolsters its own interests. Besides Google's hegemony on search, questions of algorithmic bias (O'Neil 2016), equality and the notion of 'truth' regarding the dissemination of information (Noble 2018) in a (post)digital society are also at stake. With contemporary discourse over control of information in the public sphere more relevant than ever, the politics of search engines [still] matters, as they play a preeminent role in determining whether media enables 'democratizing forces or [is] to be colonized by specialized interests at the expense of the public good' (Introna and Nissenbaum 2000:170).

2.0 Research interest

Upon reading Eli Pariser's *The Filter Bubble* (2011), I decided to investigate if and how I was being personalised while using Google Search since 2009, forming the departure point for this thesis (2014). I also became aware of the 'corporatisation' of my results when conducting research, however my focus was not on the company. John Durham Peters mentions in his chapter, *God and Google,* that there has been a broad selection of journalistic as well as scholarly texts written about Google in the past 13 years.

> John Battelle, *The Search* (New York: Penguin, 2005); Ken Auletta, *Googled: The End of the World as We Know It* (New York: Penguin, 2009, 2010); Jeff Jarvis, *What Would Google Do?* (New York: Collins, 2009); Ken Hillis, Michael Petit, and Kylie Jarrett, *Google and the Culture of Search* (New York: Routledge, 2012); Nicholas Carr, *The*

[4] One of the first critical investigations of 'googlization', Vaidhyanathan attempted to answer three questions in his book: 'What does the world look like through the lens of Google? How is Google's ubiquity affecting the production and dissemination of knowledge? How has the corporation altered the rules and practices that govern other companies, institutions, and states?' (Vaidhyanathan 2011).

> *Shallows* (New York: Norton, 2010); and Steven Levy, *In the Plex* (New York: Simon and Schuster, 2011) (Peters 2015:324).

Subsequently, there are entire areas of research that are focused on the power of Google and 'googlization' (Battelle 2003; Salkever 2003; Elmer 2004; Lyon 2007; Rogers 2013; Vaidhyanathan 2011). *The Dark Side of Google* (Ippolita 2013) critiques Google's business model in the 'datauniverse' and *Fuck Off Google* by the Invisible Committee (2014) shows Google to be an explicitly political project.

Deep Search: The Politics of Search Beyond Google (Becker and Stalder 2009) is an anthology that addresses Google as a search paradigm and investigates its social and political dimensions along with personalisation, PageRank and legislation issues. Elisabeth van Couvering's thesis, *Search engine bias: the structuration of traffic on the World-Wide Web* (2010), provides an overview of the historical development of search engines and is focused on bias. Other researchers have explained the problematics of hidden search algorithms, whether that be for commercial gain (Kaplan 2014), dissemination of news and information as 'filter bubbles' (Pariser 2011) or trade secrets (Pasquale 2015). Further scholarship has shown the 'relevance of algorithms' (Gillespie 2014), algorithmic 'visibility management' (Flyverbom et al. 2016), the 'algorithmic ideology' of Google Search (Mager 2013) and how search algorithms are oppressive and discriminatory (Noble 2018:28).

In the context of search engines as academic inquiry tools (Trevisan 2014), I thought it was crucial as a researcher to engage in a wide-ranging discussion about the challenges involved in search engines (or search methods) as objects of research *and* tools of inquiry. Google keeps its search algorithms a closely guarded secret and an entire industry (Search Engine Optimization-SEO) has been built around second-guessing them. Therefore, methodologically, it is difficult to collect search data in an environment that is in constant flux although there have been a few empirical studies that successfully do so (Feuz et al. 2011; Jiang 2014; Noble 2018). Due to this lacuna in regard to the protocols and organising properties of media technologies such as search engines, theoretical as well as empirical research on search seemed timely and essential. Taking up Jacob Ormen's call to 'document the development of search' in preserving culture (2013:189), I decided to structure my investigation by applying Google's own motto, 'organising the world's information and making it accessible and useful' to (re)search.

3.0 Aims, objectives and research questions

In the opening paragraph of the first chapter of his book, *The Sense of Dissonance*, the sociologist David Stark declares that 'search is the watchword of the information age' and by typing 'a few keywords at the toolbar, we can access enormous databases' (2009:1). Drawing on John Dewey's 'open ended inquiry', which concerns identifying the problem instead of problem solving, Stark elaborates that there are degrees of search but also 'challenging situations' — where 'you must search even though you do not know what you are looking for' (Stark 2009:2). For him, the word 'research' is the difference between 'searching for the already known' and the process of exploration, '[b]ecause, at some level, science is not about making truths but about opening up the *terrae incognita*, the inquiring posture of a good ethnographer and a good quantitative analysis are not so very different' (2010 Stark and Harrington).[5]

[5] The first usage of the term is by Ptolemy in around AD 150 in *Geographica*, a treatise on cartography that reflected the known world at that time, whereas in later centuries, colonial mapmakers often applied the term to

When I began this PhD as a social scientist researcher who inquires—in the Deweyian sense—I did not know what I was looking for, nor what I would find with search engines. Yet I had identified the research problem—I was disappointed in Google Search because I was in search of online serendipity. During the course of this research on search, search engines became my object of study as a means of organisation *and* inspiration when I discovered how search and research fuse: the Scientific Citation Index (SCI) for *research*, which is the inspiration for the PageRank *search* algorithm. Based on my etymological enquiry this is what the term 're:search' represents for me—search (engines) produce research about search engines by the very technology that facilitates the enquiry. Therefore *Re:search – the Personalised Subject vs. the Anonymous User* investigates search methods, ranging from search's prehistories via the 'personalisation' of Google to developing alternatives with Tor (The Onion Router) for querying the web. My aim and primary research question was to find out *how Google Search organises (us)ers*? In order to respond to my main research question, I first needed to find answers to corresponding subquestions, which informed and structured my enquiry along the way.

How did search happen in the past? explores the early modern 'address offices' that collected queries from clients and delivered results, whilst registering the physical addresses (data) of people in ledgers so that they could be identified and found. Furthermore, I investigate the analogue and digital attempts to *organise the world's information and make it accessible and useful*. In order to understand *how Google search works* I then unpack the two sets of meanings implicit in 'Re:search': research and search through *A Media Archaeology of Citation* that shows antecedents and the first ten years of PageRank and, with *The Personalised Subject*, the years 2009-2016 with its shift to RankBrain.[6] The sub question of *how can search be reimagined?* calls for the exploration of *Worlds of privacy* and anonymity by applying Tor and employing DarkNet search engines to explore 'non-indexed worlds' as *The Anonymous User*.[7] Drawing on these empirical studies I then set out to understand *what are the effects of search engines on (us)ers?* by way of a comparative analysis, *Black Box vs. Black Bloc.* On this basis, I set out to answer my main research question, *how Google Search organises (us)ers* by investigating *Black Utopia: Surveillance Capitalism.*

4.0 Methods assemblage

Most employees at Google (past and present) sign non-disclosure agreements, so I couldn't carry out interviews with these specialists in the field. Therefore, I opted for a method assemblage (Law 2004) comprised of three specific methodologies: a 'critical ethnography of the self', an 'experiment in living' and 'data visualisation as transcription'. With my 'critical ethnography of the self' (Wang 2008), I let myself be personalised and gathered data on myself. My self-designed empirical 'experiment in living' (Marres 2012) in my office at Copenhagen Business School enabled me to capture two forms of address when searching online—one as a

justify the eradication of indigenous peoples, because it was an 'unknown land' —not claimed by another colonial power. I use the term metaphorically to express an unknown field of research.

[6] The Oxford dictionary states that 'search' means 'try[ing] to find something by looking or otherwise seeking carefully and thoroughly.' Moreover, 'search' comes from the Old French 'cercier' or modern French 'chercher' that are derived from the Latin circare, meaning to wander or traverse and shares the same etymology with 'circle' and 'circus'. Dave Eggers' book, *The Circle,* is a parody on Google and makes appearances throughout this thesis.

[7] The verb search also means to 'examine (a place, vehicle, or person) thoroughly in order to find something or someone' and includes an investigation or to carry out an examination of records and to search through an area or place. Chapter 7 addresses legal rights regarding the 4th Amendment of the US constitution's 'search and seizure' statute and the 'expectation of privacy' when using Tor.

personalised subject with my IP address recognised by Google contrasted by being anonymous online with Tor, where my IP address is hidden. I carried out 'interviews with algorithms' and produced 'data visualisations as transcriptions' that enabled me to reflect on my results (hyperlinks). These three methods, combined and intertwined, facilitated my understanding of the behind-the-scenes constellations of agents, protocols, algorithms and myself that determined my search results.

This methods assemblage also includes 'novel territories'—citations from works by contemporary authors of fiction (Thomas Pynchon and Dave Eggers), cinema (Ex Machina) or syndicated television series (House of Cards, Mr Robot) that allude to or reference search engines. Furthermore, during my research I came across works of art, installations, interventions, netart, postmedia, postinternet and postdigital practices that provided insight into search engines through visual or interactive means. Beside academic literature on search engines, the crosspollination of media between newspapers and their digital imprints, along with videos of seminars, lectures and conferences further augment my research.

5.0 Contributions

In disciplinary terms, the thesis is positioned at the interstices of organisation studies, media theory and artistic research, taking elements from these three fields of knowledge that inform my undertaking.

This PhD research contributes to the confluences of media, technology and organisation studies (Beyes, Holt and Pias 2019) because search engines are media and devices of organisation (Beyes, Conrad and Martin 2019) and the Google corporation the contemporary 'media a priori' (Peters 2015:9). It also builds upon findings of how digital media are habitual (Chun 2016), by enacting behaviours in (us)ers such as 'ubiquitous googling' (Ridgway 2021), which advances recent research on the epistemological and political challenges of 'mediality' (Beverungen, Beyes and Conrad 2019). As denoted in the publication *The organizational powers of (digital) media*, mediality 'serves to reflect on the material and technological conditioning and structuring of experience, agency and interaction' (ibid:624) yet this distributed organising is 'mostly invisible and intransparent' (Beyes and Pias 2019; Hansen and Flyverbom 2015). I add to media theory by building upon recent publications that provide fresh insights into the media arcane (Beyes and Pias 2019) and my 'experiment in living' that engaged with two modes of address and attempted to visualise the black box.

I demonstrate a historical overview of analogue and digital search engines in Western culture in past centuries, advancing Rieder's 'archaeology of citation' (2012) on the 20-year development of PageRank by adding how RankBrain is a continuation of this 'media archaeology' (Parikka 2012; Ernst 2013). I also seek to make a minor contribution to media genealogy—the way in which history is inscribed in media or materials and bodies (Kittler 1999) by constructing a lineage between older and newer media—from the use of print media (registers, journals, posters) at the address offices, to paper machines (Krajewski 2011) to contemporary 'hupomnemata' (Foucault 1983)—computers, search engines and databases. Furthermore, I contribute to the discussion on reimagining search, merging media theory with the work of privacy and anonymity scholars (Marx 1999, Nissenbaum 1999, 2015; Brunton 2015; Sweeney 2004; Forte, Andalibi, Greenstadt 2017), along with encryption researchers (Chaum, Dingledine) who have also informed my understanding of these techniques and practices.

Such interdisciplinarity perhaps requires some further unpacking. The recent publication *Organize* expounds upon Kittler's remark (1999) that 'media determine our situation' because 'media organize' (Martin 2019), yet they also 'condition life through organizational effects' (Beyes, Conrad, Martin 2019) such as with personalisation and my method 'critical ethnography of the self'. Media, such as search engines, are 'predicated on organizational constellations' that recursively reflect how archives or databases have been structured and these technological assemblages are 'fields of knowledge and social institutions' (Horn 2007:8 cited by Beyes, Conrad, Martin 2019). Therefore, I build upon this discourse by (re)introducing the term 'cyberorganization'—'information as becoming' (Parker and Cooper 2016) and by elucidating how Google Search organises (us)ers through its databases of intentions, facilitated by the social constellation of 'surveillance capitalism' (Zuboff 2015, 2017).

Yet the intimate relation of media and organisation is not new. The media historian Tantner shows how the address offices constitute the prehistory of search engines, where clients' physical addresses were written down in registers organising information and creating profiles of their clients (2015). I build upon his scholarship by elucidating how the users' contemporary IP address functions as an organisational protocol that on the one hand identifies the user to Google and on the other, enables anonymity with Tor hiding users' IP address. In doing so I seek to expand Galloway's research on *Protocol* (2005) and *Black Box, Black Bloc* (2010) by making a comparison between the hidden 'black box' workings of Google and the obfuscation tactics of the 'black bloc' with Tor. I yield insight by drawing on previous empirical studies and a range of theoretical texts by media and privacy scholars to speculate on the effects of Google's computational agency with my 'subjectivities of search' and human 'agencies of anonymity' (Tor, protocol, programming) through intervention.

Artistic Research enabled me to transport methods from the field of contemporary art into the emerging field of Organisation, Technology and Media, as part of my interdisciplinary methods assemblage (Law 2004). I refer here to (visual) artistic research, which 'spawns all kinds of knowledge' as 'epistemic engines' (Maharaj 2009). I adopt this kind of epistemic strategy by (re)territorialising contemporary vocabularies in art discourse as keywords with which I carry out these empirical search experiments and as artworks themselves, as with my *Re:search - Terms of Art*. Media theorists in the past (Manovich 2001) have also looked to artistic practices of cinema or 'cognitive mapping' as a Marxist aesthetic (Jameson 1998), nowadays often described as 'information aesthetics' (Galloway 2011).

I seek to draw upon these 'Critical Cartographies' with my method 'data visualisation as transcription' —a means to visualise black-boxed algorithms that offers an alternative 'way of seeing' (Berger 1972; Cox 2017), where my results reflect the 'gaze of the algorithm' (Noble 2018) back to the viewer (researcher). Furthermore, data visualisations of my search results are either based on 'locative data' with Google Search or 'off the map' with Tor and these 'practices of representation' are ways of not only 'intervening into the world' but shaping it (Beverungen, Beyes, Conrad 2019:624), by making invisible infrastructures more tangible. In this sense my small set of data visualisations in a 'society of control' (Deleuze 1992) attempts to respond to Galloway's call for 'a poetics as such for this mysterious new machinic space'.

Ultimately, I connect elements of these three disciplines to my three methods contained within my methods assemblage, structured by Kittler's 'commands, addresses and data' to describe these technologies, which I apply to the scholarship of three media theorists (Franklin, Kittler, Galloway) and their terms (prescriptive, inscriptive and transcriptive). This diagram contributes

to the project of understanding (Post) Digital Cultures by introducing a methodological framework that is interdisciplinary, incorporating the organisational, mediological and artistic entanglements of online human interaction with search engines. (Figure 4)

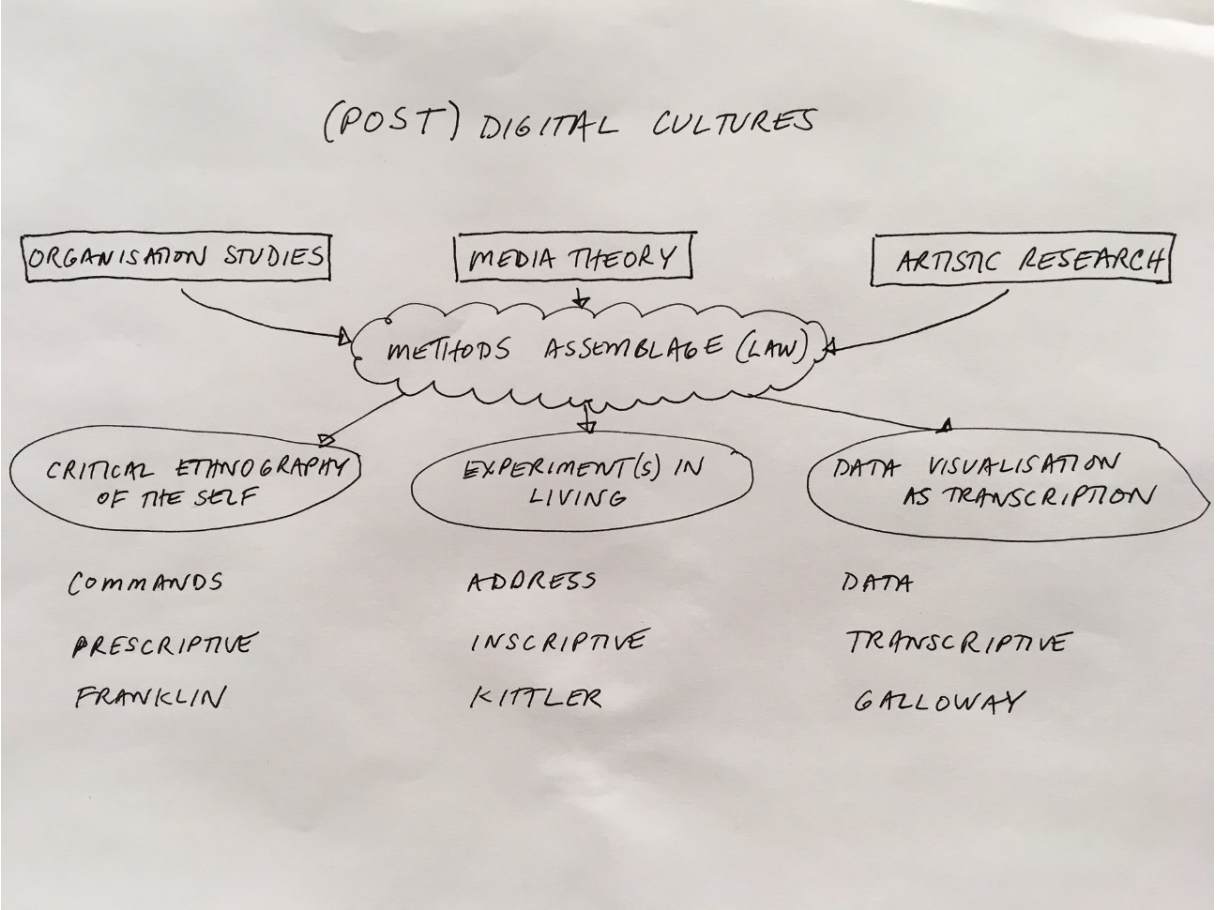


Figure 4: Methodological Framework: (Post) Digital Cultures

6.0 Structure of thesis

Instead of a conventional 'literature review' *Address Offices*: *A Prehistory of Search Engines* uses one book as a conceptual structure, *Die ersten Suchmachinen* or *The first search engines* (Tantner 2015), mapping out its major concepts and contributions in chronological order as chapters in this thesis. The *bureau d'adresse, Intelligence Office* and *Fragamt* within European cities during the 17th-19th centuries collected residential addresses of clients and used human 'crawlers' or servants to search and gather information, creating storage technologies of personal data in public and secret registers, or 'anonymity machines' (Tantner 2015). I describe some of these 'Borgesian Universes' that offered not only products, furniture and employment services but eventually became 'virtual marketplaces' (Blome 2007). I also illuminate the actions of 'Black Utopias', where the gathering of user data became a concern for clients and brought up issues of privacy, policing and surveillance. Lastly, I show that some offices were also 'Emancipatory Utopias', organising cultural and educational activities and even fulfilling an oracular function, reputedly by being able to answer all queries, even serendipitously.

Organising the World's Information focuses on indexing 'the world's information' and provides an overview of 'analogue' human endeavours by private and public actors. I describe the *Mundaneum* (Otlet and de Fontaine 1910), a 'mechanical, collective brain', HG Wells' 'world encyclopaedia' (1936-38), a vision of knowledge and peace, and Vannevar Bush's 'Memex', an imaginary machine that scanned, recorded and disseminated information (1945). Further indices include Eugene Garfield's Scientific Citation Index (SCI) (1964), or 'Hypersearch', which measures scientific publications through linkages, along with their authority, calculating an 'impact factor'. I explain how the *Whole Earth Catalogue* (1968) combined cybernetics and counter culture (Turner 2006) by creating an index of alternative back-to-the-land products. I then describe the hyperlinking of the web (Berners-Lee 1989) and early online web search in the 1990s that included different web crawlers and browsers, notably Mosaic (1994) and the first search engine that was able to crawl 'full-text' search, AltaVista (1995). Drawing on the research of Van Couvering (2010), I explain how academic institutions developed most early search engines yet, by the end of the 1990s, commercial portals from media conglomerates dominated the sector.

A media archaeology of citation begins with a chronological overview of PageRank's antecedents, bibliometry (with Garfield's SCI) and sociometry. I explain some of its technological developments in the first decade of the 21st millennium, interweaving Brin and Page's own text, *The Anatomy of a Large-Scale Hypertex Search Engine* with interpretations and critiques by media theorists (Rieder 2013; Halavais 2009; Franceschet 2010; Peters 2015). I then show how 'research' and 'search' merge by elucidating how PageRank determines relevance with its novel innovations in indexing, IR (Information Retrieval), linking and the figure of the random surfer. Furthermore, I describe how by clicking on hyperlinks, enticed by Ad Words, the labour of users builds Google's proprietary database of intentions (Battelle 2006) and delivers free content through social networking (Benkler 2006) in the hidden immaterial factory of cognitive capitalism (Boutang 2012). I then explain how 'algorithmization of the hyperlink' in turn produces 'network surplus value' (Pasquinelli 2009) for Google.

Reflections on Methods elucidates how I carried out my 'truth games of hide and seek' with algorithms and collected data through my 'methods assemblage' (Law 2004). With my 'critical ethnography of the self' (Wang 2008) I expound upon Foucault's *Technology of the Self* (1982) in regard to writing, querying and the memory tools (hupomnemata) for collecting data with my human computer interaction. I reflect on the 'inventiveness of methods' (Lury and Wakeford 2012) and my self-designed 'experiment in living' (Marres 2012), using my office at CBS. I searched with Google and Tor with the same set of chosen keywords and collected data on myself and produced 'data visualisations as transcription', *Re:search - Terms of Art*. Finally, I explain how *The Cybernetic Hypothesis* (Galloway 2014) influenced my methods in regard to my small data set, which is in contrast to the fetishization of tools and correlations produced with 'big data' sets.

In *The Personalised Subject* I investigate how Google Search is a habit (Chun 2016) that makes information 'accessible', concomitantly identifying and collecting user data by their IP (Internet Protocol) address. I analyse the results from my methods through three lenses: how ranking is determined by ads (Advertisement) and personalisation (Authenticity), along with my unique results (Authorship), assigning each a specific form of capitalism. Advertisement is regulated by the infrastructure of 'platform capitalism' (Srnicek 2016), where atomistic commensuration monetizes these microtransactions, deemed 'Googlenomics' (Varian 1999; Levy 2009). Here, the IP address not only enables user identification, but the categorisation of users into certain

groups with ‘collaborative filtering’ techniques. Seeking to understand unique results, the last section ruminates on the latest developments of PageRank into RankBrain along with Google’s black box politics of obfuscation and ‘visibility management’ (Flyverbom et al. 2016), regulated by ‘algorithmic capitalism’ (Bilic 2017).

Worlds and Technologies of Anonymity presents a genealogy of anonymity contexts, beginning with how ‘cyberspace privacy’ applied encryption technologies (Chaum 1981) such as PGP (pretty good privacy) (Zimmerman 1991). I discuss how, with the ‘privacy turn’ at the beginning of the last decade, the Nymwars debated pseudonymity (the ability to have hidden identities when online) and how the Snowden revelations (2013) exposed Five Eyes surveillance on citizens and resulted in an increase in privacy technologies. I describe how Tor (The Onion Router) is an anonymity p2p browser that is a means to search online without divulging a user’s IP addresses and facilitates exploration of the Dark Net. ‘Onionland’ (Bartlett 2014) is a melange of sites ranging from anarchist forums and illegal activities to the Dark Web Social Network (Gehl 2016) that can only be accessed with Tor in combination with a VPN, or TAILS (The Amnesic Incognito Live System).

The Anonymous User reflects on reimagining search and my attempts to be anonymous using TAILS, in combination with Tor and applying the ‘Grams’ and ‘Torch’ search engines in Tor’s ‘onion services’. Structured by my results, I address a range of ‘anonymous users’ who search and various platforms of the ‘Dark Net’ including myself, intertwining sociologist Gary Marx’s ‘rationales of anonymity’ and ethnographic studies to structure the discourse, along with media and privacy theorists. In the section ‘(De)anonymised users’, I show instances of academic Tor exploits and specifically one case where ‘researchers’ collected a user’s IP addresses without a search warrant. Furthermore, I refer to precedents in US law concerning anonymity online, the difference between content and routing information regarding ‘electronic surveillance’ and users’ ‘expectation of privacy’ with Tor.

Black Box versus Black Bloc synthesizes *The Personalised Subject* and *The Anonymous User* by employing Alexander Galloway’s eponymous essay to structure the effects of Google Search and the Tor Browser, centred around the ‘data subject’. I discuss how Galloway decodes the ‘Black Box’ as ‘an opaque technological device for which only the inputs and outputs are known’ and the ‘Black Bloc’, as ‘a tactic of anonymization and massification often associated with the direct action wing of the left’ (2010:3). Here, ‘subjectivities of search’ and ‘agencies of anonymity’ are organised according to degrees of human-algorithmic interaction. I then compare these ‘collaborative collectives’ as two categories of user’ search activities: *The Personalised Subject* and *The Anonymous User*.

Interweaving Tantner’s ‘Black Utopias’ with Shoshana Zuboff’s ‘surveillance capitalism’ and my own analysis from the preceding chapters, *Black Utopias*: *Surveillance Capitalism* makes an analogy between Google Search and the former address offices. I describe how both are contingent on personal data—with the register or, nowadays, databases (Burkhardt 2015). I explain how the extraction process of ‘surveillance capitalism’ that enables Google’s ‘logic of accumulation’ of data or ‘behavioural surplus’, is in turn used to organise users. Furthermore, I discuss how search histories are seamlessly tied into buying patterns across other technology platforms that enrich user profiles, feeding diverse industries with data as ‘mission creep’ (Christl 2017:10). I show how these commodities of ‘surveillance assets’ or behavioural data are prediction products sold in new marketplaces.

Agencies: *Ad, State, Computational,* discusses the implications and consequences of Google Search organising (us)ers. I begin by elucidating how users seek answers to their questions by 'ubiquitous googling' (Ridgway 2021), consulting 'the oracle' Google that profits from advertisements (Varian 2012; Zuboff 2015) and with 'algorithms of oppression' (Noble 2018). I explain how agencies of the state such as the National Surveillance Program (NSA) de-encrypts users' transmissions, including 'remote search' worldwide and I also address the recent calls for legislation to enforce 'anti-trust' measures on Google, the EU GDPR and *Right to be Forgotten*. Referencing my 'subjectivities of search' I relate how the computational agency of Google's algorithms produces gatekeeping, clickbaiting and 'calculated publics' (Gillespie 2014) that organises (us)ers recursively with modes of 'cyberorganization' (Parker and Cooper 2016), which controls the flow of information back to users.

In *Emancipatory Utopias: User Agency* I draw on Tantner's 'Emancipatory Utopias' to reflect how information sharing occurred anonymously, without nepotism and that privacy concerns are not novel by examining the precedent of secret registers or 'anonymity machines'. Human agency also exists in the present, from 'search raters' at Google to users intervening, just as algorithms do. Instead of illuminating the black box, 'tech-savvy' users apply various tactics of resistance through protocol (Galloway 2004), where programming, code or 'Agencies of Anonymity' facilitate circumventing data capture as they search. Instead of being 'shareveillant' (Birchall 2018) subjects, I conclude by analysing modes of collective resistance to Google's surveillance capitalism (Tor, redistribution of the sensible, right to opacity, obfuscation tactics), along with reimagining (re)search through notions of serendipity, the art of fire and the choice of an alternative search engine.

Chapter 1:

Address Offices - A Prehistory of Search Engines

Figure 5: Tantner's overview (2013)

1.0 Oikonomia

Michel de Montaigne's *Essais* of 1595 merged anecdotal digressions about the home and its organisation with intellectual ruminations on daily life. Montaigne's self-reflection also imparts what he learned from his father who enacted a certain domestic rule regarding his 'oikos'—employing a servant, later a butler, to keep a written account of household goings-on (Tantner 2015:33-34). This could be considered a precursor to what Foucault terms the 'Technology of the Self' —a daily act of writing capturing the occurrences of the world and the desire to organise the world's information in one place—the daily journal, as a memory aid but also a means of correspondence addressed to oneself, which I will address in Chapter 4. In 'One defect in our government' (1595), Montaigne writes that his father also advocated introducing the following practice: every city needs a certain place assigned for anything that might need repair and to have their business entered by an officer appointed for that purpose (ibid). These could be considered the first musings on an office able to field different kinds of enquiries.

This is also the 'Urszene' (primal scene) of *Die ersten Suchmachinen* (The first search engines) by Anton Tantner, who connects Montaigne's father's daily journal, 'oikonomia', with the 'bureaus d'adresse' or 'intelligence offices' of early modern Europe.[8] (Figure 5) Operating

[8] An edited version of his dissertation, the book is a survey of various 'offices' in Paris, London, Vienna, Innsbruck, Graz, Bratislava, Brno, Lviv, Prague, Schleswig and Berlin.

throughout European cities, people who were physically distanced from each other could be placed in contact with one another through these 'offices', which helped residents, foreign visitors or newcomers find what they were looking for. Tantner demonstrates how these 'offices' became the 'prehistory of search engines'.

2.0 Publicke Register for Generall Commerce

On March 5, 1611 the poet and translator Sir Arthur Gorges and the manager Sir Walter Cope were granted a charter for 21 years to establish the *Publicke Register for Generall Commerce* in London.[9] In their statement they mention Montaigne's essay 'Of a Defect of our Policies' (1595). Gorges and Cope's *Publicke Register* was intended to provide the things that people were in search of via a neutral third party and to bring them together under one roof. They declared that the office would enable unusual goods to be sold, even if there were few buyers. This statement foreshadows Chris Anderson's 'long tail' theory in which the internet enables the exploitation of niche markets (Tantner 2015:105).

At that time carriers were being robbed when delivering money. Through a network of *Publicke Registers*, they envisioned a cashless money transfer that would enable payment notes and it would even be possible to use another name [pseudonymity] (ibid). In this way Cope and Gorges' 'intentions' seem more reminiscent of the establishment of a bank network (ibid:107-8). This prescient pre-telegram system of 'wireless' money transfer resembles Western Union, digital transfers and Bitcoin, where people use pseudonyms. However, the *Publicke Register* was also planned to mediate the sale of goods (Tantner 2015:106) because later on, in remarkably similar terms to those used today when executing a search query, they state:

> Always provided, that no man bee compelled to make Entrie or Search in the said Office at all, nor when such Entry or Search shall be voluntarily made, shall pay any more for such Search or Entrie then shall please himself (Gorges 1611).

Ostensibly, one would be able to decide the value or worth of placing a request for an 'Entrie or 'Search'.[10] Gorges also stressed that brokerage at the office would be confidential and secure, not cheating other men or revealing their wealth.

Moreover, Gorges noted that the names of the lender and borrower, as well as the nature of the deposit, would not be registered. Instead, another 'private paper' would be kept only for the duration of the negotiations, which upon conclusion of the transaction, would be handed back to both parties (Tantner 2015:107). The staff of the office would in turn have to swear an oath in order to guarantee the secrecy of the transaction. Moreover, another accusation faced Gorges– that 'authorities' might be able to peer into the 'public register' (ibid). It is important to note that Gorges stated that the rules of honour would be respected along with discretion—hypothetically of course— because to this day, there is no proof that the *Publicke Register* was ever actually realised.

[9] The monarch issued charters, documents or legal instruments, also called 'letters patent', or 'octrooi' (in Dutch) and 'Privileg' (in German), to individuals, groups or legal entities. Many of the offices mentioned in this chapter were officially granted patents or charters by their respective sovereigns though some were never realised.
[10] With the exchange system of online search today personal data is willingly relinquished for free services.

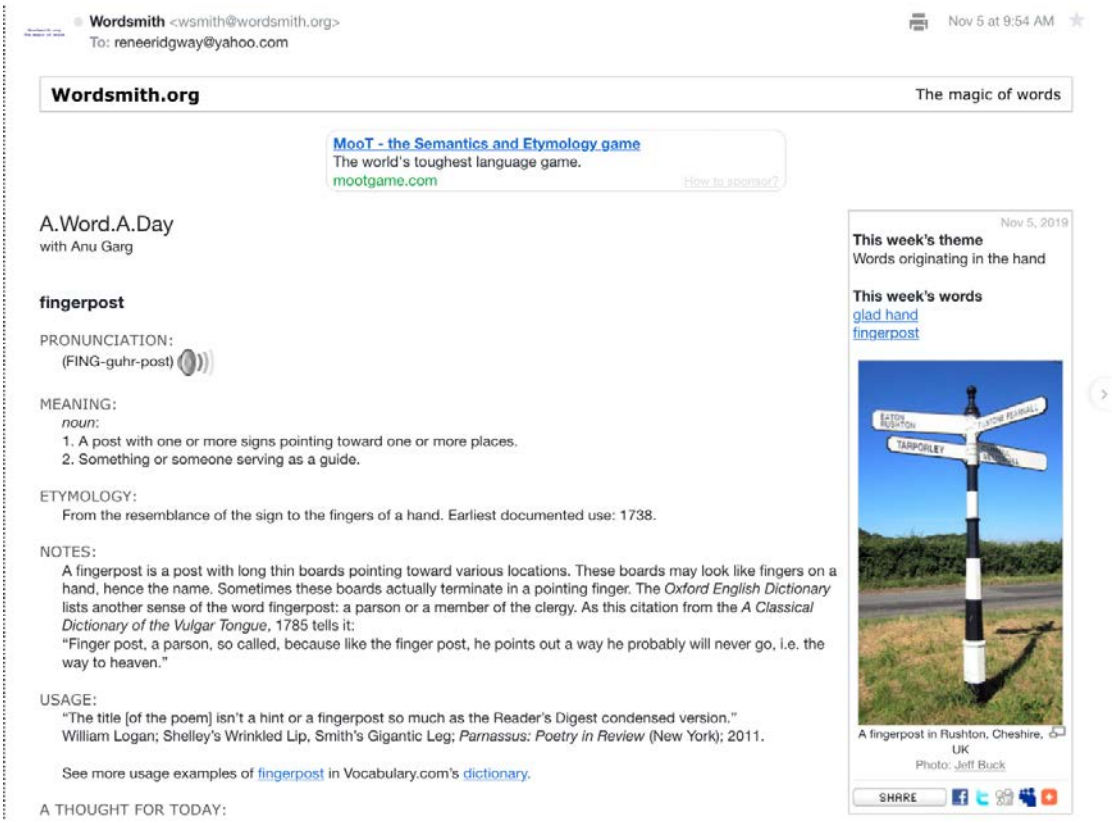


Figure 6: 'fingerpost' from Wordsmith.org

3.0 Bureau d'adresse

Considered one of France's first journalists, in 1630 the medical doctor Théophraste Renaudot applied for a 'privilege' from the monarchy, citing Montaigne's anecdote along with claiming that for the perfection of society there needed to be a place where people could come together publicly. He used the metaphor of the telescope, which could collect all the scattered pieces of the social body together.[11] Additionally, Renaudot used two other metaphors that Tantner argues are better comparisons to the bureau d'adresse: signs one finds at a crossroads (enseignes ou adresses) or 'fingerposts' and the table of contents in a book (ibid:43-44). (Figure 6). In 1630, after securing the support of Cardinal Richelieu, Renaudot was able to open his bureau d'adresse in Paris.

In Renaudot's document *Inventaire addresses du Bureau de Rencontre*, 'address' meant not just the notion of looking for a specific location or contacting the office but also addressing a place where the seeker has their wishes fulfilled (ibid:268-269). Emphasis was placed on the the notion that the address (office) was a physical place one could turn to in order to meet others and find items one was looking for (ibid:267). There is also the connotation that 'to address something' is to give it a certain amount of importance. Still, in this usage, as a verb and not a noun, the action of 'addressing' is also an act of 'interpellation', where one is addressed by someone else and accorded the status of 'subject'. In Chapter 8 I will discuss the social constraints and the technological conditions that inform this 'interpellated subject' as denoted by Louis Althusser (1971).

[11] A treatise on the telescope, lenses and focal points and how they connect to the pre-history of search engines is beyond the 'scope' of this thesis.

When foreigners arrived in the city they immediately had to register at Renaudot's bureau d' adresse, helping newcomers find accommodation and employment. It also procured information for individuals searching for an address and arranged travel for those seeking companions en route. Providing help to all interested parties in the form of clothing and food, the bureau d'adresse also supplied medical treatment for the needy. The poor received free legal advice whilst wealthy and important people found trained servants. Women, however, were not allowed to enter.[12]

3.1 Human Crawlers

The bureau d'adresse employed servants or 'human crawlers', who conducted searches and acted as agents and go-betweens, controlling the resources without divulging the address of their employees (ibid:280-281). Employed by the wealthy, these 'intelligence-servants' or insiders—'middlemen', 'middlewomen' or 'information harvesters'—purchased goods through buyers, acquired information en route and brought it back to the office. Markus Krajewski elucidates how the collecting, bundling, processing and transmission of information by servants that hunt data is analogous to the web crawlers of today's search engines, 'constantly searching for new information, which they gainfully process' (2010:13). Tantner draws on Krajewski's research to make a comparison between Montaigne's head 'butler' who tended to the family's 'oikos' and the well-known butler 'Jeeves', who performed search enquiries to obtain information and delivered results, yet had a double function.

> One of the search engines competing with Google in the 1990s was called 'AskJeeves.com'; this name also alluded to the ambivalence of these institutions. On the one hand, a human search engine is a helping hand, on the other however, she or he is a bearer of secrets, consistently suspect of being an informer, a spy, and of betraying these secrets to other powers (Tantner 2014:128).

A paranoia prevailed amongst the wealthy that servants might spy on them and they were often accused of sharing secrets with others and passing on private information (ibid). However, efforts were made to ensure the secrecy of information and that the servants were trustworthy. With regard to the former, the parties only knew the name of servant, who filtered information and conveyed 'data' to the involved parties. Because these servants held coveted information, they could decide what to share and when, passing on information to the bureau d'adresse, or 'Intelligence offices' or later even the police (ibid:134). It was only through the offices that servants could find their next job placement and it was the register which held both the data and their reputation. This is analogous to today's online 'reputation economy' where search engine results reflect one's reputation, working much like the authority and backlinks of Google, with PageRank determining value (Chapter 3) and creating profiles (Chapter 9).

[12] Women were excluded as potential customers on moral grounds. Female honour was important as women might be taken into unsafe homes and it was unclear whether prostitution accusations circulated. As an alternative the 'bureau d'adresse' suggested that women send their husbands (Tantner 2015:70).

Figure 7: Renaudot's *La Gazette* bounded in book form (1644).

3.2 Public Registers

Whereas Montaigne's father once maintained the 'oikos' of events of the home with a daily journal, Renaudot's 'address office' gathered offers and queries from these 'human crawlers' or 'intelligence servants'. Clients presented inquiries and listings orally at the office and a staff member transcribed the entry into the register, which replaced the verbal transactions of the inn or 'comptoirs', where people had previously gathered to share information. In this way the offices became data centres, where transactions were carried out according to a protocol and the method of accounting was written down in a register or *Protokollbuch*. As clients consulted the registers and histories of these exchanges, the pages recorded the data of the parties who were seeking and offering, along with the relevant financial details. This collation of information into registers is comparable to the databases constructed by Google, which I will discuss in Chapter 5 and 9.

The register recorded clients' search requests as accurately as possible along with their prospective suppliers and customers' inventory, including the conditions of sale. In principle, the transfer of information in the registers was ideally carried out on a first come, first served basis (Tantner 2015:63). If the transaction was successful, the prospectors would then receive a copy of the entry and the relevant address (ibid:62-63). Buyers were connected with sellers, seekers with finders and, for a small fee, anyone who was searching could consult the register as well as placing an ad, in a manner analogous to the present-day log-in and returned hyperlink. Whether renting property, engaging in transactions of capital or selling objects, it all went through the register (ibid:270-271). Renaudot stressed that the bureau d'adresse intended to receive all persons who wanted sell, buy, lease, exchange, borrow, learn or teach something without favouritism. Therefore, the major activity of the bureau d'adresse was as a public convenience as it did not actively seek clients (ibid:66). Shortly after its opening in 1630 the bureau d'adresse had already helped 3,000 people and by 1634, more than 50,000 (ibid:55).

3.3 Secret Registers

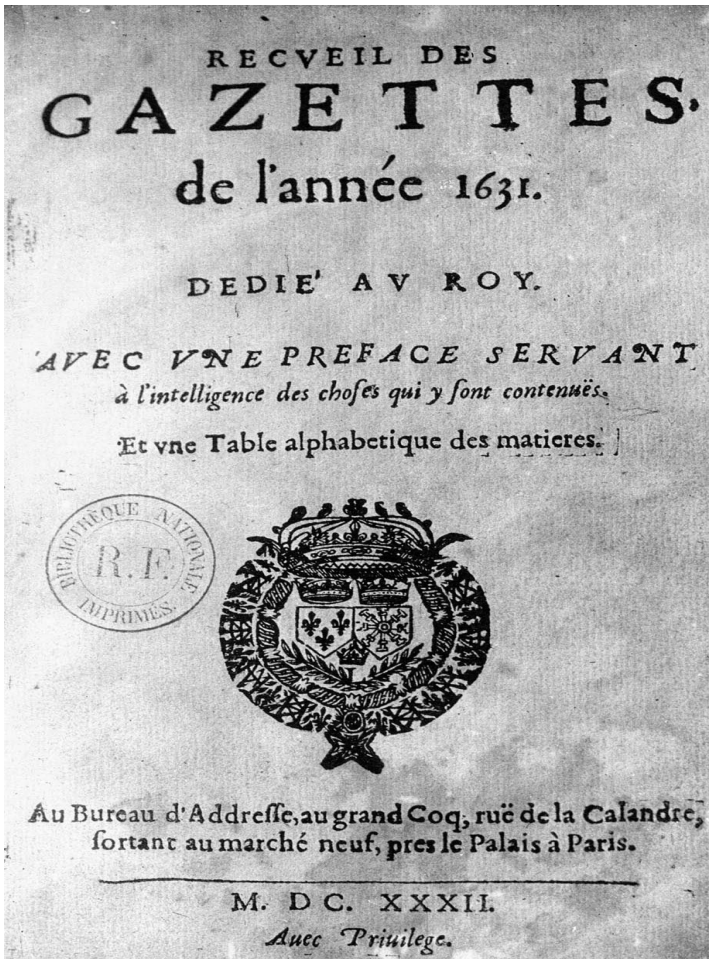
With their unrealised 'Publicke Register' Gorge and Copes not only proposed bringing buyers and sellers together to make money but designed a double entry bookkeeping system, which offered anonymity to clients. Tantner describes this anachronistic intervention of data protection as preventative and he surmises that in order to deal with those critical of this method, in 1639 Renaudot also devised two registers for work placement that distinguished people: one for 'common' people and the other for the 'honourable' (2015:64). A brochure from 1639 describes the following procedure:

> There was a secret register in which the names and residence of clients were registered along with their offer or request and a public register that would not cite the names and addresses of the client, yet was exposed to the eyes of anyone (ibid:67).

According to Tantner this method served as a type of 'anonymity machine'. By filing the requests of clients in the secret register it became a form of encryption. This is analogous to PGP encryption that generates two keys, a public one that is distributed and a private one, which I will discuss in Chapters 6 and 11.

Only interested parties would get to view the public register and if they found something worthy, paid three sous (cents) to know the contents of the secret register. In this way, insight into secrets became available only for those who could afford it. Maintaining privacy was then a privilege and *confidentiality* was an advertised keyword that concealed the face behind the information (ibid:206). However, it seems Renaudot was forced to discontinue his practice of a secret and public register because in 1647 he published a plan for a 'renewal of privilege' for the bureau d'adresse, stating that the messages from the secret registry would be transferred to the public register within 24 hours (ibid:69).

From the beginning, Renaudot's 'bureau d'adresse' placed value on discretion for his customers regarding job placement. They were notified when the deal was near completion only if they were honourable and of good repute, and the identity of the enquirer was known only by the employee of the office. There was even the possibility that anonymity could be upheld for those who were embarrassed or did not like dealing with private middlemen (ibid: 52-54,57). Instead of a 'traditional intermediary' one could nominate a third party for follow-up questions or to carry out the request (ibid:66). With its anonymous job search activities, Renaudot considered the 'protocols' of his bureau d'adresse advantageous compared to the nepotism occurring between friends and servants who could be spying (ibid:69). It is therefore important to note that the need for an anonymous exchange of information, which did not depend on personal relationship networks, was also one of the offices' primary functions (ibid:277-278).


RECVEIL DES
GAZETTES,
de l'année 1631.
DEDIE' AV ROY.
AVEC VNE PREFACE SERVANT
à l'intelligence des choſes qui y ſont contenuës.
Et vne Table alphabetique des matieres.

Au Bureau d'Addreſſe, au grand Coq, ruë de la Calandre, ſortant au marché neuf, pres le Palais à Paris.
M. D C. XXXII.
Auec Priuilege.

Figure 8: Frontispiece of Renaudot's *Recueil des Gazettes* (1631)

3.4 Virtual Marketplaces

Eventually, with the elimination of the middle[wo]men, Renaudot compared the bureau d'adresse to a market, serving as a virtual exchange for goods of all kinds and as a sales agency offering many kinds of objects ranging from farm tools to luxury items including paintings, medals and ancient coins, manuscripts, rare plants, strange animals and mathematical instruments, estates or homes, jewellery, watches, antiques, books, furniture and machines. Once there was even a young dromedary for sale 'at a reasonable price' (ibid:72).

With the introduction of printed matter the site of mediation shifted and expanded beyond the brick and mortar establishment of the bureau d'adresse with registers. Where once information was recorded in a ledger, now the material was circulated in the form of a journal enabling the dissemination of information as a 'channel'. In 1633, with the support of Cardinal Richelieu, Renaudot starting publishing France's first journal/newspaper, *Feuilles du Bureau d'Adresse*, which later became *La Gazette*. (Figures 7,8) Anyone who paid 3 sous (cents) could advertise or consult the register. According to Mark Tungate, historian of advertising, Renaudot was not only the first journalist but the 'inventor of the personal ad' (2007), a foreshadowing of Google Ads. This printed space became the site of exchange between mutually interested individuals and Renaudot claimed they could find what they were searching for within a quarter of an hour.

Within a short time 12,000 transactions had been conducted—mostly with the bureau d'adresse as middleman and not showcasing anything except paintings. The heading in Renaudot's *Feuilles du Bureau d'Adresse* from June 8, 1637 read:

> Paintings: Several of them by good masters. Tomorrow viewing at the bureau d'adresse (ibid:52-53).[13] (Figure 9)

Eventually the majority of the products were announced in the journal. Tantner argues that throughout the history of the modern era there was a different European concept of place, one in which addresses have emigrated from data (ibid:270). By separating media (ads journal, posters) from their source—a physical location (ibid:92), these 'virtual marketplaces' (Astrid Blome 2006 cited by ibid:162) in larger European cities put people in contact with other people they otherwise would have never met. There was a notion of 'deterritorialisation'—the separation of objects from the register, with items not located at the office but remaining with their owners. As I demonstrate in Chapter 5, this is how many items and services are found online with search engines in an era of platform capitalism, or with Amazon or ebay, where the goods are not 'found' in a physical location.

From 1637 onwards Renaudot's bureau d'adresse had royal permission to act as pawnshop, or *mont-de-piété* (the first in Paris) and, by 1641, more than 30,000 people had taken advantage of its services (ibid:53). The bureau d'adresse not only sold second-hand items but also functioned as a lost and found for all kinds of objects (and eventually people), far and wide. Travellers could announce their arrival, their stay and their departure from the city. Besides those who wanted to travel and obtain information about various routes, the bureau d'adresse also arranged the delivery of packages and letters by messengers on foot and on horseback. As stated in the *Hartlib Papers,* in 1639 Renaudot also conceived of the bureau d'adresses to communicate with

[13] The 'bureau d'adresse' thus assumed the role of an art gallery, for example a Madonna after Rubens and Abraham's *Sacrifice* were shown (Tantner 2015:52-53).

each other as hubs in a ‘network’ (ibid:73;153) laying the groundwork for present day technological infrastructures. Beginning with Renaudot and throughout the following century, the bureau d’adresse also served as a location for the postal delivery of letters and packages that could be collected (ibid:49-50). By 1647 the bureau d’adresse had helped over 80,000 people.

Figure 9: *BUREAU D’ADRESSE, POUR LES CURIEUX. Caption reads:* Oú ils trouveront les principaux évenemens de l'année 1696. *Et les heureux presages pour l'Année presente 1697’.* Permission: Bibliothèque nationale de France, département Estampes et photographie.

Renaudot died on October 25, 1653. Renaudot's son, Théophraste II, along with Gilles Filleau, continued the bureau d'adresse and published the *Liste des Avis du Bureau d'Adresse* in 1669, a small magazine (Anzeigenblatt). 'All of Paris' could be captured 'in a moment' with this journal, which appeared every two weeks and was available for purchase in 14 places around Paris for the price of 15 deniers (pennies) (ibid:83). With the ad journals, one paid to advertise and one paid a subscription in order to obtain information. In subsequent bureau d'adresses the ad journal became more important than the registers (ibid:91) and was less expensive to produce than the posters, which were made at the expense of the 'bureau d'adresse' and plastered on the walls of buildings in central locations. Whilst 'free' to read, they needed to be continuously printed and installed. The posters were not seen by everyone, whereas the journal could be widely distributed and therefore landlords had a greater selection of potential renters (ibid:84), with their readership extending beyond the borders of France. In this way the medium of the advert journal could be viewed by many who didn't come to the office (ibid:280-281).

4.0 Borgesian Universe (Question Office)

One special case described by Tantner was the garrulous 'Question Office' (Fragamt) in Bratislava, at that time the capital of the Austrian-Hungarian empire. It was declared 'open to all' by its founder Anton Martin on April 9, 1781 guaranteeing discretion and facilitating a bewildering array of services (ibid:196). There was a panoply of items ranging from real estate, music academies, books, tableware, decorated title pages, coins, a 'Wunderkammer', labour services, a fencing master, children! and even 'Lüneburger-Käs', a type of goat's cheese (ibid:199).[14] Moreover, the media of the 'client journal' (Kundschaftsblatt) established a huge network that extended outside of Bratislava to Vienna and Pest and enabled buyers and sellers to communicate lists of all kinds of desired food products (ibid). Tantner describes the layout of these listings (published as a supplement to the Wiener Zeitung) as resembling an unordered 'Chinese Encyclopaedia-like Borgesian Provenance compilation' because there wasn't enough room in the four pages to publish descriptions for all the goods in detail (ibid). Anything one searched for could be found, or so it seemed, because everything was offered. In this way the Bratislavan 'Question Office' claimed to be a true universal facility (ibid:219-220) and perhaps, a true 'platform of capitalism', which I will address in Chapter 5.

Serving as the physical 'hub' or a 'virtual platform', people could be directed to goods or even other people, organised by personal advertisements. Martin's 'bulletin board-like' messages to his 'readership' included notification of the next arrival of Marschansker apples with a written note on the office door (ibid:206-207). The clerks working at the 'question office' could be also entrusted with personalised services, writing a variety of documents in German and Latin. Martin's 'Fragamt' even went so far as to arrange the washing of laundry (ibid:205). Along with a lending library, 'crowdfunded' projects, where people could pre-order financed inventions also existed, however, a physical 'reading room' was not allowed as it would facilitate the intermingling of genders and classes. Besides a 'Borgesian universe' Martin's 'Fragamt' also offered to locate missing persons. Individuals could be easily searched for and found by their residential address (ibid:212-213). Yet after only 3 years the rise of Buda contributed to the demise of Martin's platform in Bratislava, which only catered to 30000 inhabitants.

[14] Coincidentally, my research lab, the DCRL (Digital Cultures Research Lab) is located in Lüneburg, Germany.

Failed attempts were also made to set up provincial chains of address offices in Bordeaux, Angers, Aix-en-Provence, Lyon, Grenoble and Rouen. In Dresden Paul Marperger and Emanuel Jacobi set up an address office (Adreß Kontoirs) in December 1714 offering a sales and money exchange, along with employment and administrative services (ibid:253-254). There were accounting services for merchants and one could appoint a notary to make contracts between merchants and future servants (ibid:226-227). This 'address office' stated that it would be a centre for all kinds of information and as Tantner notes, this sounds vaguely similar to the mission statement of Google.

Following in the footsteps of Renaudot, with the creation of pawnshops, goods could not only be viewed but payments for them could also be obtained. These pawnshops eventually expanded into institutions for money lending as they began to take up the function of banks. Which begs the question, 'what role did the offices play in the expansion of the money and the market economy?' In 1723 Jacques Savary of Bruslons responded in his *Dictionnaire universel de commerce* that a bureau d'adresse acted as 'a universal resource', 'where you could make money from anything' (ibid:282). Thus, the universal resource was information and whoever controlled and organised it exerted a certain amount of authority over the population.

5.0 Black Utopias (Intelligence Offices)

Although not its primary function, in the years of 1639/1640 Renaudot's 'bureau d'adresse' adopted two arrangements that could be called 'police tasks' and this monitoring system was connected to people's residential address, analogous to the contemporary IP address, which I will discuss in forthcoming chapters. In the first instance, artisans who arrived in Paris looking for work had to register at the bureau d'adresse within 24 hours as well as accepting work that was offered. If they did not receive a certificate issued by the bureau d'adresse within 24 hours of arrival, the concerned persons were no longer allowed to stay. However, this 'monitoring system' hardly worked in practice (ibid:55). One also needed to register a personal address in order to have letters and contracts delivered (ibid:253-254). Whereas Renaudot's bureau d'adresse of 1639 used registration to reduce poverty rather than to enact policing (ibid:190) eventually what came to be known as 'Intelligence Offices' developed policies that would now be understood as 'data profiling'(Chapter 9).

In 1721 Johann Gottfried Gutkäß presented a large-scale project to establish an 'intelligence office' that facilitated the creation of a stock exchange and a poorhouse checkout, similar to police registration, which sought to profile strangers through their height and weight (ibid:229). From 1727 onwards a closely linked series of 'intelligence offices' (Intelligenzbüros) were built within a postal network (ibid:238). On September 1, 1730, Gottlob Christian Hilscher presented an Intelligence Journal (Intelligenzblatt) that would only accept written submissions. Hilscher criticized service requests that were submitted by jobseekers or servants on just a note, as often these did not contain sufficient information (ibid:232-232). He therefore explicitly requested that they included their age, their present city of residence, the names of their parents, if still alive, and, if so, their addresses along with their former places of domicile. In regard to former employment, they were asked if they were let go and what kind of professional skills they had. Along with providing their names in Latin and German and their signature, detailed profiles in the form of a 'curriculum vitae' were compiled based on these written submissions and this data was controlled and 'owned' by the intelligence offices (ibid:233-234).

These intelligence offices also began to collaborate with the newspaper *Wienerischen Diarium* and in 1721 the Viennese 'question office' introduced the service of job placement. These first detailed registrations formed profiles of people with 'register lists' (Negotienlisten), where servants were asked to provide personal information and credentials who would then be able to be monitored yet also be ensured of a reasonable wage (ibid:170-171). In March 1730 Johann Christian Crell submitted a request to set up a *Question and Information Office* (Fragamt und Berichtamt) primarily commending it as a means to separate the bad from the good servants, based on behaviour. Servants needed to submit a testimonial and have a certificate signed by the inspector. A similar control function was applied to housing services. In this way, good citizens would be separated from the 'riffraff' (ibid:231-232).

Like its Parisian predecessor, the Universal Register Office served as a sales agency, a labour office, a pawnshop and also a travel agency. Founded in 1750 in London by the novelist Henry Fielding and his half-brother John Fielding, they stated that requests would be handled with utmost care and that the concern should be submitted in writing, with precise descriptions (ibid:125). The collated data would enable the quick processing of requests and John Fielding stressed that no description could be too detailed. In other words, it seems that the more data collected, the better informed they were (ibid:125-126). The Fielding brothers went much further than just this type of data gathering, with their inauguration of the *Bow Street Runners*, the first citizen police force, taking crime into their own hands. In addition, the methods adopted by the Fielding brothers were also directed at crime prevention by protecting newcomers from the countryside from city fraud activities. To that end, Henry Fielding set up an information office exclusively concerned with uncovering crimes with the names and descriptions of offenders recorded in a register (Tantner 2014:132). This exercise of quasi-police functions was enhanced by Fielding's 1748 practice as magistrate for Westminster and Middlesex, which enabled him to question thousands of suspects.

Similarly, he made rulings about their credibility, as the staff of the Universal Register Office did in relation to the statements of servants (Tantner 2015:129). With Fieldings' office claiming to have found work for so many, in 1755 they published a separate catalogue with 24 commandments for servants that included maintaining secrecy about family matters and emphasising that drunkenness was to be avoided and this type of control had a disciplining effect on servants (ibid:127-128). They also published *The Covent Garden Journal* that provided relevant information regarding fighting crime. In general, the public praised the Fielding Brothers, as stray and depraved servants were usually blamed for crime whilst newcomers from the provinces needed protection against fraud (ibid:128). However, the Fieldings' office register could thus be considered the first police criminal records department, becoming a 'public eye' concerning servants' behaviour (Tantner 2014:132).

6.0 Emancipatory Utopias

It was not only the negative complaints regarding display monopoly, policing and surveillance activities that created public discord—the early modern offices were perhaps too utopian and therefore incited critique (Tantner 2015:78-79). In Renaudot's *Inventaire* from 1631 he mentions how the lack of an address created hardship for poor artisans and others therefore he created housing and employment opportunities for the needy (ibid:268). Renaudot also provided addresses of doctors, surgeons and pharmacists who advised, treated and prepared cures for poor people for free (ibid:268). Moreover, Renaudot encouraged the dissemination of knowledge to

everyone and the bureau d'adresse meant joint property, or even worse, the 'heresy of the Anabaptists' (ibid:78-79).[15] This is similar to the reputation of 'the commons', such as with the Tor Browser p2p collective that I will discuss in Chapter 11.

In 1633 Renaudot's bureau d'adresse also realised a sort of 'scientific academy' where lectures were advertised, well attended, summarised and published in book form without the names of the lecturers. Renaudot intended that with this participation in knowledge exchange, courses from all disciplines should be taught except for politics and religion. The oracular function was also prevalent, documented by Renaudot's 'issues to be addressed' (questions à résoudre) that encompassed everything and to which the bureau promised to provide answers (ibid:50). Another example of their prophetic function was the 'Inquiry and Information Comptoir' from 1820 in Vienna, which was regarded as being capable of predicting the future. Tantner notes that it is frequently the case that with every new medium the impetus is to generate utopias of omniscience (ibid). Today this sounds vaguely familiar, as with the oracular reputation of Google and its promise to provide answers to all queries, even regarding politics and religion. Allegedly it would also be able to supply information about the future and prediction products (see Chapters 9 and 10).

Samuel Hartlib was a refugee from Prussia and, as with other 'office visionaries' before him, was well aware of Montaigne's essay along with the charter of Gorges and Cope (ibid:110-111). One of his correspondents in Paris also delivered brochures about Renaudot's bureau d'adresse, which inspired him to apply to open his London 'Address Office' (ibid:109). Hartlib's notion of an office was even more comprehensive than his Parisian counterparts, in that it could fulfil all human needs and enable the execution of a 'well-ordered society.' Reminiscent of Google's mission statement, it would include an educational institution that would be a gathering place for all available knowledge (ibid:111). He even went so far as to describe this as a 'machine' that would order all chaos and bring to light shortcomings of society (ibid:111-112).

Gottfried Wilhelm Leibniz was well aware of his predecessors—and in 1675 envisioned an even more comprehensive establishment, including an 'Office for Inventors' whose works would be published in *Semestria Literaria* (ibid:148-150). He described a theatre of all things imaginable, which included a menagerie, a herb garden, a laboratory, an anatomical theatre, a cabinet of curiosities, concerts, art galleries, conversations and conferences. Leibniz also suggested a number of other facilities including a timbered house, a prison, a magazine or department store and a charity for the poor along with free medical service (ibid:150-151). Leibniz's musings extended to an academy and in 1712/13 there was a scheme for the establishment of a bureau d'adresse, which would be used to finance the planned Imperial Society of Sciences. Intending to serve all of those with ingenious ideas, it would support scholars wishing to produce critical works who didn't have publishers so that they could make a living and the general public could learn from their inventions.

With his 'reimagining of search' through the utopian conception of an office, Leibniz's societal vision proposed the extraction of certainty out of chaos. Leibniz did not however wish to eliminate chance entirely— rather he tried to fully integrate the process of information transfer into his plan. When consulting the register, 'one often finds what one seeks yet one also discovers other desires one hadn't even thought of' (ibid:152-153). Following in the footsteps of Renaudot (whose office he cited as inspiration), Leibniz described the good fortune of

[15] However, Renaudot also offered authentication certificates for ideas or proposals (Tantner 2015:51).

‘serendipity’, before the word as it is known today even existed (ibid:153). The office then was not only an ‘Emancipatory Utopia’ but also a place where searches resulted in unexpected and fortuitous encounters, or serendipity, which I will discuss in Chapter 11.

None of these offices, with the exception of Renaudot’s, were realised yet there were Emancipatory Utopias established that included art, music and books. Anonymity and discretion were also an integral part of these visions. In 1722 Anton Heinscheidt’s office in Frankfurt specifically mentioned that the journal ad should be placed anonymously and whoever wanted to know the name of an advertiser had to go to the office and pay four ‘kreutzer’, a unit of currency. In 1730 the book dealer Hans Jacob Lindinner opened an ‘Information Office’ (Berichthaus) in Zürich and published a ‘newspaper’, which offered assurances that customer privacy and secrecy would be respected (ibid:246). One is reminded of the Tor Browser that promotes secrecy and anonymity by hiding the user’s IP address or with contemporary VPN’s where one has to pay extra for privacy. In 1748 Siegmund Ehrenfried Richter established an ‘address office’ (Addreß-Comtoir) in Dresden that served not only the management of the advertising business but also the sale of goods and as a place of culture (Figure 10). In one room music lovers could test out the latest available sheet music for sale on the best Viennese instruments, free of cost. In addition, domestic and international journals and reading books of all kinds would be available in an adjoining hall for locals and foreigners’ (ibid:235-236).

Figure 10. Siegmund Ehrenfried Richter’s ‘address office’ publication (1756).

These offices were Emancipatory Utopias in the minds of those seeking to organise the world's information (Renaudot, Martin, Leibniz, a.o.) as research and academic institutions that provided a model for the early development of search engines, before the era of Google (see Chapter 2).

> The main motive of all projects was to consolidate communication, search for new and effective communication structures, and create institutions that would take over the collection, registration and distribution of existing information which then was available for everybody (Blome 2010:220).

In this way the offices fulfilled the functions of public libraries, museums, conservatories or today's local cultural centres by bringing information to the public for free. These activities also resemble contemporary online p2p 'networked' distributed publishing of scholarly material as open access such as the Internet Archive (see Chapter 11).

7.0 Reputation economies

The reputation of the offices was that they assisted citizens in offering housing, employment and goods and by answering queries so that they could obtain information, services or items. What mattered to the public was the 'relevancy' of the search results they delivered. However, there were concerns about what privacy measures were taken by the offices during the transactions and afterwards with their data in the register. Already in May 1630 there was talk of scandals being committed every day at Renaudot's 'bureau d'adresse' regarding mediated services that were quasi secretive and somewhat anonymous, yet were then made accessible to the public through the register. Tantner postulates that one of the reasons for the publication of the *Inventaire* by Renaudot, which presented the activities of the bureau d'adresse to the public, was in order to prevent slander. The register still appeared to some as dubious as it aroused all kinds of fantasies, with accusations ranging from marriage and matchmaking to abortion and prostitution. There was even talk about the figure of a pimp that was haunting the bureau d'adresse (Tantner 2015:74-75). The reputation economies of the Dark Net in today's online world, or those using anonymising technologies such as Tor to search them are thus analogous to the reputation economy of these offices in that they consist of tales of criminality.

Besides constructing the reputation economy with citation (as I will discuss in Chapter 2 and 3 with regard to the Scientific Citation Index), a shift occurred in the 18th century concerning both the function and naming of the offices (ibid:168). Throughout the centuries the offices were aware of each other's activities and reputations, reflected by the fact that applications for charters cited their antecedents—Montagne's *Essais* and Renaudot's infamous bureau d'adresse. It is worthy of note that the offices altered their names constantly over time to deliver an image that would be approved by their public, like Google rebranding itself as a 'research company' called Alphabet (see Chapter 10). Eventually, the clientele realised that these institutions gathered everyone's intentions into a register (or nowadays into Google's database) and therefore the public conception of the offices included control and surveillance activities.

The 'Intelligence Office' then became synonymous with taking over policing functions (ibid:145) along with missing person's reports and sensational criminal cases becoming newsworthy (ibid:171-172). With its policing and surveillance techniques, John Fielding of the 'Universal Register Offices' accused competing offices of forgery and warned the public against

using them (ibid:130). Fielding stated that the office should only be undertaken as a monopoly and that the income should flow to the government (ibid:132). In this way the private and thereby privileged institution of the Universal Register Office was also designed as a governmental agency contributing to a surveillance society (ibid), foreshadowing contemporary discussions on Google, the state and 'surveillance capitalism' that will be further elaborated upon in Chapters 9 and 10.

By the late 17th century, the term 'Intelligence Office' generally designated facilities with a poor reputation that were accused of defrauding those seeking employment or sending money to the colonies (ibid:121-122). Moreover, certain Register Offices were also accused of encouraging prostitution, as maids were promoted to masters on 'show-days' and thereby treated as commodities (ibid:134). It was not only prostitution (das horizontale Gewerbe) that was encouraged but also 'good old commercial pimping' (gute alte Gewerbe der Zuhälterei) as satirised in John Reed's play *Register Office* (1761). Within the noble or bourgeois families that were able to afford servants, this suspicious attitude was an especially pronounced phenomenon around 1800, which helped precipitate the downfall of the offices. Notably, the Fielding Brothers Universal Register Office was an exception to this rule, declared by even its harshest contemporary critics as exemplary and reputable (ibid:134-135).

8.0 Denouement

During the time of the Habsburg monarchy 'competition' between the 'Intelligence Offices', the name for which they had obtained charters, and 'Question Offices' played out (ibid:172-173). Eventually, the Habsburg offices went by more public sounding names 'Frag- und Kundschaftsamt' and instead of operating as employment agencies or pawnshops, these offices began to publish journals or advertisement magazines called 'client journals' (Kundschaftsblätter) and 'ad magazines' (Anschlagstafel) (ibid:265-266). Besides the changing of names, relevancy was a crucial issue with office media (registers, ad journals, posters) because the updating of transactions needed to be current and the information valid (ibid:279-280). As with his predecessors, updating data after the conclusion of transactions proved to be a problem and Israel Saul's solution in 1827 was to list rental properties of Wroclaw in the register or 'log book' with a three-month expiration date, which could be extended when needed (ibid:258). The validity of the ad journal was limited to a period of two weeks and if no transaction had been reached, it could be reprinted. In this manner the updating process became quicker and the information more 'relevant' or in today's search terminology 'fresh' (Chapters 5,9 and 10).

As with other cities, the office disappeared into the advertising department of the affiliated newspaper as shown by Michael Ambros's Innsbruck 'Fragamt' (ibid:175). His 'question office' first converted to the intermediate step of an ad journal (Anzeigenblatt) that was connected to an ad office (Anzeigenamts) which then became a newspaper edition (Zeitungsredaktion) (ibid:195). With the aforementioned office of Israel Saul announcements were not only made in the local newspaper's advertisement section (especially in Silesia) but were printed as large folio format posters and placed on prominent street corners (ibid:254-255). In contrast to the 18th century, in the 19th century these offices no longer published ad journals themselves but ran their ads in local newspapers such as the *Wiener Zeitung* (ibid: 262-263).

Founded as noble institutions that served the public, two centuries later the offices were viewed by the public as centres of ill-repute where fraudulent transactions occurred (ibid:264-265). The activities the offices continued to acquire a negative reputation because of their data gathering and profiling activities. Their 'intelligence force' aspect was increasingly seen as anachronistic, but remained amongst the offices' 'tasks' until 1849, when the official gazettes of the respective governments took their place (ibid: 239-240). Eventually the offices became competition to printers with their own productions or they published ads in local newspapers or eventually became newspapers in their own right (ibid:172-173). In the mid-19th century literacy was much more widespread and it was assumed that the newspaper readership of the 'imagined community' (Anderson 1983) encompassed not only the city and its surroundings but beyond.[16] Moreover, wealthy families could do without the assistance of search enquiries or servant services at the offices and instead turned to the same ads in newspapers (ibid:134).

Tantner ends his book with a telling excerpt from *Der arme Spielmann* (Poor Mr Spielmann), a novella written by Grillparzer in 1848 that describes the fate of the offices.[17] The story unfolds with Spielmann's dead father's secretary suggesting to him a brilliant plan for the realisation of an office for information, copying and translation services—for only 3000 guldens (ibid:264-265). Spielmann's beloved Barbara thinks it is a worthless idea because everyone learned how to read and write in school. The curtain closes when the secret is revealed in the right place and at the right time: the protagonists (Spielmann and Barbara) read in the local newspaper that the secretary has fled town leaving a trail of debt. Instead of finding information at the office, or the ad journal that no longer exists, one finds it serendipitously in the newspaper (ibid).

Perhaps, the denouement of the offices is similar to how search engines and online publishing on the internet has taken over the market from newspapers and printed journals. In any case, Tantner's study of early modern address offices shows that the concept of 'search' was around long before Google and demonstrates how these prototypical search engines used a variety of printed media to organise information. In his conclusion Tantner states that it is legitimate to regard these early modern offices as antecedents of today's internet search engines, or even as the first 'analogue' search engines (ibid:270-272). He draws on the concept of 'controlled anachronism', a term coined by Nicole Loraux (1993), which attempts to square the friction of the term anachronism and to give due recognition, in this case the 'search engine' within another epoch—to the early modern period (Tantner 2015:272). Therefore, as Tantner rightly remarks, investigating these now seemingly 'whimsical' institutions of the past makes it possible to raise new perspectives on contemporary problems and challenges and that alone is worth their recall (ibid:139).

[16] As demonstrated by Benedict Anderson's study *Imagined Communities* (1983), with the expansion of print capitalism (Blome 2010:208) those literate and living elsewhere in the world were also able to participate in society through text and language.

Chapter 2:

Organising the World's Information

1.0 Collecting information

Chapter 2 continues to answer my research subquestion, *how did search work in the past* by investigating past visions and attempts to organise the world's information. As elucidated in Chapter 1, 'human crawlers' gathered information and passed it on to interested parties or the address offices that created storage technologies for public and secret registers of personal data, 'indexing' the information (Tantner 2014:123). However, these offices were not the only actors collecting and storing data on subjects. In *Big Data* Mayer-Schönberger and Cukier (2013) present manifold examples of 'large and systematic campaigns of data collection ranging from censuses in ancient Egypt and China, to Renaissance bookkeeping, to 19th-century navigation' (Venturini et al. 2018:4196). Other examples include 'census bureaus' that actualised government control through constructed statistics, such as Louie XIV's administrator Jean-Baptiste Colbert's 'enquêtes', King Philip II's 'elaciones topográficas' and the harvesting of information by the Habsburg dynasty (Tantner 2014:123). The German word *Statistik,* coined by the political scientist Gottfried Achenwall in 1749, provided a blanket terminology for the 'science dealing with data about the condition of a state or community' (Stalder 2010) and 'would do for populations what cartography did for territory' (Davies 2017). In this way the organisation and processing of data by kingdoms and later state institutions in an 'analogue environment' was, moreover, an 'extremely labour-intensive affair that only massive bureaucracies were capable of conducting' (Stalder 2010). The US census in the late nineteenth century, which was supposed to be carried out once a decade, reached a 'critical juncture' when the overflow of data 'could not be finished before the next census was to be held' (ibid). They were also extremely difficult to carry out as many people were indigent, however, names and house addresses where noted whenever possible.

2.0 Universal paper machines

Harkening back to the offices, street addresses were collected in order to identity the inhabitants of cities and were noted in the register. This method of identification can be considered analogous to the 'call numbers' of books that indicate their physical location on the shelves of a library. Yet it is the shift from physical books on shelves to the systematic organisation by alphabetical order or subject catalogue that reflects changes in the 'search procedure' as 'a new type of mobile management' (Krajewski 2013:38). Instead of the book's location being the most important criteria, it is the conscription number, or call numbers that are a precise form of address, providing a 'logic of search' which progresses from walking along shelves [serendipitously] to accessing books via the card catalogue (ibid:39-40).

Markus Krajewski examines the card catalogue as an index pertaining to the origin of 'universal paper machines', imparting to the reader why a 'card index' is a machine with the 'possibility of rearranging its elements'—a process he describes as a 'chained mechanism' (ibid). (Figure 11) Yet this process embodies a particular type of machine—a search engine. Whereas a library catalogue contains the 'addresses of available books', a human, a librarian or 'readers themselves' mediate queries with the library expected to provide an answer and, in this way,

'the library serves as a collective search engine' (ibid:55). However, the 'communication structure of a collective search engine', in media theory terminology, is a 'network dialogue' (ibid), where 'the difference between the collective search engine and the learned box of paper slips lies in its contingency and the resulting possibility that queries in one's own terms can be posed in this strange arrangement' (ibid:57).

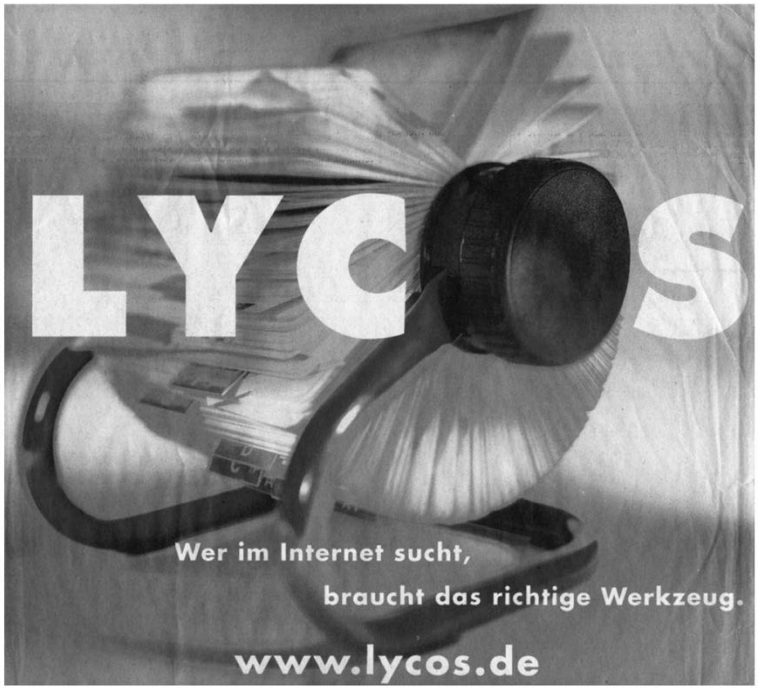


Figure 4.1
A search engine. ["If you search the internet, you need to have the right tool."] (From *Frankfurter Allgemeine Zeitung*, September 8, 1999, p. N3.)

Figure 11: Lycos was an early search engine and used the analogy of the rolodex (itself a portmanteau of roll and index). Screenshot from Markus Krajewski's book (2013:56).

The registers of the offices captured the circulation of ideas and money as 'an isomorphic logic of the representation of slips of paper and banknotes, pointing to their merging in card/data/banks'; card catalogues are also a means of note collecting and transactions (ibid). Scholarly research is comprised of methods, from the ability to find what one is looking for to notetaking and writing ideas down. In this way the scholar's 'small library' '*sorts addresses* so as to *address* thoughts' in contrast to the card catalogue that serves as a 'formal representational structure—namely an 'interface', which has a public function to guide users when they visit a library (ibid, emphasis mine).[18] Krajewski also points out another key difference between research and a search engine. The scholar determines whether or not to record a piece of information and to include it—as an idiosyncracy—whilst a 'search engine is designed to register *everything* randomly' (ibid, emphasis mine). Furthermore, it is the scholar's machine as a text generator, which enables the recombinatory linkages of ideas and 'forgotten memories' for cross-referencing—that reflects the 'excerpts with chains of references' (ibid).[19] The importance of these linkages is not only prevalent in research but search, which I will address shortly.

[18] The 'site map' is the modern-day cartography of websites.
[19] 'The reformulation of excerpts into new texts transforms the copyist into an artist. In Hegel's *Phenomenology of Spirit*, the absolute spirit appears as a 'hidden box of index cards' because his excerpted foundation vanishes in the unstated erasure of references' (Kittler 1997:197 cited by Krajewski 2013:62).

Figure 12: 'Employés du Mundaneum' - Image d'archives. Collection Mundaneum, Mons (Belgium).

3.0 Mundaneum

As shown by Krajewski's research, an index is needed in order to organise the world's information and there needs to be a categorical system for the user to access the contents. In 1895 Paul Otlet and Henri La Fontaine imagined the *Mundaneum*, a universal catalogue consisting of the documented knowledge of the entire world, considered by some as the world's first (analogue) search engine.[20] By the 1920s their vision was housed in Brussels where rows and rows of brown, wooden cabinets were filled with index cards containing individual elements of knowledge: books, newspaper articles, photos and other documents that had been extrapolated from all kinds of sources, including collated archives and libraries located elsewhere.[21] Originally called the *Universal Bibliographic Repertory*, people were able to

[20] 'The project was something like a paper Google, but developed decades before the Internet and without the benefit of computers' (Laaf 2011).

[21] Before the Nazi's dismantled it and replaced it with pro-Nazi art, it was a meeting place for scholars but also a huge archive. It is now housed in Mons, Belgium. Google scours the world looking for 'roots of the web' and makes cultural investments relative to its series of products in order to improve its public image. In return, 'The *Mundaneum* […] agreed to use Google's social networking service, Google Plus, as a promotional tool' (Laaf 2011). When I visited in November 2016 the staff informed me that they turned down this type of financing and

submit their queries from afar, using the quickest means of the day —the telegram—but mostly hand-written letters comprised the requests, sometimes 1500 per year. With this organisational system, 'librarians', who were predominantly women, searched the index cards and found answers to the queries.[22] (Figure 12)

For the *Mundaneum,* or 'mechanical, collective brain', Otlet and La Fontaine drew on Dewey's decimal system of classification (1894) as their basic structure. In 1904, with the help of researchers from Europe, England and the U.S., Otlet and Fontaine produced their own Universal Decimal Classification (UDC). This 'map' of knowledge unified personal classification systems and Dewey standardisations, expanding upon the catalogue in order to unite the various sciences to become 'in effect a highly complex database management system' (Rayward 2002:4). Although Otlet's project to catalogue, classify and structure knowledge was considered visionary, his universal taxonomy system along with his praise of the colonisation of the Congo and attempts by Belgium King Leopold II at 'civilising Africa' as well as his 'apparently benevolent interest in advancing "The African Issue" was fuelled by a firm conviction of the superiority of European culture and intelligence' (Constant 2020).[23]

This 'mechanical, collective brain' also incorporated the concept of 'documentation' and how it organises the world's information, which Otlet previously described in his *Treaty on documentation: the book on the book, theory and practice*.[24] In 1911 Otlet lectured on the future of books as 'containers of ideas', augmented by graphics and diagrams as well proposing how it could be 'dissected' (Otlet 1909:19 cited by Rayward 2002:3) such that 'each intellectual element, in corresponding to a physical element, will create a structure such that any combination of ideas, notions and facts will be possible' (ibid). For Otlet, the book needed to be 'transformed in some way' and this 'body' or corpus had various headings so that 'different searches using the same graphic elements on the cards are possible' (ibid:5). Besides organising and recording information, 'automatic retrieval' at any point in time mattered, with Otlet stating that '[documentation is] a vast intellectual mechanism designed to capture and condense fragmentary and scattered information and to disseminate it wherever it is needed' (Otlet 1909:11 cited by ibid). Various types of media including radio, images, microfilm and sound recordings along with traditional text elements comprised this 'documentation', which 'involved a complex of processes for the analysis, synthesis (what he also referred to as 'codification') and distribution of information through a network' (Van den Heuvel and Rayward 2011:4).

Beyond the technique of documentation, the manner in which information could be viewed, recombined and organised mechanically was the moment in which this new form of the book became the database (Rayward 1994). 'In effect, Otlet envisaged not only new ways of organizing knowledge to create a special kind of database, but also new ways of communicating or interacting with the database' (Van den Heuvel and Rayward 2011:4). Otlet envisioned a 'mechanical, collective brain' as a way to shift from the paper document or record, contained

exchange deal in order to have more control and autonomy over the *Mundaneum*. The location in Mons is not far from one of Google's largest data centres in Europe, to which I couldn't gain access. See Chapter 9.

[22] This 'city of knowledge' was conceived with 'electric telescopes' that enabled 'users' to search and browse the card catalogue.

[23] 'At several occasions, Otlet published racist statements dressed up as scientific facts, starting at the beginning of his career with *L'Afrique Aux Noirs* (1888) where he argued that white people or 'westernized' blacks were to be tasked with 'civilising' Africa. Similarly, in *Monde* (1935), near the end of his life, he claimed the biological superiority of white people [....]It neatly fitted the Enlightenment project that he was dedicated to and aligned with his self-identification as a liberal, a universalist and a pacifist'(Constant 2020).

[24] *Traité de documentation: le livre sur le livre, théorie et pratique*. Editions Mundaneum (Bruxelles 1934).

within an archive, to a database of objects of organised content. A paperless chain of disseminating research, information and thereby knowledge, this 'universal network' connected and combined a variety of contemporaneous media: index cards with telephones, telegrams with search queries and (200,000) postcards. Besides 16,000,000 index cards, by 1934 the *Mundaneum* contained an enormous quantity of images, with Otlet rearranging their relationships by organising a quasi-circular rotating structure to which other links were connected. (Figure 13)

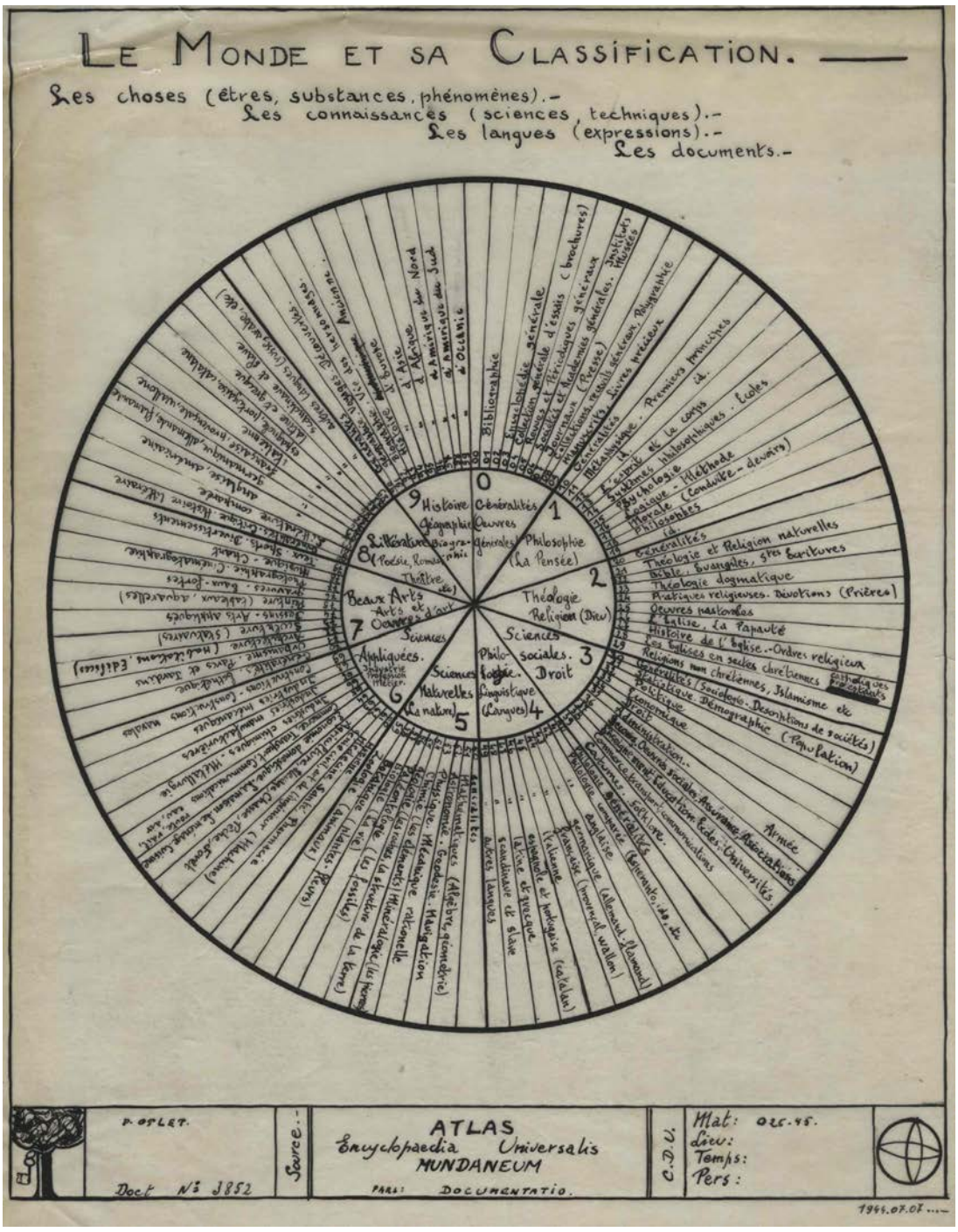


Figure 13: 'Le Monde et sa Classification', Atlas, Encyclopaedia Universalis, Mundaneum by Paul Otlet. Collection Mundaneum, Mons (Belgium).

This ‘hyperlinked’ structure wasn’t just to connect various pieces of data with one another—Otlet wanted the links themselves to have meanings, or ‘associative indexing’. Moreover, Otlet’s ‘mechanical, collective brain’ was designed to not only store all of the world’s information as a centrally-structured body involving libraries, archives, museums and other types of ‘offices of documentation,’ but to organise it in a standardised way for processing the information. By rethinking the format of a book, Otlet envisioned a collaborative process that would improve upon the production of knowledge and its dissemination. According to Van Den Heuvel and Rayward, books, as well as the diagrams and images drawn by Otlet, ‘were effectively interfaces that Otlet created in trying to visually get a grip on problems of scalability, representation, and perception of relationships between classes of knowledge objects that might be of interest today’ (ibid:1). The mechanical element of searching and spreading information along with its integration into public knowledge reflected the future of the book as a *machine à penser* (machine to think with) (Otlet 1911:291 cited by ibid:3).

4.0 World Brain

Echoing Otlet’s and La Fontaine’s award for the Noble Peace Prize in 1913, H.G. Wells also wanted to bring unity and establish peace by organising the world’s information and disseminating this knowledge. In 1937 his *World Brain: The Idea of a Permanent World Encyclopaedia* questioned the very subject of what constitutes an ‘encyclopaedia’ and to what extent it reaches a wider audience, instead of being written by ‘gentlemen’ for reading by ‘gentleman’. Originally published in the new *Encyclopédie Française,* Wells’s treatise argues that universal access rests on the distribution of this ‘world brain’ of knowledge, which hasn’t kept up with the automobile and the aeroplane. The responsibility then lies with the bodies that reflect the ‘intellectual life of mankind’: universities, libraries and schools. The notion of accessibility, along with the ‘foreshadowing’ of a ‘real, intellectual unification’ of the human race would be encompassed in this ‘all-human cerebrum of knowledge, ideas and achievements’ (Wells 1937).

In organising the world’s information, Wells conceived of a certain ‘order scheme of reference and reproduction,’ which could encompass a vast amount of ‘manageable well-ordered facts’ (ibid). What he envisioned was an updated index of human knowledge, compiling heterogeneous resources with micro-photography to create ‘a concentrated visual record’ (ibid). Wells used the term ‘Permanent World Encyclopaedia’ to describe the technological developments that made possible a ‘synthesis of bibliography and documentation with the indexed archives of the world’ (ibid). The storage of such information would require a small amount of space, in this case ‘microfilm’, which could be reproduced and then enlarged with a magnification lens and projected onto a wall to be viewed by students, or the general public.[25] In this way this compact ‘efficient index’ would have a scope that could have mondial access, sharing human endeavour but also serving as a peace-making device.

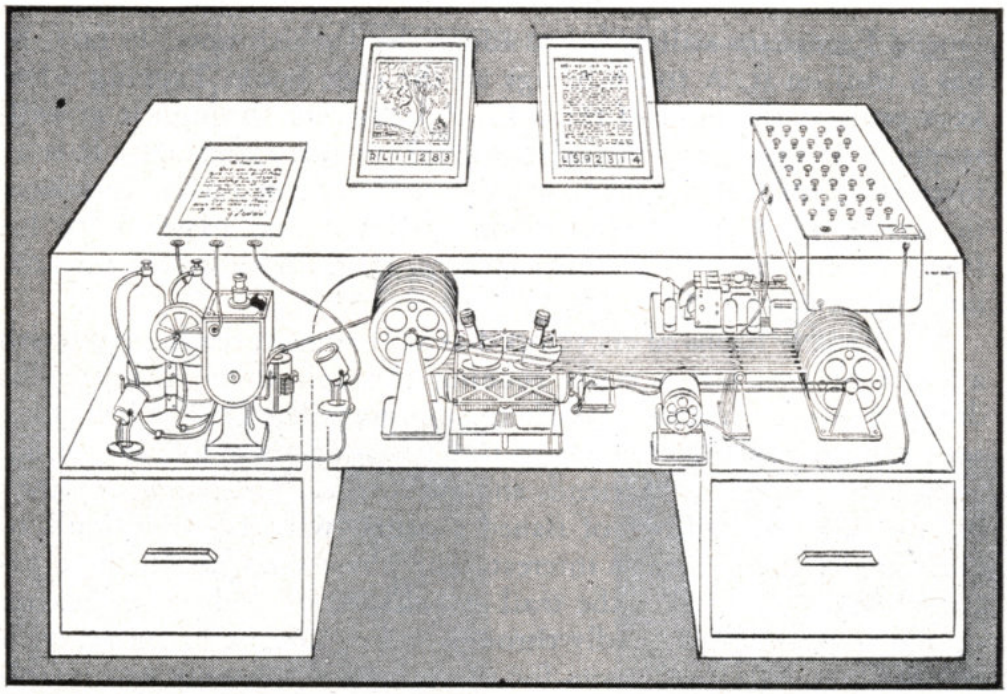
[25] As the anonymous author notes in a footnote from a now defunct Swedish project, ‘If you just change “microfilm” to “computer files”, it is amazing how accurate Wells’ prophecy was.’ This page is a local copy of the original created at the Swedish Origo project. https://sherlock.ischool.berkeley.edu/wells/world_brain.html

5.0 Memex: envisioning a search engine

Vannevar Bush was a controversial figure, heading military research that was united with universities whilst the head of the US Office of Scientific Research and Development (OSRD) during the Second World War, later MIT. At the end of WWII, in 1945, Bush concomitantly expressed concern over the use of technology to make 'machines of madness' and instead imagined a collective memory apparatus that would promote understanding. His seminal text, *As We May Think* described the 'proto-hypertext' mechanism:

> Consider a future device for individual use, which is a sort of mechanized private file and library. It needs a name, and, to coin one at random, 'memex' will do. A memex is a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory (1945:9).[26]

Bush then explicates how he envisioned *Memex*, an acronym for 'memory extender' (Bory et al 2016:5), as somewhat similar to a desk, yet with a keyboard and a series of buttons and levers. The reading material (business letters, newspapers, books and images) is projected onto slanting translucent screens and the question of storage is resolved with the addition of 'microfilm', which can be conveniently purchased, easily inserted and holds the content. Yet this is only one small part of *Memex*, as the rest is for the mechanism. (Figure 14)



MEMEX in the form of a desk would instantly bring files and material on any subject to the operator's fingertips. Slanting translucent viewing screens magnify supermicrofilm filed by code numbers. At left is a mechanism which automatically photographs longhand notes, pictures and letters, then files them in the desk for future reference.

Figure 14: Memex design published with Bush's article, As We May Think*, Atlantic Monthly* (1945)

[26] The Department of Homeland Security in the US now uses the name *Memex* for their search engine that only searches the 'Dark Net' and looks for criminal activities, which I will discuss in Chapter 7.

> On the top of the memex is a transparent plate. On this are placed longhand notes, photographs, memoranda, all sorts of things. When one is in place, the depression of a lever causes it to be photographed onto the next blank space in a section of the memex film, dry photography being employed. There is, of course, provision for consultation of the record by the usual scheme of indexing. If the user wishes to consult a certain book, he taps its code on the keyboard, and the title page of the book promptly appears before him, projected onto one of his viewing positions. Frequently-used codes are mnemonic, so that he seldom consults his code book; but when he does, a single tap of a key projects it for his use. Moreover, he has supplemental levers (Bush 1945).

Reminiscent of a copy machine, combined with the ability to search documents, the memex machine would make knowledge more accessible, transforming scattered documents into a library, or a global encyclopaedia. However, at that time Bush was wary of indexing information because of the difficulty of retrieving information given the artificiality of indexing.

> When data of any sort are placed in storage, they are filed alphabetically or numerically, and information is found (when it is) by tracing it down from subclass to subclass. It can be in only one place, unless duplicates are used; one has to have rules as to which path will locate it, and the rules are cumbersome. Having found one item, moreover, one has to emerge from the system and re-enter on a new path (ibid).

Although links can be connected through different means, this 'associative indexing'—'the basic idea of which is a provision whereby any item may be caused at will to select immediately and automatically another' (ibid). This harkens back to the *Mundaneum* but becomes the 'method', a 'hodology' or path that leads to relationships and connections between items, which I will address in Chapter 3. 'It affords an immediate step, however, to associative indexing. This is the essential feature of the memex. The process of tying two items together is the important thing' (ibid).

Bush goes on to sketch out the sequence of documents, which, formed together, become a 'trail' that can then be connected with other 'trails', comprised of numerous items and even comments from the user who was building the trail! Using the levers, the specific trail could be called up and, because it was on microfilm, inserted into another's memex machine.[27] In this way, Bush conceived of an invention able to deal with vast amounts of stored documents, items or images centred around 'filtering' as 'associative trails' or, as he predicted, 'selection by association,

[27] 'Wholly new forms of encyclopaedias will appear, ready made with a mesh of associative trails running through them, ready to be dropped into the memex and there amplified. The lawyer has at his touch the associated opinions and decisions of his whole experience, and of the experience of friends and authorities. [Fusing research and search, Bush might be referring to *Shephard's* for law citations.] The patent attorney has on call the millions of issued patents, with familiar trails to every point of his client's interest. The physician, puzzled by a patient's reactions, strikes the trail established in studying an earlier similar case, and runs rapidly through analogous case histories, with side references to the classics for the pertinent anatomy and histology. The chemist, struggling with the synthesis of an organic compound, has all the chemical literature before him in his laboratory, with trails following the analogies of compounds, and side trails to their physical and chemical behavior. The historian, with a vast chronological account of a people, parallels it with a skip trail which stops only on the salient items, and can follow at any time contemporary trails which lead him all over civilization at a particular epoch. There is a new profession of *trail blazers*, those who find delight in the task of establishing useful trails through the enormous mass of the common record. The inheritance from the master becomes, not only his additions to the world's record, but for his disciples the entire scaffolding by which they were erected' (Bush 1945).

rather than indexing, may yet be mechanized' (ibid). Moreover, these 'trails' are analogous to the human brain, like an intricate web of thoughts that are connected by association and called up by cells.

> Trails that are not frequently followed are prone to fade, items are not fully permanent, memory is transitory. Yet the speed of action, the intricacy of trails, the detail of mental pictures, is awe-inspiring beyond all else in nature (ibid).

Although he admired the mechanical workings of the human mind, *Memex* held the items permanently, which could be 'resurrected from storage' at any time, foreshadowing search engine and information retrieval. Bush was forward-looking in regard to the development of mapping neural networks of the brain as a paradigm for AI (artificial intelligence), which I will address in Chapter 5 and 10 in relation to Google's machine-learning search algorithms.

6.0 Hypersearch: Scientific Citation Index

In 1952 Eugene Garfield first wrote down his thoughts on 'The Mechanization of Indexing' and Hollerith machines and in 1955 published an article about the 'citation index', mainly designed to solve information retrieval problems. He was inspired by the law citation index, 'Shephard's Citations' (1873), a system that was chronologically organised—the most recent cases were most relevant and therefore held the most authority. Garfield also realised the value of the evaluation process, comparing it to a key that unlocks the store of law and arguing that if such an index were applied to technological patents and scientific publications, '[a]n article on any scientific subject would be the key to all others' (Garfield 1954 cited by Mayer 2009:64). By focussing on scientific journals and patents, he believed that his Scientific Citation Index (SCI) could 'provide an entry point to the entire scientific landscape, as a tool of analysis and an instrument of evaluation like a "hyper search engine"' (ibid). According to Katja Meyer, the SCI produces a socio-structural model of publication behaviour, turning the footnote into the most important source of information (2009:65), with the SCI not measuring content but links or references and their reputation.

The SCI is based on the number of citations attributed to an article and contributed to the development of 'bibliometrics' (also called scientometrics) that concerns the study of scholarly publications and journals applied to research outputs. In 1965 Garfield's database contained 1057 journals, almost 50,000 patents and more than 2 million 'access points to the 'world's science and technology literature' on punch cards that enabled automated processing and statistical analysis (Garfield 1965 cited by ibid). It is however, the infamous 'impact factor' that became the 'most popular and controversial journal bibliometric indicator', measuring 'for a given journal and a fixed year as the mean number of citations in the year to papers published in the two previous years' (Franceschat 2010:5). Derek de Solla Price's work (1965) showed the probability effect in which the more a paper or article is cited the more it will be subsequently cited, resulting in a significantly greater likelihood of references being made to works that were already popular. 'Thus, the number of references, referred to as 'edges' in graph theory, generates visibility and recognition' (Mayer 2009:66). Applied to the academic world, this methodology of hierarchically organising information became the status quo for authorship and along with it, authority. Moreover, this innovation is key to understanding how academic

research is linked to contemporary ‘link culture’ and *search* engine development in regard to PageRank that I will illuminate in Chapter 3.[28]

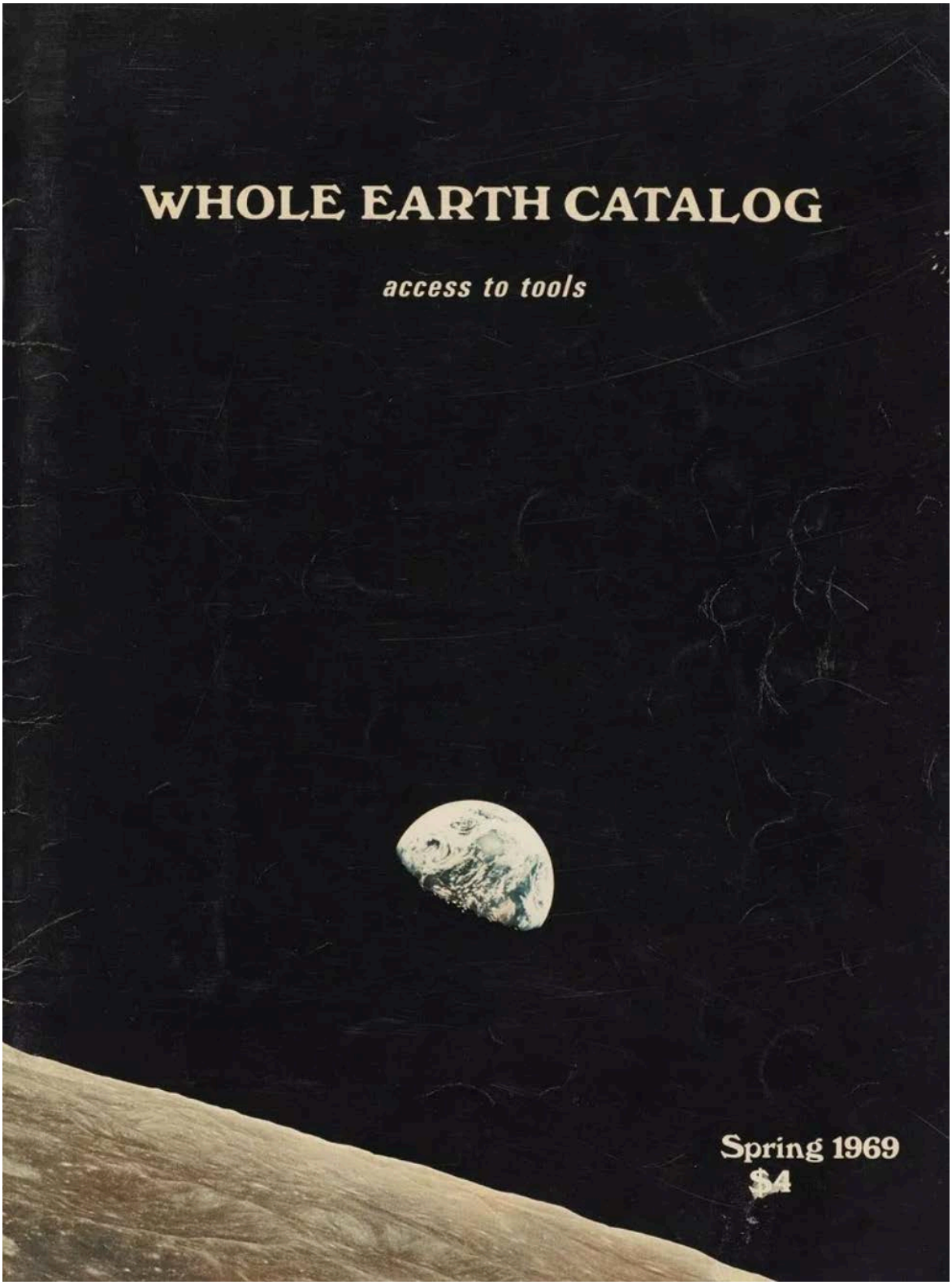


Figure 15: Cover of WEC, Spring 1969

[28] Michel Foucault is still the most cited author in western social sciences.

7.0 Whole Earth Catalogue

By 1968, as exemplified with the publication of the Whole Earth Catalogue (WEC) by Stewart Brand and the Portola Institute in San Francisco, manual indexing was not limited to legal citations and academic references. The cover featured an image of the earth taken from the ATS-3 satellite in 1967 that Brand had lobbied to obtain by printing buttons with the text, 'Why haven't we seen a photograph of the whole Earth yet?' in 1966 (Figure 15). The answer was to organise the world's information into the WEC, which had 'access to tools' as its subtitle. Reflecting the back-to-the-land movement, its contents were not a traditional mail-order catalogue but a 'smorgasbord of books, mechanical devices, and outdoor recreational gear' (Turner 2006:71). This panoply 'offered a cacophony of artifacts, voices, and visual design' organised into seven categories: Understanding Whole Systems, Shelter and Land Use, Industry and Craft, Communications, Community, Nomadics and Learning (ibid). Phrased otherwise, the WEC 'offered their readers the chance to encounter information and perspectives that might change their thinking or behaviour' (ibid:92). The content also reflected the communes, communities and scenes in which Brand operated, with the 'counterculture movement' playing a key role, combined with academic input from various figures along with seeds of the blossoming technological revolution.

Fred Turner's *Counterculture to Cyberculture* (2006) epitomises Brand's branding of the WEC as a 'research organization' by his declaration that it was 'designed as a system. I knew about systems. I had studied cybernetics' (ibid:79). Turner's book avidly portrays how the counterculture transformed into cyberculture, along with the role that cybernetics played in the holistic approach to WEC. As Turner puts forth, changes originated from those outside the computer industry (ibid:106). This counterculture movement imbibed 'information's transformative potential' with a confluence of various hobbyists, computer engineers, programmers, 'hackers' and academic labs, including the Homebrew Computer Club (ibid:103). It embodied an entrepreneurial spirit of DIY (Do-it-Yourself) as a 'networked forum'—executing principles of system theory as well as organisational properties within its structure as a 'whole system' and a 'tool' for readers that could improve the world with its usage (ibid:84). Gathering together participants in communes (the movement at the time contained around 10,000,000 Americans), the media of the WEC was diverse enough to offer information regarding manifold tools and techniques to trade. In their communes, readers could search the catalogue for what they needed—from eastern philosophy to military gear for survival in rough climate and conditions.

> When these groups met in its pages, the Catalog became the single most visible publication in which the technological and intellectual output of industry and high science met the Eastern religion, acid mysticism, and communal social theory of the back-to-the-land movement. It also became the home and emblem of a new, geographically distributed community (ibid:73).[29]

Although geographically separated, the searching and reading of such a catalogue could occur anywhere, with readers submitting orders by post or physically attending the gatherings announced in its pages. The textual networked forum would also supplant images of previous events and readers could write in with suggestions and publish their comments and requests

[29] It also exemplified a DIY 'home schooling' for the counterculture who were simultaneously synthesising Native American and Settler ways of life, being linked to each other through the medium of the WEC.

under the umbrella of a united 'imagined community' (Anderson 1983) or a 'town square' (ibid:89). Turner elucidates this holistic cybernetic system of exporting rhetorical universalism as ideology, which is transcribed into every page of the WEC. Searching for ways to inject ideal models of community into the world at large, along with technologies and information systems, it also attempted 'to legitimate mainstream forces of consumption, technological production, and research as hip' (ibid:84).

With the WEC, it was 'the ideal relationship between information, technology, and community' that drove collaboration. Particular figures, such as Douglas Engelbart, stand out. The vision of the Augmentation Research Center (ARC) at the Stanford Research Institute (SRI), Engelbart's place of employment, was collaborative, between each of the members but also between human and machine. Echoing Bush's vision of the *Memex* search engine, individuals would be able to retrieve texts and to manage and organise information themselves. Within this social organisation, the more people participated, the greater the comprehension of each user in the system, along with the 'process of collective feedback facilitated by the computer' (Turner 2006:108). Engelbart gave birth to not only the infamous mouse of individualised freedom, but the QWERTY keyboard, the CRT terminal and a computing system for office work called the On-Line System, or NLS. 'The system offered its users the ability to work on a document simultaneously from multiple sites, to connect bits of text via hyperlinks, to jump from one point to another in a text, and to develop indexes of key words that could be searched' (ibid:108). As Turner is keen to point out,

> like the hyperlinked texts of Engelbart's system, the Whole Earth Catalog presented its readers with a system of connections. In the Catalog, no text stood apart from every other; each was part of an informational or social system, and each offered a doorway through which the reader could enter one of those systems (110).

Another key figure was Alan Kay, who was employed at Xerox Parc research laboratory and envisioned an interactive desktop computer, the 'Dynabook'—as a 'language machine where content was the description of things' (ibid:112). According to Turner, when Kay 'saw the Catalog, it offered him a vision of how an information system might organize that content' (ibid). The linkage was that the Catalog was an information tool, or an analogue computer that organised content, yet it was also a hyperlinked system. Kay described the WEC 'as a print version of what the Internet was going to be' (ibid). Key to Kay's understanding of the conception of the Catalog was that it was serendipitous, offering ideas to the reader (user) when they didn't know what they were searching for.[30] In hindsight, the WEC set out to change how the world's information was organised by making it accessible and on the other hand, providing tools allowing mass consumers, i.e. users, to 'intervene'. The hacker's ability to be in control of interactions with computers, including the tools that could organise the world's information and make it universally accessible, imbued the user with agency, a 'tools can change the world' mindset which I will address in Chapter 8 and 11.[31]

---

[30] Therefore, the reception of every new iteration of the Catalog was marked by excitement and copies of all of the books at the Whole Earth Truck Store became part of the library at Xerox PARC.

[31] In 1972 it was Brand again who became the 'trendcaster' having picked up on the Californian 'hacker' vibe with his Rolling Stone text, 'Spacewar: Fanatic Life and Symbolic Death among the Computer Bums' where he pitted the bureaucratic planners or 'squares' (those having normative values) against those of the figure of the 'hacker' who were 'cultural revolutionaries' not just technicians. 'Like P. T. Barnum, [Brand] had gathered the performers of his day—the commune dwellers, the artists, the researchers, the dome builders—into a single circus. And he himself had become both master and emblem of its many linked rings' (Turner 2006:101). The etymology of search

It was not only the world at large and the biosphere but the intimate sphere of the self that was intended to interface with the catalogue. 'The reader could order the "tools" on display and thus help to create a realm of "intimate, personal power" in her or his own life (albeit by entering the commercial sphere first)' (ibid:83). Particular items required specific skillsets that transformed the readers as subjects through the WEC ideology, a point I will return to in Chapter 8 with the Interpellated Subject. Not only was the reader transformed but the development of the 'tools' that the Whole Earth Catalogue offered—from mainframe computers to personal computers, were themselves affected. Historian Paul Ceruzzi describes the transitions between 1959-1969 stating that 'small computers, microprocessors, keyboard-based interfaces, individual usability, and the sensation of interactivity, were all in place by 1972' (ibid:105). This development stems from shifts in habits and patterns by users, where following Thierry Bardini (2000), Turner explains that the 'dynamic of personalization' has been ongoing since the 1940s.

> After all, wasn't it the Whole Earth Catalog that had set out to liberate technology from its corporate and governmental contexts? And wasn't it the Catalog that had promoted the notion that the right tools, properly used, could help reform society? Could perhaps even save the "mass" economy by "personalizing" it? (ibid:138)

During the process the emphasis lay on the transformation of the individual 'user's' consciousness into a creative being (ibid:84,93). Yet it was not only about creativity. It 'opened windows onto the universal order of things, the items in the Catalog promised to be "personal" technologies as well' (ibid:92). With these citations Turner verifies that the 'dynamic of personalization' was already an inherent part of counterculture and the cyber community and that they complimented each other (ibid:106). Hobbyists, WEC and computer engineers were all interested in personalised tools and products. Eventually search engine start-ups and companies took on key roles in the 'personalisation' of information, which I will discuss in Chapter 5.

At the forefront of those who were empowered to change the face of society and improve the lives of individuals by WEC tools and technology were a couple of individual hackers (Steve Wozniak and Steve Jobs). 'Later on, at a Stanford commencement speech (Google's founders alma matter) Jobs claimed that the WEC was one of the 'bibles' of his generation, as a sort of paperback Google, 35 years beforehand' (ibid:109).[32] The image of the 'blue marble' mirrored Brand's attempt to create a holistic approach to life with the WEC where the individual played a central role, looking down upon the earth as a 'god', responsible for whatever would come out of the know-how (information) that was contained within the catalogue's pages.[33]

> At one level, the Catalog was a "Whole Earth" in its own right. That is, it was a seemingly comprehensive informational system, an encyclopedia, a map. Simply by picking it up and flipping through its seven sections, the reader could become an astronaut looking down from space on a textual representation of a new earth (ibid: 83).[34] (Figure 16)

---

is the same as 'circus' and this circus or 'readership capitalism' predicted the 'networked computing' of the 1990s, which in turn laid the groundwork for the beginnings of the 'New Economy' that later took hold.

[32] Full speech available here: https://www.youtube.com/watch?v=D1R-jKKp3NA&t=12m45s

[33] Stewart Brand also created one of the first virtual networks, the Whole Earth 'Lectronic Link, otherwise known as *WELL*.

[34] However, in spite of its holistic approach, Stewart Brand, as the editor in chief, decided virtually everything that went into the catalogue and what was left out. Although drawing on their traditions and customs, there were no Native Americans (except Brand's wife who was half Ottawa), women were visualised as sex objects or relegated

Figure 16: *Earthrise*, taken on December 24, 1968, by Apollo 8 astronaut William Anders

8.0 Academic search: WWW, Hypertext, FTP, Browsers

In 1968 Doug Engelbardt demonstrated the first 'hypertext' 'oN-Line System' at the Fall Joint Computer Conference in San Francisco, where he typed in a word that simultaneously appeared on a screen at the Stanford Research Institute in Menlo Park.[35] As well as producing nearly all the equipment (raster scan video monitor, mouse, computer graphic interface) for this 'Mother of all demos' (Levy1994), Engelbardt and his team showed the possibilities of human/computer interaction and personal computing for non-programmers. In March 1989 Tim Berners-Lee, then employed at CERN, proposed an information management system called an 'Information Universe' (Berners-Lee et al. 1992).[36] In November the following year he submitted a document to CERN with Robert Cailliau called 'World Wide Web': Proposal for a HyperText Project' that explained how a user could browse and find accessible information through linking.

Concerned about the loss of information, Berners-Lee wanted to implement at CERN the possibility of interconnecting websites and decentralized nodes for organising information onto one screen, designating documents and other resources as a shared system that are then

---

to nursing children (social reproduction) and there were few references to the Vietnam war. 'In the fall of 1974, not long after the Catalog had officially ceased publication, Brand did turn over an issue of the magazine that grew out of the Supplement called CoEvolution Quarterly to the Black Panthers to edit as a special issue. The magazine they created simply copied the format of their newsletters, and none of the editors or authors of that issue became regular contributors to Whole Earth productions in later years' (Turner 2006:97).

[35] Funded by ARPA, NASA and the US Air Force

[36] The first website went online on August 6, 1991: http://info.cern.ch/Proposal.html. The first webpage address: http://info.cern.ch/hypertext/WWW/TheProject.html

'hyperlinked'. Berners-Lee merged 'a communication network linking various computers, an idea originated with the American Arpanet' or a 'distributed network' with 'hypertext', 'in which any type of information would be reachable from anywhere and directly linkable with any other content' (Bory et al. 2016:5).[37] His 'proposed mesh' became what is now known as the World Wide Web, drawing upon Theodor Nelson's coinage 'hypertext' (1967) and his 'jumping metaphor'.[38] (Figure 17)

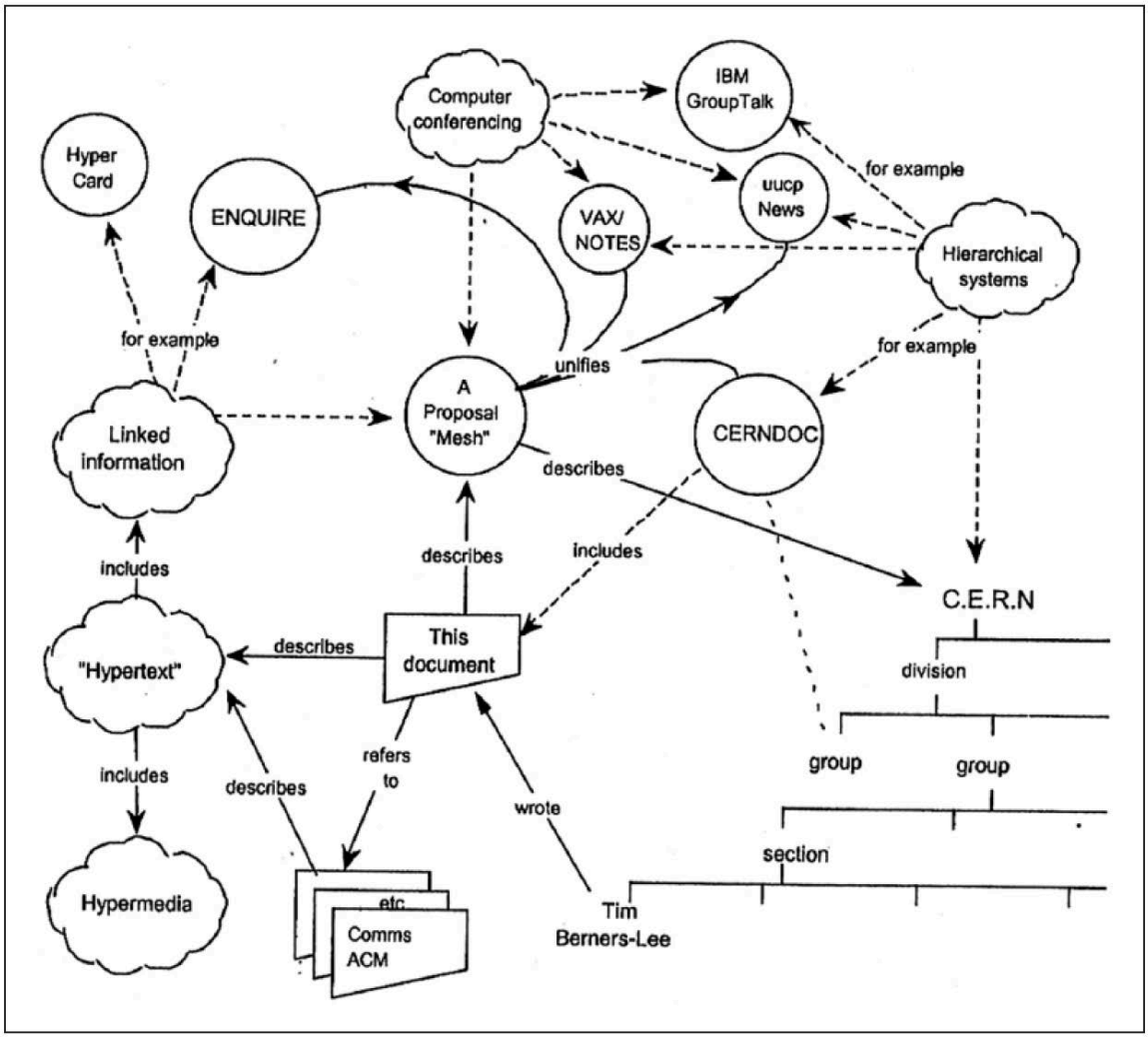


Figure 17: Berners-Lee proposal March 1989 with a jumping-link model. 'A Proposal "Mesh"' is what he then called 'the web' (Berners-Lee 1989:1). https://www.w3.org/History/1989/proposal.html

In their writings both authors promoted the 'non-linear organization of information', 'the manipulability of documents' along with 'the difference between document and information retrieval' and lastly, 'the need for shared and compatible standards for all computing systems' (Bory et al. 2016:7). Berners-Lee saw the web as a new space, not as a replacement for Nelson's hypertext but as a 'supplementary system, something to be added to any workstation to easily

[37] Berners-Lee had already written a programme in 1980 called Enquire, 'a personal information management program that is considered the precursor of the WWW' (Bory et al. 2016:5).
[38] 'Although familiar with hypertext, having written Enquire in 1980 and knowing HyperCard, Berners-Lee used this term for the first time in his 1989 proposal: "I first made a small linked information system, not realizing that a term had already been coined [by Ted Nelson in 1965] for the technique: Hypertext"' (Bory et al. 2016:7).

retrieve information' (ibid:7-9). It was the promotion of 'user-friendliness' as a 'faster and easier way of reaching existing documents and data' (ibid:6) that sold the idea of the web to CERN, with Berners-Lee twisting George Eastman's (Kodak) famous slogan to '"click the mouse, we do the rest," without shaping the way in which information is produced' (ibid:7).

Later that year HTTP (HyperText Transfer Protocol), the coding language that enables the user to jump from webpage to webpage, emerged. 'HyperText is a way to link and access information of various kinds as a web of nodes in which the user can browse at will. Potentially, HyperText provides a single user-interface to many large classes of stored information such as reports, notes, data-bases, computer documentation and on-line systems' (Lee and Cailliau 1990).[39] Its architecture was based on 'four technical components: Hypertext Transfer Protocol (HTTP), uniform resource identifiers (URIs, sometimes called 'locators' or uniform resource locators [URLs]), Hypertext Markup Language (HTML), and hyperlinks (encoded in HTML)' (Berners-Lee and Fischetti 1999; Jacobs and Walsh 2004 cited by Plantin et al. 2017:301-302).[40] Comprised of content and HTML (HyperText Markup Language) webpages allowed connectivity between various parties all over the world and anyone could access the World Wide Web, which became available for people outside the academic community in 1991. However, in the late eighties the 'manual' still played a key role in navigating what was then known as the 'pre-web'. The manual 'pre-web' concerned the publishing of personal bookmark files as web pages and storing these collections of URLs that accrued value, along with sharing them between groups of people.[41]

> The collaborative filtering and tagging sites that are popular today descended from this practice, and the updating and annotating of links to interesting new websites led to some of the first proto-blogs. Most importantly, it gave rise to the first collaborative directories and search engines (Halavais 2009:22).

This type of search consisted of '[t]he first indexes on the internet [that] were created by hand, often by the users of the systems as a guide to others' (ibid:21). In the early 1970s, whilst a student at MIT, Abhay Bhushan developed FTP (File Transfer Protocol) that enabled the secure transfer of files between ARPANET's servers and computers. Subsequently File Transfer Protocol (FTP) made possible the moving of files between computers and entailed choosing the FTP from a list of servers. The user acquired this list of public servers from friends or colleagues and there was often 'a text document that could be downloaded that briefly summarised the content of each of the files on a given server' (ibid). Eventually it became cumbersome to search the FTP servers and as their number enlarged so did their inconsistencies. 'While the increase in content was a boon to those who used the internet, it became increasingly difficult to locate specific files' (ibid).[42]

Simultaneously universities were joining the network and the creation of *Archie,* a file-transfer-protocol (FTP) crawler designed at McGill University in Canada in September 1990, changed

---

[39] 'A program which provides access to the hypertext world we call a browser' (Berners-Lee and Cailliau 1990).
[40] 'However, this 'Open Web' as it became known competed with the already established dial-up services from the 1980s, such as CompuServe, Prodigy, and American Online that were 'walled gardens' (Plantin et al. 2017: 301-302).
[41] URLs today have value, as shown by my data visualisations, *Re:search - Terms of Art* (Appendix D), or how Tor's hidden services are distributed between users in the know, as with the Hidden Wikis for the 'Dark Net' (Appendix I)or the public sharing of URLs via Twitter.
[42] FTP continues to be used today as a way of transferring files.

everything. It searched and 'indexed' FTP sites and directories, which is why it could be considered the first search engine. *Archie* was not searching the entire document nor discovering servers that were linked together but instead focussed on the titles of the files.

> Nonetheless, it represented a first effort to reign in a quickly growing, chaotic information resource, not by imposing order on it from above, but by mapping and indexing the disorder to make it more usable (ibid:21-22).

Many have compared this description to the way the Web has grown, almost in a haphazard manner, collating things randomly from disparate sources or 'like a library that consists of a pile of books that grows as anyone throws anything they wish onto the pile' (ibid:22-23).

In 1992 a browser called *Lynx* evolved that used hypertext links in documents. *Erwise* was a graphical browser using the 'libwww' or the library of the World Wide Web, developed by a group of master students in Helsinki in 1992 but was never funded to advance further. One of the in-between steps transitioning from browsing files to the early beginnings of the internet was the *Gopher* system, seen still by some as an alternative to the World Wide Web, which 'facilitated working through directory structures, and insulated the individual from a command-line interface' (Halavais 2009: 22-23). Yet *Gopher* lacked what was called 'hypertext' so, in 1992, *Veronica* was placed on *Gopher* servers as a crawler to search 'menu-structured directories' (ibid).[43] Created in 1992, *ViolaWWW* stands for 'Visually Interactive Object-Oriented Language and Application' and was the first browser to add extended functionality such as embedded scriptable objects, stylesheets and tables. Eventually *ViolaWWW* lost out to another graphical browser called *Mosaic* that was released on April 22, 1993. *Mosaic* was a kind of central nervous system, which provided users with full-colour, graphic webpages and, more importantly, a visual understanding of networked webpages that were both fun and intuitive to surf (Calore 2010). This browser enabled not only geeks but also users from around the world to have access to the web and it was subsequently ported to Microsoft Windows, making it popular. Traffic increased on the World Wide Web from around 500 known servers in 1993 to around 10000 in 1994 with *Mosaic* the predominant means for searching the web.

9.0 The rise of commercial enterprise: Web directories, meta search, portals

Very soon thereafter the problem of locating relevant documents within the burgeoning information space emerged. The web lacked a built-in indexing and categorizing mechanism as it was only threaded by one-way hyperlinks. As the space expanded exponentially, simply following links almost immediately became impractical and books once again became the inspiration. Keywords exist in book indexes, where words appearing on pages are listed at the end of the book and are also an 'inverted index'—they reverse the process so that the reader can find which pages contain a given word. Applying this structure from books, search engines would need to build an index in order to process the information (data) but there were questions as to which parts should be indexed: page titles, metadata, hyperlink text, or full text (Yuwono et al. 1995 cited by Halavais 2009:17).[44] Thus, indexing and categorising were created as an extra service within this space, rather than being part of the protocols that created the space in the first place and the providers of such indexes have been prominent actors from early on.

[43] Archie's girlfriend in the comic book series.
[44] 'If indexing the full text of a page, is it possible to determine which words are most important?'(Halavais 2009:17).

Considered to be the web's first 'robot', in June 1993 the *World Wide Web Wanderer* was initially designed to measure the size of the web and it existed until late 1995. Written in Perl, the *Wanderer* wandered through the web of hyperlinks, indexing titles and created an index called *Wandex*. With crawling, these 'indexers' discovered new documents and chose what to index whilst building archives and deciding how items should be structured and organised through parsing. Two students at Stanford (Jerry Yang, David Filo) released Yahoo! in 1994 that offered a very familiar looking directory compiled by experts which resembled a library catalogue. While this traditional format made the relatively unknown space of the web seem less alien to many, it quickly ran into deep problems, both in terms of scale, with the impossibility of keeping up with the growth of the web and by its ontology, the categorical system could not contain the complexity and dynamism of the information space it claimed to organise. In 1994, PhD student Brian Pinkerton's *WebCrawler* was the first search engine to index 'full text search' and originally had its own database, with ads in separate areas on the page.[45] 'Receiving its millionth query near the end of 1994, it clearly had found an audience on the early web, and, by the end of 1994, more than a half-dozen search engines were indexing the web' (ibid:22-23). In 1995, *AltaVista* also appeared as a full-text search engine built on automated information gathering and indexing and because it was faster and more comprehensive, it quickly overtook the human-compiled directory. It also established the now standard interface paradigm of a relatively empty page with a simple search box, in which users could enter a query and receive a ranked list of search results.

During the 1990s search engines developed faster processing power and more storage space. The key factor was rapid growth—like the economy—that facilitated the expansion of the index, or database. The design of these infrastructures had to be implemented and competition existed between companies over who had the largest database—the size of the engine or directory index—and who could update search queries most quickly, or the retrieval speed (Van Couvering 2010:97). The early search engine era consisted of two alternative models of service provision: the web directory (*Yahoo!, Magellan*) pioneered editorial ratings and *LookSmart* provided groups of sites that were categorised, and in some cases rated, by an editorial team (ibid). 'The Open Directory Project, by releasing its volunteer-edited, collaborative categorization, provided another way of mapping the space' (Halavais 2009:24-25). These implementations offered the user the ability to browse their directories and to be able to search 'effectively', as there were ready and available indexes that had been selected, filtered or 'curated', sometimes by hand (ibid).

> The second model was much more complex technically, and involved automated technology to browse websites, store them in an electronic index, and automatically retrieve them based on user queries. These were more properly called engines. (Van Couvering 2010:97). (Figure 18)

In 1996 there were already 'metasearch engines'—aggregators that culled results from various sources. Drawing on search queries from users, they then sent out their own queries to search engines and algorithmically structured the results, or 'ranked' them and displayed them back to the user. Once again, the university was the site of innovation with Daniel Dreilinger's *Savvysearch* at Colorado State University aggregating around 20 different search engine results. There was also *Metaseek*, a context-based search engine for images, and *HotBot* that combined results from *Inktomi* and the *Direct Hit* database where the user could search within the search

[45] Nowadays it a metasearch engine with sponsored and non-sponsored ads. https://www.webcrawler.com/

Table 3: Early period search engine dates, institutions, and founders.

Engine/ Directory	*Date went live**	*Institution (Location)*	*Developer(s)*	*Position at time of development*
Yahoo (directory)	Feb 94	Stanford University (Palo Alto, CA)	Jerry Yang David Filo	Computer Science (CS) PhD students
WebCrawler (engine)	20 Apr 94	University of Washington (Seattle, WA)	Brian Pinkerton	PhD student in CS
Lycos (engine)	Jul 94	Carnegie Mellon University (Pittsburgh, PA)	Dr Michael Mauldin and Bob Leavitt	Postdoctoral research fellow in CS
Infoseek (engine)	13 Feb 95	n/a (Sunnyvale, CA)	Steve Kirsch	Serial technology entrepreneur – founded Frame Technology and Mouse Systems. BA and MS from MIT.
OpenText (engine)	Apr 95	n/a (Waterloo, Ontario, Canada)	(uncredited, possbly OpenText VP of Information Retrieval Larry Fitzpatrick)	Early provider of search interfaces to products such as Oxford English Dictionary
Magellan (directory)	Aug 95	n/a (Sausalito, CA)	Isabel & Christine Maxwell	Daughters of publishing magnate Richard Maxwell, originally published a print guide to the Web

[29] Not included in this chart are Archie, a pre-Web search engine for FTP sites developed by McGill University student Alan Emtage in 1990 and Veronica, a similar engine for Gopher sites, developed at the University of Nevada in 1993. Also excluded are the first two Web search engines, the WWW Wanderer, the first spider to crawl the web, developed by Matthew Gray, a researcher at MIT, in 1993, and Aliweb, developed in 1993 by Martijn Koster while he worked for Nexor in Nottingham, England. Neither of these technologies were commercialised.

Excite (engine)	*29 Sep 95*	Stanford University (Palo Alto, CA)	Graham Spence Joe Krausz Ben Lutch Ryan McIntyre Martin Reinfreid Mark Van Haren	Recent CS graduates (apart from Krausz who graduated in political science)
AltaVista (engine)	15 Dec 95	Digital Equipment PARC (Palo Alto, CA)	Dr Louis Monier	Research fellow
Inktomi (engine)	20 May 96	University of California at Berkeley (Berkeley, CA)	Dr Eric Brewer Paul Gaulthier	Assistant professor of CS and graduate student
LookSmart (directory)	28 Oct 96	Reader's Digest (Melbourne, Victoria, Australia)	(uncredited)	(uncredited – presumably the publishing team acting through ordinary channels?)

*Dates refer to when the search engine became publicly accessible.
Source: Data derived from original press releases and news reports, compiled by author.

Figure 18: Elizabeth van Couvering's Table 3, a comprehensive timeline overview of the above-mentioned search engines (2010:96).

results (Beigi et al 1998). Yet another 'metasearch engine', *Northern Light,* 'clustered' results from both private and public information resources and custom search engines, which was considered an innovation at the time. In 1998 *AltaVista* began incorporating 'editorially rated sites', harking back to the earlier explorative days of *Magellan*, only with a new interface. As with the pre-history of the search engine described in Chapter 1, *Ask Jeeves* was a search engine that referenced the name of a personalised butler. '*Ask Jeeves* attempted to make the query process more user-friendly and intuitive, encouraging people to ask fully formed questions rather than use Boolean search queries' (Halavais 2009:23-24).[46] It is now Ask.com.

It is important to note that most of these early search engines came out of research in academic environments or non-commercial settings. 'Search engine technology primarily developed from the academic discipline of information retrieval, which itself is something of a hybrid between library science (now often called information science) and computer science' (Van Couvering 2010:95). The late 1990s was, however, a time of great diversity, with dozens of search engines competing for market share, using advertising as a business model with venture capital.

> This first period of search engine history, then, is characterised by technological innovation within research centres followed by commercialisation using advertising and licensing as business models and capitalisation through venture capital and the stock market. The market was competitive, consisting of multiple companies with different technologies (ibid:100).

The 'middle period' begins with the boom of the dot-com era in late 1997 and continues until 2001. It is 'characterised by the change in focus from search engines to "portals" and the involvement of traditional media and telecommunications giants in the sector' (ibid). From 1997-1998 the navigational aspect was still present along with a directory service, yet this terminology of 'portal' became used more frequently. One example of the portal conglomeration and business models that were being incorporated into the act of searching was *iWon*, with every search query an entry to a lottery rewarding its audience and paying out to its users. Aggregators such as *Dogpile* and *MetaSearch* queried other search engines for their results and *Direct Hit* ranked popularity.

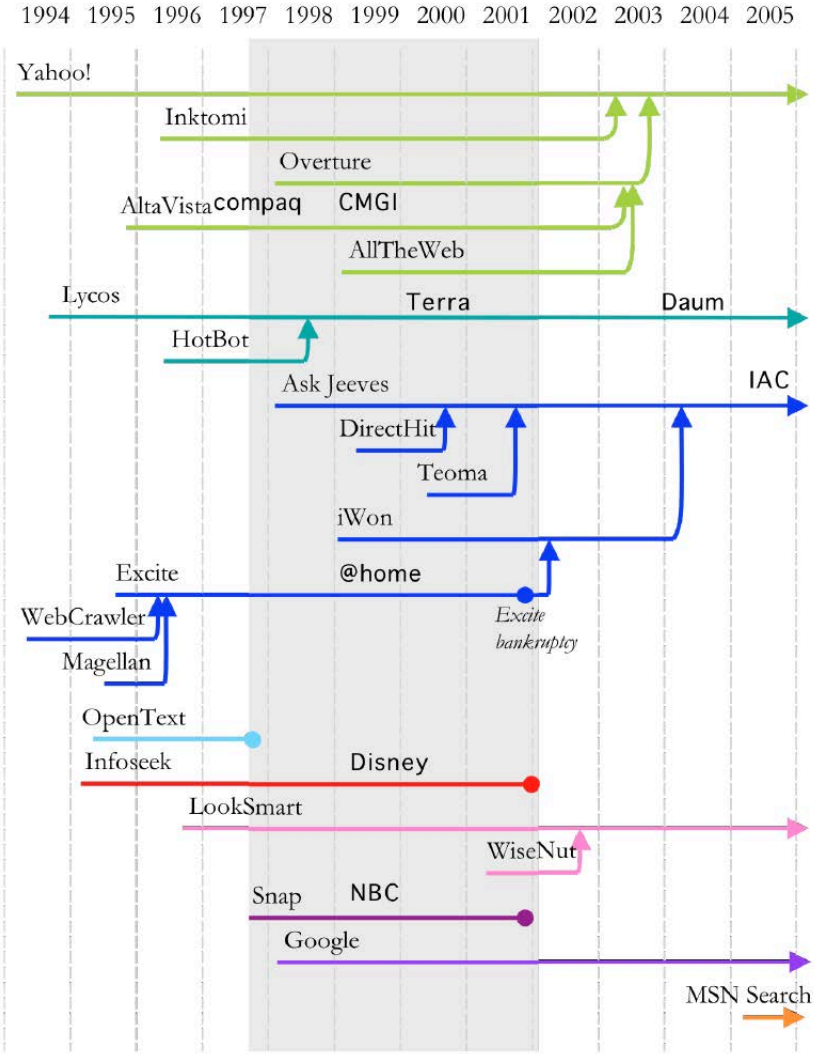
> Eventually many of these directory-based portals became major players, particularly Yahoo, who experimented with a number of search engine partnerships, beginning with combining *Inktomi*'s search technology with their existing directory in 1998, and subsequently acquired some of the largest general-purpose search engines, including *AlltheWeb.com*, *AltaVista*, and *HotBot* (Halavais 2009:24-25).

During the late 1990s there was a large amount of money invested in venture capital, as the search engine market was divided up into shares. Various ways to produce revenue were being tested out in these models; acquisitions occurred, and the field became a bit smaller (ibid:27). Van Couvering describes this period in the history of search engines as two-fold in that there was competition yet also mergers: the concept of the 'walled-garden' (2010:100) where

[46] Boolean search allows users to combine their search queries with terms such as AND, NOT, OR and NEAR to refine their queries and obtain more relevant results. Known as Boolean operators they can limit, define or widen search queries, though nowadays most search engines have these Boolean parameters as default. Operators have corresponding symbols, where AND is equal to "+" and NOT is equal to "-" and OR is default, meaning whatever you type in generates returns.NEAR is equal to putting your search query in quotes, in a specific order.

companies thought they could control one part of the internet, even though they later offered 'limited, curated sets of web- or Internet-based resources while preventing access to the rest' (Plantin et al. 2017: 301-302). The other was the integration of various forms of content and mergers that reflected attempts by media conglomerates to co-opt the arena while neglecting search (Van Couvering 2010:100).[47] (Figure 19) This is made apparent by the merger of the search engine with other services— content from advertisers that included a range of topics e.g. shopping, travel, email, music and finance (ibid:102). Venture capital still played a role, and 'despite the diminution of the actual search engine from the core of the business to lossleading commodity, there continued to be new technical innovations in search' (ibid). As acquisitions and mergers proliferated in the late 1990s the field diminished yet there was still investment from venture capital before the dot.com bubble crash of 2000-2001.

Figure 6: Search engine mergers and acquisitions in the three periods of search history.



Source: Data from company websites and press reports, compiled by author.

Figure 19: Elizabeth van Couvering's Figure 6 illustrating search engine mergers and acquisitions (2010:96).

[47] Yahoo! is perhaps a partial exception because it still survives today.

By the end of the 1990s the model was to have advertising finance search and, as more people came online, the content on the web was changing. The usage of search engines rivalled that of email with people querying 'current events, health concerns, products, government services, natural disasters, their new neighbours, prospective employees or dates, and a myriad of other topics ranging from the mundane to the utmost serious' (Hargattai 2007:769). For a search engine, the question was how to discern those users looking for porn, which was also a dominant activity, from others searching for information or just 'surfing' the web.[48] In 1998 an early theorist of the web, Phil Agre, contextualised the problem with search at this moment with a comparison of connections between two media: the web and the telephone.

> Assuming that every page on the web had eight hyperlinks leaving it, and that the targets of these links were picked at random from all the possible sites on the web, the structure would be entirely unnavigable. Unlike the telephone, the web is very malleable, able to take on the characteristics not just of a point-to-point network or of a broadcast network, but a multitude of shapes between the two (Agre 1998; cited by Halavais 2009:59).

It is not only the web's malleability but also its navigability, which is facilitated by hyperlinking. This interconnection between texts would come to embody a type of politics, where authority determines the relevancy of information through the merger of search and research, as I show in the next chapter with Google's PageRank algorithm.

[48] However, more people were searching for pornography and reciprocally pornography sites were also in search of viewers and wanted to be found by their audience. 'Often credited as the first money-making business online, adult websites became a big part of [the] online economy' (Gilmore 2016).

Chapter 3:

A (Media) Archaeology of Citation

1.0 What is an algorithm?

The past two chapters explained *how searched worked in the past*, demonstrating how the address offices and registers, and later on private and public endeavours, organised the world's information. In the late 1980s academics developed search technologies such as the hyperlink and by the mid 1990s commercial interests had already been woven into the fabric of the web. In 1998 two academics introduced a search engine that could not only crawl and index quickly but also deliver 'relevant' results with their innovative PageRank algorithm. In this chapter, in order to answer my research question *How does Google Search work?,* I begin by asking, what is an algorithm? The name refers to the 9th century Persian mathematician Abū Ja'far Muhammad ibn Mūsa, in Arabic called *al-Ḵwārizmī*, meaning 'the man of Ḵwārizm', who was a translator of arithmetic and algebra.[49]

> It is probably more accurate to say that it developed from or with the word algorism, a formal term for the Hindu-Arabic decimal number system, which was sometimes spelled algorithm, and which itself is said to derive from a French bastardization of a Latin bastardization of al-Khwārizmī's name, Algoritmi (Gillespie 2016:18-19).

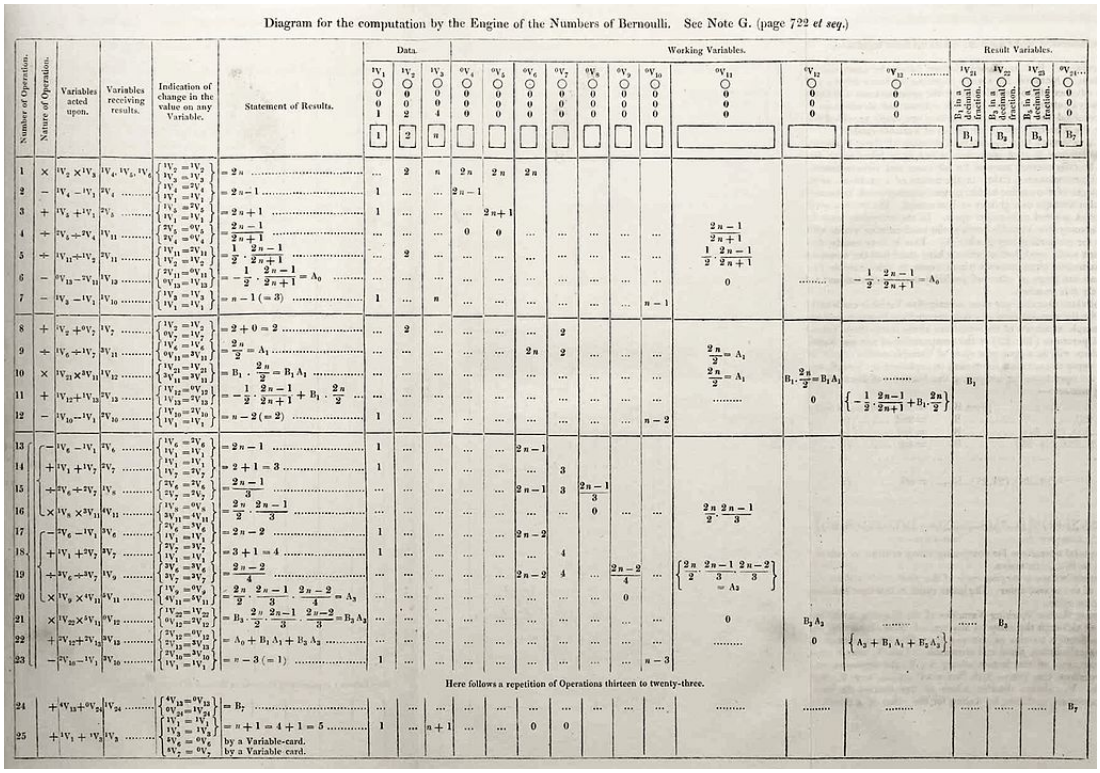


Diagram for the computation by the Engine of the Numbers of Bernoulli. See Note G. (page 722 *et seq.*)

Figure 20: Diagram of 'algorithm G' for the computation of Bernoulli numbers, from *Sketch of The Analytical Engine Invented by Charles Babbage* by Luigi Menabrea, with notes by Ada Lovelace (1843).

[49] 'Like his crater and the zero concept he [al-Khwārizmī] championed, the term algorithm will turn out to be important in part because it is vacant, a cypher, a ghostly placeholder upon which computational systems now stand' (Gillespie 2016:18-19).

In the 19th century Ada Lovelace realised that there was more to just 'calculating machines'. Around 1843 she developed a method for calculating a sequence based on Bernouilli numbers in relation to Charles Babbage's 'Analytical Engine'; according to some, this entailed the first executable algorithm written on paper and constituted 'the "first computer program, and the notes the first expression of computer theory' (Krysa 2013:353). (Figure 20) Although Babbage's machine didn't exist yet, in contemporary terms Lovelace's paper algorithm was 'the software required to operate the hardware of Babbage's machine' (ibid).[50]

As Gillespie points out, we should be aware of the algorithm's diverse meanings 'as well as the discursive work the term performs' (2016:18). Software engineers use the term as a 'conceptual sequence of steps, which should be expressible in any computer language, or in human or logical language' (ibid:21). Although rule-based and able to solve mathematical problems, an algorithm is often considered by the public as 'a set of instructions for carrying out procedures step-by-step, and range from quite simple to very complex' (Tufeksci 2015:206). While for organisational scholars, 'algorithms can be understood as series of generalized procedures for turning disorganized data inputs into manageable outputs through a series of logical rules' (Flyverbom and Madsen 2015). According to a data scientist, an algorithm is a recipe where the ingredients are already curated, which are 'comprised of essentially the data you train your algorithm on' and there is an agenda imposed along with it being optimised for success when executed (O'Neil 2016). Gillespie concurs that an algorithm is a 'recipe composed in programmable steps' yet emphasises that the generation of a model is necessary for 'the formalization of a problem and its goal, articulated in computational terms' (2016:19).

Technical specialists further articulate the algorithm as 'merely the procedure for addressing the task as operationalized: steps for aggregating those assigned values efficiently, or making the matches rapidly, or identifying the strongest relationships according to some operationalized notion of "strong"' (ibid:20). An 'array of parameters and thresholds' are 'tuned' by the designers that reflect 'a tiny assessment or distinction':

> In search, this might mean the weight given to a word based on where it appears in a webpage, or assigned when two words appear in proximity, or given to words that are categorically equivalent to the query term. These thresholds can be dialed up or down in the algorithm's calculation of which webpage has a score high enough to warrant ranking it among the results returned to the user (ibid:21).

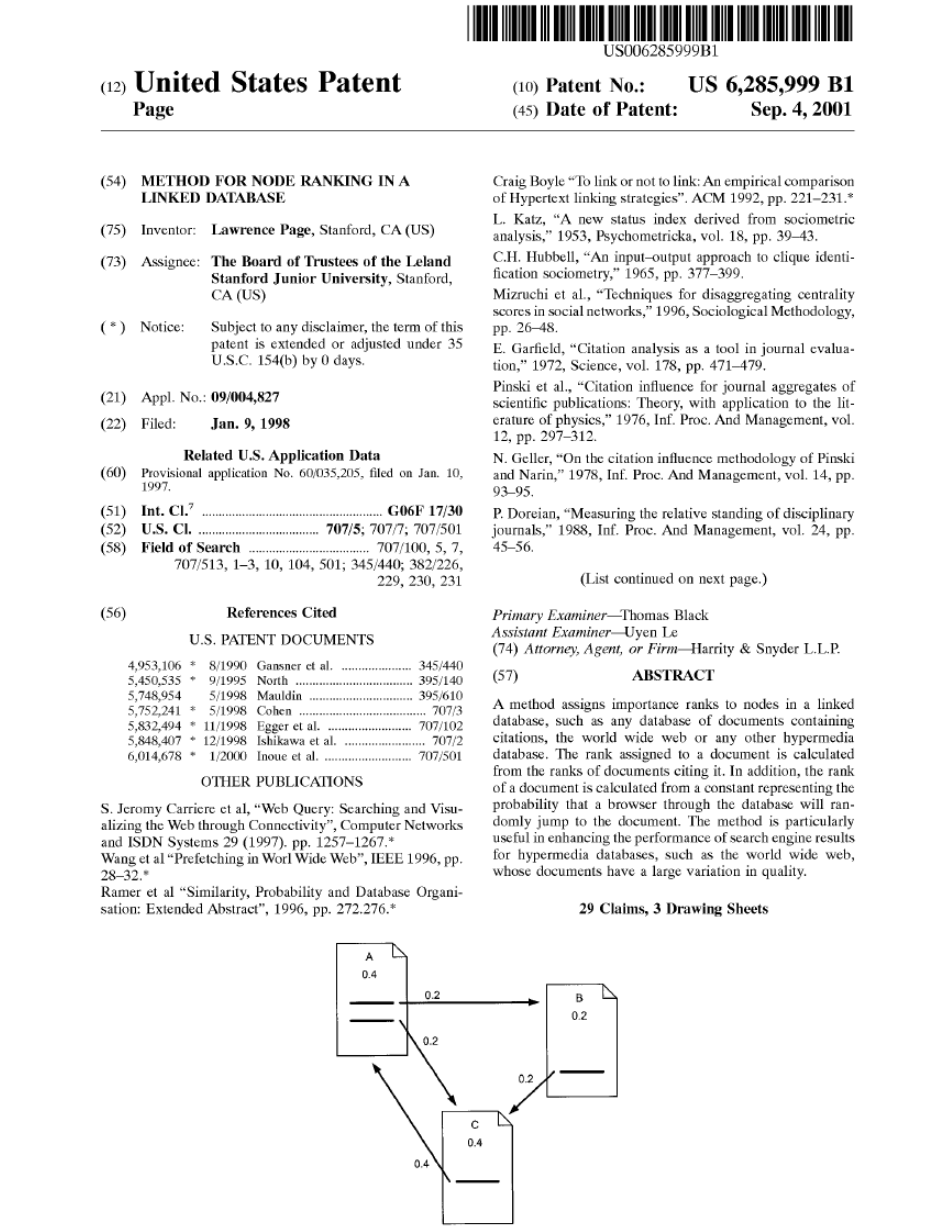
For Google's PageRank it is the target of providing users with 'the most relevant search results for their query', 'efficiently calculating the combined values of preweighted objects in the index database', which is quantified by user satisfaction 'in terms of percent clicks on the top results' (ibid:19-20).

2.0 What is in PageRank? (Bibliometry)

In his article *What is in Page Rank?* Bernhard Rieder deconstructs PageRank not by asking what an algorithm is but rather by enquiring, 'what is *in* an algorithm?' (Rieder 2012:1 emphasis mine). Articulated otherwise, what are the ingredients and what is the procedure for producing a result from the recipe? Rieder's contribution to the field of software studies focuses on the 'conceptual commitments' that PageRank embodies framed through computation, not just as

abstract Turing machines,[51] but rather the algorithm itself as an object-in-the-world:

> [R]eal software, lodged in tight networks of other software, all written for a purpose; knowledge, ideas, skills, tools, hodology, habits, and values that permeate practices embedded in layers of social organization, cultural configurations, economic rationales, and political struggles (ibid).



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 Worl Wide Web", IEEE 1996, pp. 28–32.*
Ramer et al "Similarity, Probability and Database Organisation: 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, Psychometricka, 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**

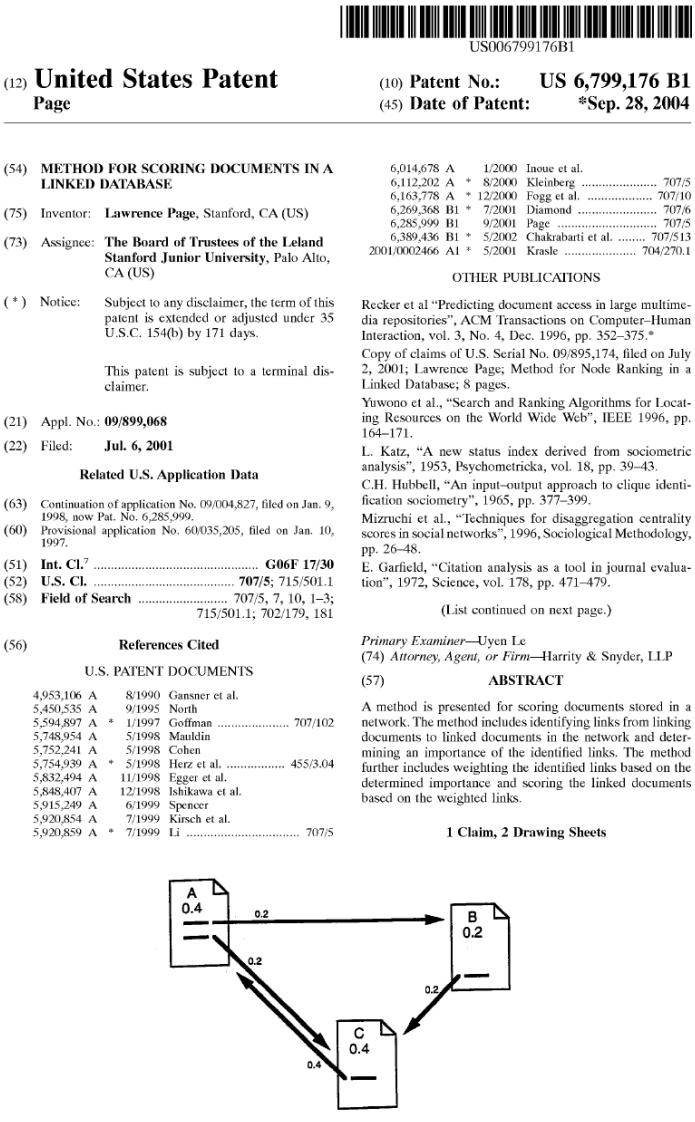
A method assigns importance ranks to nodes in a linked database, such as any database of documents containing citations, the world wide web or any other hypermedia database. The rank assigned to a document is calculated from the ranks of documents citing it. In addition, the rank of a document is calculated 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 21: US patent 'Method for Node Ranking in a Linked Database' by Larry Page (1998)

[51] In a 1936 text, mathematician Alan Turing's hypothetical machine could simulate any computer algorithm, no matter how complicated it is, by transforming a string of 1's and 0's into a different string of 1's and 0's. This is the essence of computation and the blueprint for digital computers to solve problems. When a computer can do what a Turing machine can do, it is called 'Turing complete'. Quantum computers challenge this notion.

Specifically, in regard to habits and hodology, the paths that the PageRank code takes in its development make use of references. Rieder begins with two patents that are 'more explicit in their citation practice than the academic publications' (ibid). On January 8, 1998 Larry Page filed a patent 'Method for Node Ranking in a Linked Database' (US6285999) which became PageRank—pun intended. [52] (Figure 21) Also shown by Larry Page's second patent 'Method for Scoring Documents in a Linked Database' (U.S. Patent 6799176), Page not only references his previous patent but the article 'Citation analysis as a tool in journal evaluation' by one of the founders of bibliometrics, Eugene Garfield, mentioned in Chapter 2. (Figure 22) The patents influence ranking but also reflect Intellectual Property Rights (IPR), one of the ways that enabled Google to transform itself from an academic start-up to a billion-dollar advertising agency. I will address both in more detail in Chapters 5 and 10.



US006799176B1

(12) **United States Patent**
Page

(10) **Patent No.: US 6,799,176 B1**
(45) **Date of Patent: *Sep. 28, 2004**

(54) **METHOD FOR SCORING DOCUMENTS IN A LINKED DATABASE**

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

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

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 171 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **09/899,068**

(22) Filed: **Jul. 6, 2001**

Related U.S. Application Data

(63) Continuation of application No. 09/004,827, filed on Jan. 9, 1998, now Pat. No. 6,285,999.
(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**; 715/501.1
(58) **Field of Search** 707/5, 7, 10, 1–3; 715/501.1; 702/179, 181

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,953,106	A		8/1990	Gansner et al.
5,450,535	A		9/1995	North
5,594,897	A	*	1/1997	Goffman 707/102
5,748,954	A		5/1998	Mauldin
5,752,241	A		5/1998	Cohen
5,754,939	A	*	5/1998	Herz et al. 455/3.04
5,832,494	A		11/1998	Egger et al.
5,848,407	A		12/1998	Ishikawa et al.
5,915,249	A		6/1999	Spencer
5,920,854	A		7/1999	Kirsch et al.
5,920,859	A	*	7/1999	Li 707/5
6,014,678	A		1/2000	Inoue et al.
6,112,202	A	*	8/2000	Kleinberg 707/5
6,163,778	A	*	12/2000	Fogg et al. 707/10
6,269,368	B1	*	7/2001	Diamond 707/6
6,285,999	B1		9/2001	Page 707/5
6,389,436	B1	*	5/2002	Chakrabarti et al. 707/513
2001/0002466	A1	*	5/2001	Krasle 704/270.1

OTHER PUBLICATIONS

Recker et al "Predicting document access in large multimedia repositories", ACM Transactions on Computer–Human Interaction, vol. 3, No. 4, Dec. 1996, pp. 352–375.*

Copy of claims of U.S. Serial No. 09/895,174, filed on July 2, 2001; Lawrence Page; Method for Node Ranking in a Linked Database; 8 pages.

Yuwono et al., "Search and Ranking Algorithms for Locating Resources on the World Wide Web", IEEE 1996, pp. 164–171.

L. Katz, "A new status index derived from sociometric analysis", 1953, Psychometricka, 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 disaggregation 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.

(List continued on next page.)

Primary Examiner—Uyen Le
(74) *Attorney, Agent, or Firm*—Harrity & Snyder, LLP

(57) **ABSTRACT**

A method is presented for scoring documents stored in a network. The method includes identifying links from linking documents to linked documents in the network and determining an importance of the identified links. The method further includes weighting the identified links based on the determined importance and scoring the linked documents based on the weighted links.

1 Claim, 2 Drawing Sheets

Figure 22: US patent 'Method for Scoring Documents in a Linked Database' by Larry Page (2001)

[52] Page received the patent on September 4, 2001, when search engines became crucial in altering how information was prioritised, coinciding with the events of 9/11, as pointed out by Pynchon in *Bleeding Edge*.

Citing previous research, the application of the very same circularity of Garfield's SCI hierarchies are then instantiated with the iterative measurement process of PageRank itself. In his book *Search Engine Society* Alexander Halavais observed that there are 'certain affinities between citations and hyperlinks, and both undergird the distributed conversation that is at the heart of scholarship' (2009:102).[53] Ironically, it is the same general recursitivity that Brin and Page's article (1998) describes, which I will come to shortly, as it explains their novel search engine design. Their text has been studied 'extensively by mathematicians and computer scientists but is also regularly commented on by scholars in the humanities' (Rieder 2012). As stated in the Introduction, the title of this PhD, 'Re:search' is a conjunction of 'research' and 'search', reflecting the way in which these terms are intertwined, both through their nomenclature along with how PageRank has become the primary, bibliometrically based, means for hierarchising information on the web. Harking back to Bush's 'associative linkages' and Garfield's 'association-of-ideas index' from Chapter 2, the Scientific Citation Index was the basis for an evaluative metrics that could measure the 'impact factor', as duly noted by various authors (Halavais 2009; Gillespie 2014; Rosenberg 2016; Chun 2016).[54]

> From a mathematical standpoint, the impact factor, calculated for scientific journals, is an extremely basic measure: it takes the number of citations the last two volumes of a journal received the following year and divides it by the number of articles published in these two volumes (Rieder 2012:5).

In this way the entanglement of 'research' in academia with 'search' engines becomes apparent, along with how authority is measured in both fields. 'In a citation network, the more number of citations a paper has, the more important it is considered to be' (Chun 2016:103).[55] This is how academia measures the value of a researcher's published articles yet the impact factor does not 'take into account the importance of the citing journals' and therefore, in 1976, Gabriel Pinski and Francis Narin developed another method to rank journals (Franceschet 2010:5.2). They critiqued Garfield's SCI, noticing that prestigious journals should be considered more worthy and should not be given the same 'weight' as those on the margins and they designed an 'input-output' solution with the weight of incoming citations divided by the number of outgoing citations (Rieder 2012:6).[56] Their proposition was a recursive index that measured the importance of the text yet also took into consideration eigenvector calculations. In other words, their thesis stated that 'a journal is influential if it is cited by other influential journals' and, according to Franceshet, this is the 'same circular thesis of the PageRank method' (ibid).[57]

---

[53] 'Some universities base academic advancement on the citations found in the Web of Science, an indication of both its status in the academic community and the relationship of search technology to the institutionalized power of the academy' (Halavais 2009:99).

[54] Garfield's Scientific Citation Index (SCI) was an index of citations that was 'manually extracted' and contained over 613 volumes and stored 1,4 million citations on magnetic tape (Rieder 2012:5). As Rieder notes, 'this index is perhaps the first "big data" file available in the social sciences, and over the following years a significant number of researchers participate in analyzing it with various computational methods' (2012:5).

[55] 'The *h*-index is an author-level metric that attempts to measure both the productivity and citation impact of the publications of a scientist or scholar. The index is based on the set of the scientist's most cited papers and the number of citations that they have received in other publications' (https://en.wikipedia.org/wiki/H-index).

[56] Rieder notes that '[i]f we were to introduce a "damping factor" into this model, such as Katz's connection weight reduction scheme, in order to attenuate the "free flow of citations in the referencing market place", we would essentially end up with PageRank' (2012:6).

[57] Eigenfactor (http://eigenfactor.org/), a PageRank inspired bibliometric indicator, is among the most interesting recent proposals to collectively evaluate the status of academic journals (Franceschet 2010:6.0).

This ‘normalising factor’ is crucial to understanding how sociological research was built upon in regard to designing iterative ranking algorithms.

3.0 Eigenvector centrality (Sociometry)

Rieder shifts focus to the ‘margins’ with an archaeological probe of statistics, beginning with sociometry, which was founded in 1937. Using a questionnaire, or survey, social scientists asked people to choose their favourite individuals, the results were supposed to reveal the underlying psychological structure of society. Society could then be mapped and visualised though a ‘sociogram’ or a ‘point and line’ diagram, what would nowadays be called a ‘network visualization’ (Rieder 2012:3). Shifts in representation, such as the use of a square matrix or reordering by row and column were methods used to produce views of this sociometric data, resembling the standardised techniques of statistics. However, this is just a reordering of order, transforming numbers into other numbers tautologically—it is looking at the same data, only differently (ibid).

Groups of people that were related to each other were measured by ‘indegree, closeness, betweenness, as well as eigenvector centrality’, which are still used in contemporary ‘network analysis’ (Franceschet 2010:5.3). Specifically, it was the ‘eigenvector centrality’ that ‘applied PageRank to a social network’, where ‘a person is prestigious if he is endorsed by prestigious people’ (ibid). The antecedent of this circular argument performed by PageRank was probably first articulated by John R. Seeley in 1949, applied to a context with children that he defined thus: the popularity of a child as a function of the popularity of those children who choose the child, and the popularity of the choosers as a function of the popularity of those who chose them and so [on] in an “indefinitely repeated reflection”’ (ibid). Using mathematics to formalise social networks, the psychosociologist Alex Bavelas introduced ‘distance’ and ‘closeness centrality’ to the matrix structure in the 1940s, yet it wasn’t until 1953 that the ‘first application of graph theory to network analysis appeared’ (Rieder 2012:3). The use of graph theory, simply put—point and line methods— in regard to contemporary network analysis, is the ‘theory of the social’, or the mapping of the relationship between individuals that has been brought to bear on all types of intellectual endeavour.

Returning to PageRank, the connections between the nodes (and authors) and how they are quantified then becomes crucial; the influences between these individuals reflect their ‘social power’ and most importantly, how it is calculated (ibid:5). Both of Page’s patents mention Katz’s 1953 article ‘A new status index derived from sociometric analysis’ in *Psychometricka*. Instead of counting votes, which is a common way to measure popularity, Leo Katz proposed a new computational method for measuring social status from sociometric data (ibid). The saying ‘its not what you know but who you know’ comes to pass in determining ranking criteria. This is conducted by measuring the hodology, or ‘paths’ between all of the nodes, ‘then attributing a lower weight to longer paths through ‘damping’, and finally calculating a metric for every node based on the combined weight of their connections to all other nodes’ (Katz 1953 cited by ibid). This type of measurement is known as ‘Katz centrality’ and was improved upon by Philip Bonacich’s contribution in 1972 that not all paths were measured, only the shortest ones, called eigenvectors, which reflected the locality of the network (ibid).[58]

[58] Rieder is swift to point out that ‘references to Bonacich’s work are conspicuously absent from the PageRank patents’ (2012:5).

The economic calculations behind sociometry (and eventually leading up to PageRank) were first indicated by Granovetter in *The Strength of Weak Ties* (1973), where he ponders why '[s]ociometry, the precursor of network analysis, has always been curiously peripheral—invisible, really—in sociological theory' (Granovetter cited by ibid:6).[59] (Figure 23)

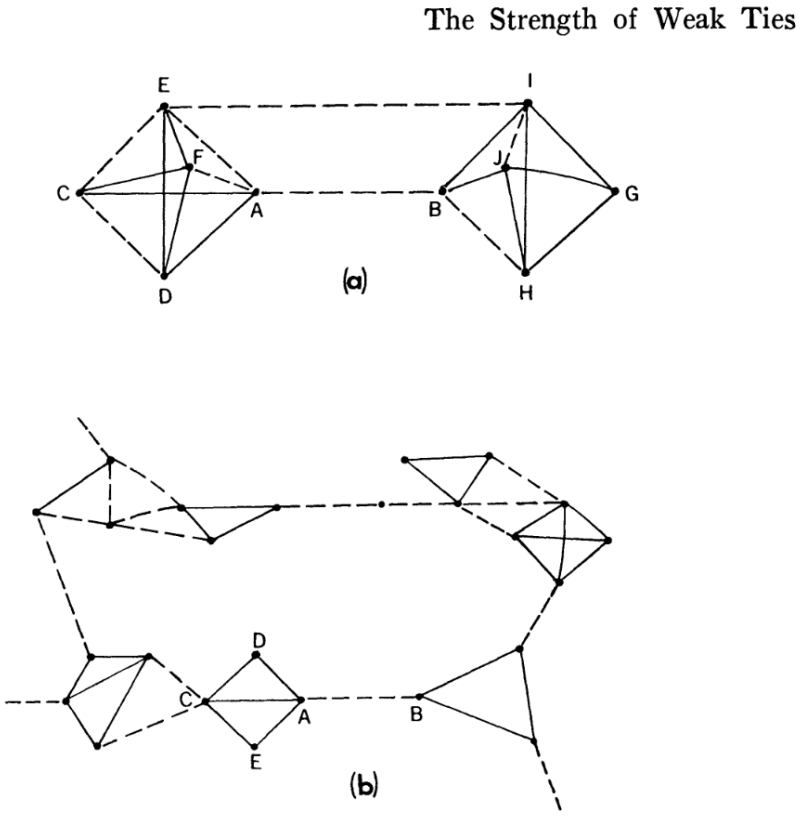


FIG. 2.—Local bridges. *a*, Degree 3; *b*, Degree 13. ——— = strong tie; — — — = weak tie.

Figure 23: Granovetter's diagram (1973). The dotted lines show the weak ties.

Granovetter created one of the 'most influential social maps' that makes the 'relationship between personal experience and social structure' comprehensible (Chun 2016:42). In this diagram Chun points out that Granovetter 'countered the assumption' that constant static nodes are the most powerful, instead 'disseminating rare information' […] 'moves us from the mass, measured in terms of sheer force, to the N(YOU), gauged in terms of unique of lines of influence' (ibid). In other words, networks with weak ties have reduced path lengths that enable a quicker diffusion of information.

However, in Rieder's own citation archaeology he discovers the leap from social theory to 'social exchange theory'—articulated by Richard Emerson in 1976 as an approach that 'might be described, for simplicity, as the economic analysis of non-economic social situations' (Emerson cited by Rieder 2012:6). Furthermore, he emphasises strands in his archaeology that provide a narrative which upholds what he deems the 'innocence of the link':

[59] 'Granovetter's paper *The Strength of Weak Ties* is highly influential research, with about 37,000 citations according to Google Scholar (by March 2016)'. https://en.wikipedia.org/wiki/Mark_Granovetter

> whether it is spontaneous attraction, rational choice or simply an 'inspirational' account of scientific citation, the application of the metrics to actual ranking, with concrete and tangible consequences, can only be justified if the link is kept reasonably pure. In this vision, the main 'enemy' is therefore the deceitful linker, whether they come in form of scientific citation cartels or their contemporary cousins, link farms. It is not surprising that a central argument against citation analysis as a means for research evaluation builds on a critique of actual citation practices (ibid:7).

Critique aside, bibliometrics, sociometrics, graph theory, citation metrics and social exchange theory all comprise the archaeology of the PageRank algorithm that Rieder explains through his own recursive citation analysis.

> 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 (ibid).

Moreover, he argues that these factors cannot be overlooked, as well as 'how an understanding of metrics as descriptions can help us in dealing, analytically, with their becoming operative prescriptions' (ibid). The prescriptive characteristics of 'real technology' were already mapped out in 1989 by Ursula Franklin, which I will address in Chapter 4, yet it is the ingredients that go into the prescription, or recipe of the algorithm that is of importance. Herewith, the very design of these directional devices (algorithms) plays a crucial role, 'becom[ing] both invisible and inevitable, in the sense that margins of discretion are defined by the system itself' (ibid). The recipe of the algorithm then sounds easy enough to reproduce but it's a well-kept secret that generates immense revenue for one of the world's highest earning corporations. (Figure 24)

Figure 24: image from Danny Sulllivan's Twitter feed [60]

[60] In September 1998, Brin and Page rented a garage in Menlo Park, Susan Wojcicki's, who became the 16th employee at Google and developed Google's Image Search. She was responsible for two important Google

4.0 Hypertext: The Anatomy of a Large-Scale Hypertextual Web Search Engine

Although PageRank's recipe is shrouded by IP rights, in 1998 Brin and Page published *The Anatomy of a Large-scale Hypertextual Search Engine,* the only text that explains (in its various versions) their novel search engine on their own terms and lends insight into their own citation archaeology.

> The description of Google's PageRank system, the earliest component of its complex search algorithm, was published first as a technical paper (already a crafted rendition of its mathematical workings), but was subsequently mythologized— as the defining feature of the tool, as the central element that made Google stand out above its then competitors, and as a fundamentally democratic computational logic—even as the algorithm was being redesigned to take into account hundreds of other criteria (Gillespie 2014:180).

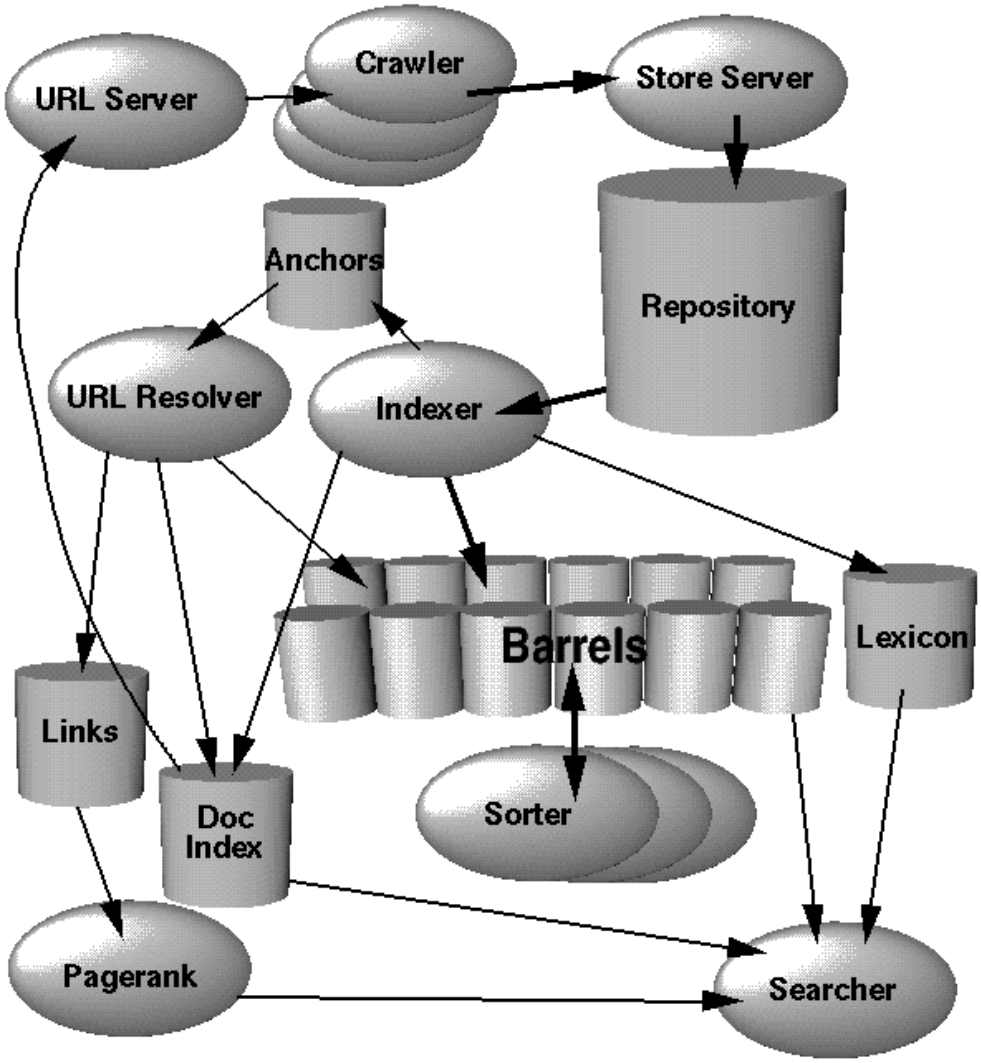


Figure 25: 'System Anatomy'. Brin and Page's Figure 1: High Level Google Architecture.

acquisitions, DoubleClick in 2007 and You Tube in 2006, where she is presently CEO since 2014. Danny Sullivan is a co-founder of Search Engine Land (SEO) but since 2017 he is Google's public 'Search Liason'.

In their text, Brin and Page state that they choose the name Google for their company 'because it is a common [mis]spelling of googol, or *10 100* and fits well with our goal of building very large-scale search engines' (1998:107), which I will return to in the Epilogue.

Brin and Page forecast that a 'comprehensive index of the web will contain over a billion documents', though they argue against 'completeness', because a complete index does not always determine the quality of the search results and besides porn, there's a lot of 'junk' out there (ibid:108). Keeping this in mind, they explicitly state that 'Google's data structures are optimized so that a large document collection can be crawled, indexed, and searched with little cost' and therefore, as the size of the web grows, Google would 'scale cost effectively' (ibid:112). (Figure 25) This would entail optimising data structures that enable quick indexing and the ability to process 'hundreds of gigabytes of data efficiently', along with processing the ever-increasing amount of user search queries, which, at that time was 'hundreds of thousands per second' or 'hundreds of millions of queries per day' (ibid:107-108).

Besides the scale of the index and the speed of processing documents as the amount of queries increased, all of the data had to be stored somewhere. In their text they describe how Google is optimised for 'fast and efficient access' as well as taking into consideration the concept of Moore's law, where 'a doubling every 18 months in processor power,' makes it economically feasible (ibid:108).[61]

Brin and Page also explain that only one of four search engines 'finds itself'; in other words, the search returns contain the name on the first page. Therefore, their design needed to filter relevant results, e.g. to only include 'the very best documents since there may be tens of thousands [nowadays billions] of slightly relevant documents' (ibid:109). Seen through the lens of academia and the Scientific Citation Index (SCI), they cite other researchers who expressed optimism that using hypertextual information would improve search results '[Marchiori 97] [Spertus 97] [Weiss 96] [Kleinberg 98]' (ibid:108). Particularly their reference to Marchiori's *The Quest for Correct Information on the Web:Hyper Search Engines* (1997) is of importance here because of his contribution of 'hyper information', later on referred to as 'hypersearch', which recognised not only the web object, but the web structure which it is part of.

Marchiori's 'hyper information' measured 'the potential information content of a Web object with respect to the Web space', scoring not only the content of text on webpages but the more dynamic behaviour of hypertext—pages it is linked to—thereby 'augment[ing] the informative content with the information contained in the pointed Web objects' (1997). His proposed method would measure the 'depth' (the minimum number of links that have to be activated, or "clicked") of these hyperlinks and 'how far the information is from the user (how many clicks s/he has to perform)' (ibid). In other words, '[b]y fixing a certain depth, we thus select a suitable finite "local neighbourhood" of a Web object in the World Wide Web' (ibid). 'Hyper information' would work as a 'post-processor' of the main search engines available at that time and analyse the informative content of a web object, i.e. all 'web objects that are reachable from it via hyperlinks', whilst navigating the World Wide Web (ibid).[62]

[61] Moore's Law was defined in 1965 as a doubling every 18 months in processor power.
[62] 'The search engines for which a post-processor was developed were: Excite, HotBot, Lycos, WebCrawler, and OpenText. This list includes all of the major search engines, but for AltaVista and InfoSeek, which unfortunately do not give the user access to the scores, and thus cannot be remotely post-processed' (Marchiori 1997).

5.0 Relevancy

5.1 IR system

As Rieder critically points out, the 'Google search engine is of course both a link analysis and an IR System' and I begin by addressing the latter (2012:8). The term Information Retrieval (IR) is attributed to Calvin Moers in 1950, where he defined IR as the 'human cognitive processes' in epistemological quests that drew on 'theories of information categorization', as well as creating automated catalogues and 'retrieval in the field of computer science and artificial intelligence' (Van Couvering 2009:95). IR spans different fields, however its combination with citation analysis developed for finding scientific literature began in the 1990s with the rise of search engines. This convergence between 'Information Retrieval' and 'searching the web' occurred at the moment when the web was expanding exponentially and creating millions of documents, which needed to retrieved and ordered. PageRank was designed to bring the searcher to another destination as a means to solve the 'lost in hyperspace' problem, by determining what is 'relevant' to the user.

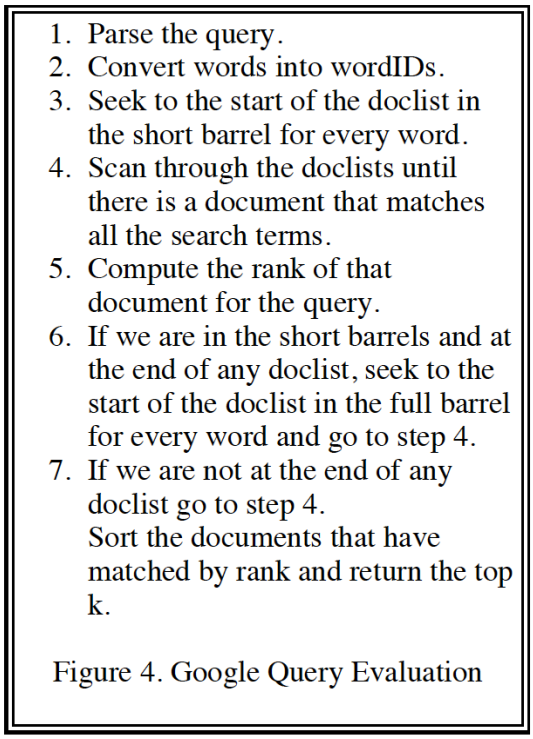

1. Parse the query.
2. Convert words into wordIDs.
3. Seek to the start of the doclist in the short barrel for every word.
4. Scan through the doclists until there is a document that matches all the search terms.
5. Compute the rank of that document for the query.
6. If we are in the short barrels and at the end of any doclist, seek to the start of the doclist in the full barrel for every word and go to step 4.
7. If we are not at the end of any doclist go to step 4.
   Sort the documents that have matched by rank and return the top k.

Figure 4. Google Query Evaluation

Figure 26: Brin and Page's Figure 4 (1998)

Brin and Page's concept of 'relevancy' was based on the discovery that '[p]eople are still only willing to look at the first few tens of results' (1998:108).[63] Brin and Page also point out that the biggest problem facing users is the quality of the returned search results and efficiency, with the solution being the Google query evaluation process (Figure 26). In order to obtain quality search results based on a query, they defined 'relevancy' through hit lists, keeping location information

[63] This revelation is crucial because it takes the position of a user who has to be able to search through large indexes of information and also inspired me to look on the 2, 3rd or even the 50th page of Google Search results.

and the use of 'proximity' for search which helps increase relevance for multiple word queries. PageRank works in conjunction with automated programs called spiders or crawlers and applies an 'inverted index' or 'reverse index' that improves search speed as it indexes words or terms into a database of text elements. The PageRank algorithm also considered the frequency and location of keywords within a web page and how long the web page has existed (ibid). Also 'full, raw HTML of pages is available in a repository', the font size and whether the words are in bold, or are weighted differently is important (ibid:110). Appendix A provides a more detailed explanation of 'how keyword search works'.

The section of Brin and Page's text titled *Intuitive Justification* offers insight into what is behind the workings of PageRank and why specific pages are given a certain 'weight' based on terms. Karen Spärck Jones is accredited with the invention of 'inverse document frequency' in IR systems (1972), where the

> importance of terms is weighted according to the proportion of documents in the corpus in which they occur; the intuition being that terms which occur in many documents are poor index terms. This is the partial basis of all weighting schemes adopted by widely used Internet search engines (Tait 2007).

In regard to IR systems, the information scientist Tefco Saracevic states that 'there is no such things as relevance without a context. Relevance is always contextual' (2013). Rieder concurs, stating that the 'classic IR idea of relevance' is always thought about 'in relation to a specific "informational need" —complemented by the sociometric concepts of status and authority' (2012:8). According to Anders Koen Madsen, PageRank creates a 'market of relevance' that depends on a calculative space similar to the one economic sociologists have detected in a 'market of goods' (2012:12). Just as markets assign prices to goods, the function of PageRank assigns visibility and relevance to information in response to specific queries.

> We take the dot product of the vector of count-weights with the vector of type-weights to compute an IR score for the document. Finally, the IR score is combined with PageRank to give a final rank to the document' (Brin and Page 1998:109).

5.2 Link analysis

To return to the former aspect of Rieder's statement regarding link analysis, Brin and Page's self-referential citation method constituted exploiting links and divulges part of PageRank's secret recipe. 'In particular, link structure [Page 98] and link text provide a lot of information for making relevance judgments and quality filtering' (Brin and Page 1998:108).[64] PageRank applies an 'anchor text'—the human readable text in a URL, which is no longer than 60 characters and provides a clear and accurate description of the web pages themselves but also consumes a large amount of computing power for processing. Other text-based search engines cannot read documents such as databases, images, programmes, etc. that also have anchor texts but 'Google can and does return pages that haven't even been crawled or didn't yet exist' but had 'hyperlinks pointing to it' (ibid:109). The algorithm evaluates the number and quality of links to a page to get a rough estimate of the value of the page relative to all pages on the web.

[64] Larry Page had already published on the idea of how to rank efficiently under his last name, Page.

The more links on a page the more valuable those pages are, value being measured by the amount of times a website is visited on the web.

> When Google developed PageRank, factoring in incoming links to a page as evidence of its value, it built in a different logic: a page with many incoming links, from high-quality sites, is seen as 'ratified' by other users, and is more likely to be relevant to this user as well (Gillespie 2014:178).

The more important or worthwhile websites were likely to receive more links from other websites and this 'linking' is a direct reflection of society and a 'rational attribution of importance' that contributes to the 'universal understanding of authority' (Rieder 2012:8).

> Every link between websites is a kind of vote about worth or "authority" (Kleinberg 1997). Google reads the infrastructure of the web as the interpreter of its content. The metadata used for indexing are not contained in a document but are inferred, kind of like Saussure's network of language. Google's crawlers are interested in key words on websites, but are even more interested in network strength and density, reading for inlinks and outlinks (PageRank rarely points to broken links) (Peters 2015:327).

Besides indexing, IR and linking determining relevance, how the 'corpus was conceptualised' becomes decisive, because PageRank is not a flat corpus (Rieder 2012:8). Instead of being a document repository, PageRank is a 'social system', reflected by documents that have been placed in a 'stratified network/society before the searching even begins' (ibid). It is therefore helpful to elucidate how PageRank could have been designed otherwise.

Rieder subsequently compares PageRank to Jon M. Kleinberg's HITS (Hyperlink-Induced Topic Search) that appeared independently around the same time and proposed another 'ranking' model to determine the 'relevance' of results.

> An advantage of HITS with respect to PageRank is that it provides two scores at the price of one. The user is hence provided with two rankings: the most authoritative pages about the research topic, which can be exploited to investigate in depth a research subject, and the most hubby pages, which correspond to portal pages linking to the research topic from which a broad search can be started (Franceschet 2010:5.1).

Furthermore, Hugill et al. explain the important distinction between the two search methods:

> PageRank assigns a numerical weight to each document, where each link counts as a vote of support in a sense. It is executed at indexing time, so the ranks are stored with each page directly in the index. HITS basic features are the use of so-called hubs and authority pages. It is executed at query time. Pages that have many incoming links are called authorities and pages with many outgoing links are called hubs (2013:247).

Although both analysed the link structure of the Web, the difference was that HITS used two eigenvector metrics, one for authority and the other for 'hubs' and secondly reverses the temporality, the order of how the documents were retrieved or ranked:

> Rather than calculating a universal or a priori landscape of authority, documents matching a query are retrieved first and authority is calculated second —thus based on the link structure in the result set only and not the full corpus (Rieder 2012:8).

In this way one could question the implicit 'bias' in PageRank, having already sorted and compiled the corpus before ranking whereas with HITS, 'authority is dependent on a domain' (ibid). The query is thereby 'freer' to roam through the documents and a page having a high score for one query would not necessarily receive a high score for another. 'Ranking can be done at different stages of the search process. Depending on how the index is formatted and what information can be' (Hugill et al. 2013:246). However, the reason why HITS was not a success was because of the 'higher susceptibility of the method to spamming' (Franceschet 2010:5.1) and the fact that it demanded much more computational power to carry out search requests.

6.0 PageRank as a visibility engine

Brin and Page refer to the SCI in their text by stating that '[i]ntuitively, pages that are well cited from many places around the web are worth looking at' (Brin and Page 1998:109). Furthermore, they already hint at PageRank 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' (ibid). They also explain that '[a]cademic citation literature has been applied to the web, largely by counting citations or backlinks to a given page' and that PageRank is 'not counting links from all pages equally', rather they 'normalise' the amount of links (ibid). This 'normalising' is also drawn from SCI where the number of citations is divided by the number of pages in the entire book, as explained in their text below, represented by a screenshot. (Figure 27)

We assume page A has pages T1...Tn which point to it (i.e., are citations). The parameter d is a damping factor which can be set between 0 and 1. We usually set d to 0.85. There are more details about d in the next section. Also C(A) is defined as the number of links going out of page A. The PageRank of a page A is given as follows:

$$PR(A) = (1-d) + d\,(PR(T1)/C(T1) + \ldots + PR(Tn)/C(Tn))$$

Note that the PageRanks form a probability distribution over web pages, so the sum of all web pages' PageRanks will be one.

Figure 27: PageRank algorithm.[65]

These 'backlinks' or citations provide a type of 'peer review' with a page's significance derived from links originating from 'important pages' rather than average pages. Whereas '[m]ost search engines associate the text of a link with the page that the link is on', Google's novelty is to 'associate it with the page the link points to' (ibid). Simply put there is a PageRank assigned to each page in the index that is calculated 'by looking at the PageRanks of the sites pointing to it as well as the number of links from each of those pages to alternatives sites' (Halavais 2009:67). These searched pages were found because there were links leading to them contained

[65] Notice the usage of the word citation: 'We assume page A has pages T1...Tn which point to it (i.e., are citations)'.

on other previously encountered pages, yet they differed in importance. 'Treating all links as an equal vote is a bit like treating all roads the same when some are driveways and others are multilane highways' (ibid).

Whereas other search engines carried out rankings based on 'content scores', Brin and Page 'revolutionised the field of the Web' by 'introducing the notion of an importance score, which gauges the status of a page, independently from the user query, by analysing the topology of the Web graph' (Franceschet 2010:2.0). On January 28, 1998 Lawrence Page, Sergey Brin, Rajeev Motwani and Terry Winograd published *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 backlinks is high (Page et al. 1998). 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' (ibid). Page et al. devised a scheme on what were to happen if a 'real Websurfer ever gets into a small loop of webpages' and becomes bored, through the notion of a 'random surfer model', which was defined through a 'user defined parameter' (ibid).

In turn, Brin and Page conceived of PageRank as a 'model of user behavior' and introduced the figure of a 'random surfer', 'who is given a web page at random and keeps clicking on links, never hitting "back" (1998). Rank is awarded, in part, by the time the random surfer spends on a page, yet when the surfer gets bored and starts on another random page they are 'teleported, like a Star Trek character, to that page, even if there exists no link connecting the current and the destination pages in the Web universe' (Franceschet 2010: 3). Brin and Page note an 'important variation' of PageRank that includes adding a 'damping factor' (0.85) to a single page, or a group of pages 'meaning that after about five link clicks the random surfer chooses a random page' (ibid).[66] This 'damping factor' was applied because otherwise there would be a 'rank sink', in which the same links would accumulate rank at each pass therefore defeating the purpose of PageRank. Preventing the creation of false nodes or 'link spam' is why PageRank uses a variant of the previously mentioned 'eigenvector centrality'. The damping factor 'works as a deterrent against those mislead[ing] the system in order to get a higher ranking' (Brin and Page 1998) and got rid of these artefacts, with the phenomena of 'teleportation' described above.

The damping factor is represented by a parameter 'α' which indicates the distance between links, or 'how far the structure of a network should influence the status of a node' that determines importance. Where the value of 'α' is small, this would be similar to the balloting scenario described by Katz and only a certain number of votes would be the determining factor. Once again PageRank remained cryptic as to how the 'rank sink' and 'random surfer' exactly 'relates to the calculation of an importance score' (Rieder 2012:9). Therefore, Rieder argues that

> [t]he value Google usually communicated for α was 0.85 and at this level, the often-heard interpretation of PageRank as essentially an expression of popularity misses the target. While sheer numbers certainly do count, the idea that PageRank interprets the Web as 'a gigantic popularity contest' is simply too imprecise (ibid:10).

[66] 'Thus, the new importance score of PageRank augmented the traditional content score of search engines' (Franceschet 2010:2.0).

Instead it becomes a question of power centrality, referring back to Bonavich's contribution 'ß' as the criteria to assess the varying conceptions of what constitutes power in different social networks. In other words, at times the links can be negative—where connections to power are reflected as 'detrimental to rank'—in other cases low status nodes are beneficial because they can be exploited, or 'power comes from being connected to those who are powerless' (ibid).[67] When applied to PageRank, depending on the damping factor, there would exist different types of 'governmentality': for example if 'α' were greater than 0 it would be a democracy of one person one vote, when greater than 1 it would look Machiavellian, with patronage models of a feudal system. Rieder's full quotation is instructive, along with his emphasis on elucidating how visibility distributes power in an economy of attention:

> If we consider the Google search engine as a central site of power negotiation and arbitrage for the Web, an analysis focusing on *PageRank* —which in practice is certainly not enough—would have to conclude that its authority ranking mechanism applies, in a universal fashion, a largely conservative vision of society to the document graph in order to 'pick out the *real* leaders' and distribute visibility to them. Rather than showing us the *popular* it shows us the *authoritative*, or, to connect back to citation analysis, the *canonical*. If we consider the link indeed as 'innocent,' as a valid indicator of disinterested human judgment, *PageRank* shows us a *meritocracy*; if we take the link to be fully caught up in economic forces however, we receive the map of a *plutocracy*. The search engine as a visibility engine subjects both to the self-reinforcing dynamic of *cumulative advantage* (ibid).

7.0 Economies of Attention

7.1 Advertising and Traffic

In a nutshell, then, PageRank organises relevance by deciding which links receive visibility and in this way, directs the flow of attention on the 'information highway' of the web. Yet 'this commodification of attention occurs in a largely invisible way' (Halavais 2009:83). Economically speaking, search results are an expenditure similar to TV programming in that the service is given away for free in order to attract an audience which can then be sold to advertisers by the provider. In 1981 Dallas Smythe first articulated this strategy of collating attention into the stream of information as the 'audience commodity', offering a 'Marxist analysis of the role of communication for Fordist capitalism' (Beverungen 2015). As Smythe pointed out, television entailed a network that financed the production and then packaged the content to sell to advertisers, often described as the 'eyeball effect', where more viewers exponentially increase profits. According to Elizabeth an Couvering, search engines are not just a case of technological development but are 'navigational media' which she defines as a 'type of technically-based media actor that organises and directs audiences or users to various types of content' (2010:225). Specifically in regard to the development of search engines in the 1990s, the browser was seen as the crucial point for audience aggregation' (ibid:101). it was not only about finding an audience to consume (and pay for it) (ibid:92), but also converting attention into action, enticing users to click on links and interact with the algorithmic interface.

[67] For example, Charles H. Hubbell showed that 'receiving a negative endorsement from a member of negative status makes a positive contribution to the prestige of the endorsed person (if the same Mafioso opposes you then your reputation might rise' (Franceshet 2010:5.3).

> Search engines extract value through the freely given labor of millions of people, and, by reconfiguring it, use it to draw attention to themselves. That attention is then sold to producers who wish to advertise their products (Halavais 2009:83).

In 1998, seven months before Google, one of the key developments in the search industry in response to spam was the search engine 'GoTo.com' (formerly Overture), which implemented the purchase of search terms (keywords) by advertisers. Similar to the Yellow Pages, websites wanted to be on top of the GoTo.com SERP (Search Engine Page Results) for specific keywords. Instead of a wide range television broadcast audience, GoTo accentuated the niche customers in the long tail, matching advertisers to specific visitors to the site. 'These 'bid-placements' operated as an auction, where advertisers competed with each other and only paid when users clicked on their ad listed in search results. Van Couvering structures this advertising through three key characteristics:

> 1) it is priced on a cost-per click basis; 2) it is contextual, linked either to page content or to the users' search term; 3) it is syndicated to other websites on a revenue-sharing basis (i.e. the fee is split between the owner of the website and the provider of the paid search service)(2010:115).

GoTo's model was not to make advertisers pay with CTM or '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' (ibid:113). 'Paid search' was GoTo's innovation with its cost-per-click (CPC) advertisements and these were connected to user traffic, syndicated or otherwise. Therefore, by the end of the 1990s, if not even earlier, commercial interests had already been woven into the very fibre of modern media networks through legislation, market mechanisms and the like (McChesney 1999 cited by Halavais 2009:169).

Referencing the section *Academic Search Engine Research* (1.3.2) in their text, Brin and Page explicitly spelled out their vision on design goals that could be considered a small critique of the commercialisation of web search development.

> 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 (1998:108).

Moreover, in their *Appendix A:Advertising and Mixed Motives* they explicitly speak of 'bias' and that income derived from ads 'provide an incentive to provide poor quality search results'(ibid). Further on they assert that less advertisement would be needed to help the consumer find what they want and that this is what constitutes a 'better' search engine because advertising always wants customers to acquire new products (ibid:107). With the understanding of their search engine 'as free of the "mixed motives" that coloured other search engines whose business model depended on advertising', they even cite the critique of the 'concentration of media ownership' with Ben Bagdikian's *Media Monopoly*, 'a book that in retrospect seems to have served them more as a how-to guide than as a warning' (Peters 2015:326). In spite of these statements, it is noteworthy to show their resolve in solving 'the problem' at hand—making PageRank economically viable.

Although Brin and Page did not originally wish to mix organic search with paid ads (Levy 2011), by 2000 Google were 'cold-calling people, trying to get them to buy keywords' and even 'rolled out a new, self-service advertising product called AdWords that allowed businesses to purchase text ads on search-results pages' (Oremus 2013). In 2002 Google launched an auction-based search-advertising AdWords Select, which replaced their old AdWords with a CPC (cost-per-click) model, or CPA (cost-per-acquisition or a 'cost-per-action') model where the advertiser only pays when a user clicks on an ad.

> [W]hen people went to a search engine, they were often in search of something, and that might make them more prepared to have their interest piqued by a banner ad, or, better yet, an appropriate advertisement right there next to the search results (Halavais 2009:78).

Revenue and viewer clicks increased with Google having a search engine that was not dominated by spam and which could control what the eyeballs saw, as not all advertisements were equally attractive.[68] On August 19, 2004 Google had its IPO, valued at 27 billion dollars.

> The success of Google's highly original business model is 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 (Kaplan 2014:57).

By 2005, 'Cyberspace' (Gibson 1982), as the internet was affectionately termed back in the early 1990s, had morphed into 'Cybercapitalism' (DeLillo 2003), where search results were shaped by a highly intricate series of communication networks and commercial platforms that incorporated advertising. Once considered portals, as explained in Chapter 2, search engines were now 'connecting users to advertisers, within a burgeoning media economy' and this shift in the search engine industry meant they were able to combine 'disparate technical infrastructures' (Zimmer 2008 cited by Van Couvering 2010:124) and diverse types of media including traditional media and communications. 'If there is an emerging attention economy, the search engine is its trading floor' (Halavais 2009:71) and obtaining attention, in the form of hyperlinks, is a kind of enduring wealth, as long as the links are not broken and 'alive.'

Whereas, in the 1990s, the search engine industry was built upon the 'supply chain for audiences rather than for content' (Doyle 2002:18), by the 2000s, the keyword attention economy of PageRank directed 'the creation and exploitation of a new commodity for media: traffic' (Van Couvering 2010:92). User 'traffic' was made possible first through mainstream distribution models and later on through search platforms and hyperlinking in the mid 2000s (ibid). As Halavais points out, this battle for attention enables the 'ideology of the marketplace to be granted access to new areas', as 'search engines were becoming one of the most visited kinds of sites on the web; traffic alone made them attractive' (2009:78). Yet if search engines needed to provide access to information for all users, it was not only the 'backlinking' structuring information; the ranking of search results determines which links the user will click on, therefore creating more traffic.

[68] This is also the innovation of personalisation, where ads are customised for particular users, which I will address in Chapter 5.

In *Googlearchy,* Matthew Hindman elucidated the inequalities of PageRank through the lens of traffic, which measured the visibility of any site based on its search result ranking and the number of links pointing to it:

> Links do not just provide paths for surfers…If links help determine online visibility, how links are distributed tells us much about who gets heard on the Web…The importance of links challenges notions that online equality is easy or inevitable (ibid:132).

By the mid 2000s, political discourse was already filtered thanks to *Googlearchy*, thus 'deliberative democracy' was prohibited by the infrastructure itself—'the social, economic, political and even cognitive processes that enable it' (ibid:130). *Googlearchy* purported that 'niche dominance', where only a small portion of websites receive most of the traffic, is self-perpetuating—the sites with more links receive greater traffic whereas those with few links are harder to find and require better searching skills (2009:55).[69] With hyperlinks continuously being added and Google collecting data ad infinitum, the bias in search engine results simultaneously became more noticeable—'bias that invites users to click on links to large websites, commercial websites, websites based in certain countries, and websites written in certain languages' (Van Couvering 2010:3).

7.2 Network Surplus Value

Whereas in the Fordist era the 'audience' was sold to advertisers, in the 'new economy' (Kelly 1998) it now became the 'trusted user', in Google terminology, who was the preferred commodity. In a Post-fordist context, value emanates from instant communication online: search queries, blogposts, comments, likes and tweets. Marx's 'hidden abode' (1867) —the site of labour and its production of surplus—has now shifted to the immaterial (Böhm and Land 2012). Assuming that users online are not only watching or viewing but also producing content, what Axel Bruns describes as 'produsage' —production and usage—they are thereby 'working' (2008).[70] Building platforms of self-expression, users join the supply chain that connects them to other users, with this new model of production changing the way people generate content online and produce relevance. 'In order to create sociability in the space of flows people first have to make themselves visible, that is, they have to create their representation through expressive acts of communication' (Stalder 2010). Part of the social and economic burden is to participate in the reputation and attention economies, expressing individuality and uniqueness within social networks as well using search engines, namely Google.

With algorithms directing users' attention, users became entrepreneurs or 'homoeconomicus', selling a brand of 'creative' self-management in an online reputation economy and, as they engaged in digital activities, left traces of data everywhere. Moreover, in an era of 'big data' (Mayer-Schönberger and Cuckier 2013), the technologies in which human subjectivity is captured diverges from the industrial era, with the labour of searching delivering free user-generated content to corporations, such as Google. As a consequence of 'the network effect' more people contribute online because others also choose to do so, causing the value and power of the network to increase exponentially as it grows (Leach 2014). Following Yochai Benkler, instead of engineers determining what is most relevant, the critical innovation of Google is

[69] The websites that are the most heavily linked 'rule'.
[70] Whether the (prod)user should be remunerated for the production of content and the further questions surrounding the terms of the digital labour involved in regard to search habits is beyond the scope of this thesis.

outsourcing the decision and labour to the community of the web (users) as a whole (2006:242). It is this 'social networking', where

> people on the web decide what counts literally, with whatever they like for whatever reason, vanity, pleasure, and they produce links, which are connected to each other, and these are counted and eventually monetised (Benkler 2008).

It was PageRank that introduced the 'original notion of quality' or relevance on the web, determined by the 'collective intelligence' of those on it and the 'opinions of the millions of people that populate this universe, is exploited to determine the importance, and ultimately the quality, of that information' (Franceschet 2010:6.0). In this sense social relations become significantly more important than they ever were as an economic phenomenon, where the labour of users clicking willingly on links that interest them adds to 'social networking' (Benkler 2006) yet the criteria remain invisible. Matteo Pasquinelli elucidates the exploitation of users' cognitive intelligence through the hidden 'immaterial factory' of Google, with each link that is clicked on adding value: 'If commodity is traditionally described by use value and exchange value, network value is a further layer attached to the previous ones to describe social relations' (2009:156). Although PageRank determines the ranking of these links, Google itself does not produce content; rather it is expropriated in the form of 'cognitive rent' (ibid). 'Accordingly, Google can be described as a global rentier that is exploiting the new lands of the internet with no need for strict enclosures and no need to produce content too' (ibid:159).

By determining its own attention economy with these already 'trusted' links, PageRank captures 'living time and living labour time and transforms the common intellect into network value' (Pasquinelli 2009 cited by Bilic 2017:13). The re-appropriation of network value through the labour of user interaction and engagement is the common intelligence that Benkler (social production and networking) and Pasquinelli (network value) articulate. 'Google's implicit design decision, is "an intricate form of populist hypermedia" as Kleinberg put it', where websites make themselves found, thereby reading 'the web as the embedded intelligence of millions of users' (Peters 2015:327). With this 'cognitive capitalism' (Boutang 2012), PageRank was able to restructure links to such a degree as to create a 'link economy', turning the hyperlink into the 'currency of the web' (Pasquinelli 2009; Helmond 2013). Value is commensurated by the amount of hits, or traffic to the site and how the 'network surplus value' of the nodes is redistributed (Pasquinelli 2009).

> Indeed, PageRank produces what Deleuze and Guattari described as machine surplus value referring to the surplus value accumulated through the cybernetic domain, or the transformation of a surplus value of code into a surplus value of flux (ibid:156).

The value of these social relations is then measured and PageRanked, making the amount of attention received by a particular text or website visible thus adding even more value or trust to the user.[71] Along these lines, Brin and Page mention that PageRank carries out judgments that reflect the personal choices and affiliations of the 'user' who became central to their terminology of human computer interaction. The 'number of new users inexperienced in the art

[71] Produsage is a portmanteau of production and usage coined by Axel Bruns. Whether the (prod)user should be remunerated for the production of content that is difficult to produce and the further questions surrounding the terms of the digital labour involved in regard to search habits is beyond the scope of this thesis.

of web research' (1998:107) increased and eventually became 'trusted users', which I will further discuss in Chapter 8. In their 'Future Work' section, Brin and Page explain that they foresee applying the 'trusted user's' search history to PageRank, which 'can be personalized by increasing the weight of a user's home page or bookmarks' and 'result summarization' text' (ibid:6.1).[72] Moreover they also 'plan to support user context (like the user's location) and wish to 'extend the use of link structure and link text' (ibid). Over the years, by pinpointing locations, search histories, IP address and tracking their 'usage' with the collation of data, Google created personalised search results, which I will return to in Chapter 5.

However, Brin and Page mention that companies whose business model is co-opting users' attention and manipulating the 'unseen' metadata contained in search results for profit is a 'serious problem', yet they are keenly aware of data produced 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 1998:109).

Already presaging the coming era of 'dataism' (Harari 2001), the more people used PageRank the more it improved and Google simultaneously constructed a proprietary database. With the building of 'possibly the most lasting, ponderous, and significant cultural artefacts in the history of humankind', Google's 'Database of Intentions' constantly captured users' search queries and histories thereby enabling 'a new culture to emerge' (Battelle 2005:7 cited Noble 2018:148). Subsequently search data became a commodity, along with monitoring and influencing user behaviour in an era of 'surveillance capitalism' (Zuboff 2015), which I will address in more detail in Chapter 9.

In their Appendix A, Brin and Page explain that web search 'remain[s] largely a black art and to be advertising oriented' (1998) and that they 'believe the issue of advertising causes enough mixed incentives that it is crucial to have a competitive search engine which is transparent and in the academic realm' (ibid:107). Referring back to the intertwining of research and search through relevance, academic research was determined by a few experts or peers, where '[c]itations generally witness the use of information and acknowledge intellectual debt' (Franceschet 2010:6.0). As shown previously, the shift to hyperlinking URLs (Uniform Resource Locator) is a reflection of the authoritative pages found when surfing the web, showing 'preferential attachment' (Halavais 2009). This linking paradigm in turn reflects the analysis of attribution via the hyperlink as the essential structure or 'mesh' of the web (Berners-Lee 1989) indicated by *associative linkages* between the various parties, as mentioned in Chapter 2.

> Google treats its search algorithm […] like a happy-go-lucky pragmatist willing to crawl the *snail trails of associations* wherever they lead […] A page is valued by how other actors in the system value it, and their power to value it is determined by the value that others place in them' (Peters 2015:330-331 emphasis mine).

Nowadays the importance of a contribution is based on the 'collective evaluation' of bibliometrics—how citations from the academic community are measured and commensurated through attention.

Additionally, Brin and Page impart that besides being a 'high quality search engine', Google is 'a research tool', and they close 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' (1998:112). They point out that '[a] Web search engine is a very rich environment for research ideas' and they also state their goals of 'pushing more understanding into the academic realm' (ibid). The authors also compare the increasing amount of documents available due to the growth of the web with the lack of research documents about search engines, lamenting that so few are willing to discuss the details of the technology in academic papers. 'According to Michael Mauldin (chief scientist, Lycos Inc) [Mauldin], the various services (including Lycos) closely guard the details of these databases' (ibid:107). This is quite contradictory in regard to their own 'visibility management' (Flyverbom et al. 2016) regarding the PageRank algorithm as a black box, or for some a black art similar to an alchemical formula that performs for its creators, in this case a corporation that maintains its 'secret recipe'.

Lastly, in their 'Future Work' section Brin and Page state that much 'remains to be done'—'updating' is an area that needs attending to, along with 'smart algorithms' that discriminate what should be recrawled and new ones to be crawled (1998). This foreshadowing of artificial intelligence applied to search has now become a reality with the shift from PageRank to RankBrain, a 'machine learning' algorithm, which I will explain at the end of Chapter 5. But first a reflection on methods where I tested out Google Search and PageRank with specifically chosen keywords in a self-designed experiment, gathering data on myself in the hope to glean insight into its proprietary and hidden algorithmic ranking procedures and processes.

Chapter 4:

Reflection on Methods

1.0 Methods Assemblage

The previous chapters explained the 'pre-history' of search engines along with the first 15 years of Google Search. In order to experiment with how search works, or how search could work differently, I now turn to my methods. As argued in John Law's *After Methods: Mess in social science research*, method is performative and helps produce realities, it is not a 'more or less successful set of procedures for reporting on a given reality' (2004:43).[73] For the social sciences to create these realities, a broader understanding of methods is needed—a 'methods assemblage'. This functions as a 'reality detector' and 'reality amplifier'; there is no neatly ordered account because 'the realities they stand for are excessive and in flux' and not 'simply organised' (ibid). With its constant state of 'overwhelming fluxes of the real' (ibid:116-117), this 'reworking' and 'rebundling' of realities is a creative act that produces 'new signals' and 'new resonances' that are continuously being 'remade' (ibid:43).

Part of my 'methods assemblage' includes my 'serendipitous search' for knowledge, incorporating numerous 'offline' encounters with academics, researchers and artists at courses, conferences and symposia. Additionally, some of the references I found when searching with Google online and I have used in my PhD will contribute to the citation process (SCI), which I explained in the previous chapter. However, few published academic articles can keep up with Google's public announcements of adjustments to their secret algorithm. Instead I found online forums and other SEO affiliated marketers' blogs that elucidated how they 'seek to game an algorithm to optimise an outcome for a client' (Kitchin 2017:24), which I elucidate in Chapter 5.[74] As stated in the Introduction I also draw upon works of literature, primarily *The Circle* by Dave Eggers and *Bleeding Edge* by Thomas Pynchon from 2013 that, as 'novel territories', provide a means to imagine alternative pasts and potential futures of how search engines organise (us)ers. My methods assemblage primarily consists of the following three methods, which are all intertwined.

1.1 Critical ethnography of the self

In her text *What is 'critical' in critical ethnography?* Chia-Ling Wang suggests that the self could also be an ethnographical field to be explored critically, with a view to 'achieving a new relation in our knowledge of ourselves and in order to transform our subjectivity' (Wang 2008:4). Wang cites Foucault's lecture *What is Critique* as it relates to the 'contact between technologies of domination of others and those of the self I call governmentality' (1994:225). However, the term 'resistance' is also bundled up in the term 'critical'. Thus, Foucault's first definition of critique can be underlined with the words: 'the art of not being governed like that and at that cost' (1997:29), which formed the departure point of this thesis. My desire was to find non-commercial ways of seeking information. I am specifically addressing the inequalities

[73] The 'performative' is of course a broad subject beyond the scope of this thesis. I 'simply' refer to Law's understanding that technologies are enacted, multiple, performed and contextual, and arise from projects (Law and Singleton 2000) along with his practices, such as 'heterogeneous engineering' (1985).
[74] Search Engine Land and MOZ write about the SEO industry.

and injustice of the corporate control of daily search activities and the way that knowledge is being censored through Google's advertisement models, personalisation and surveillance capitalism (Zuboff 2015), which I will explain in the forthcoming chapters.

Following Alexander Galloway, the critical is defined as that which is 'contestational' to hegemony, such as Google Search. In *The Cybernetic Hypothesis*, Galloway attends to the problem of hegemony that is 'not simply limited to a hierarchy of domination and subordination' but whether or not 'certain claims about knowledge or reality are recapitulative or critical of the hegemonic position' (2014:126). Moreover, he emphasises the ability to identify critique 'as a specific shift in the relative value of recapitulative versus contestational claims', with his 'update' of the Kantian question:

> Is thought as such dictated by the regularity of an inherited structure, or is thought only possible by virtue of an asymmetrical and autopositional posture vis-à-vis the object of contemplation? Having inherited the computer, are we obligated to think with it? (ibid).

In Chapter 1 I explained how the 'address office' brought people together who had similar interests, organising their data in one place—a register or 'Protocollbuch'. Nowadays, evidenced through the very act of conducting research, where the habit of searching—querying, clicking on links, reading and writing—is captured by computers instead of registers, it exemplifies what Foucault described in his text *Technology of the Self* (1994). The 'technology of the self' consisted of the *ethopoeien* that 'means making ethos, producing ethos, changing, transforming ethos, the individual's way of being, his mode of existence' (ibid:237). The ethopoiein was in turn comprised of two forms, *correspondence*, not personal accounts but rather the 'recurrence of discourse by the "citational" practice under the seal of age and authority' and hupomnemata (ibid:271). With the former, I carry out correspondence with algorithms and conduct research, which also includes referencing the knowledge of others, as part of contemporary communication circuits, merging online search with 're:search', as shown in Chapter 3. The latter, hupomnema, was a personal archive in which one writes down what one has read and heard (or in my case, searched for) whilst preserving the ethos (ibid:237).[75]

Foucault states that hupomnema has a 'very precise meaning: it is a copybook, a note' and this technology was 'coming into vogue' for 'personal and administrative use' in Plato's era yet it was as disruptive to Greek society as the computer's invasion of the private sphere is today (ibid:272). He enumerates that hupomnemata were the 'technical and material framework' from which writing and the self came about and that it didn't matter 'whether a text is written or oral' because the problem concerned 'whether or not the discourse in question gives access to truth' (ibid). Hupomnemata and the 'culture of self' converge with the 'government of the self', with Foucault providing a historical comparison between 'governments and those who managed enterprises'—writing down their administration in registers—to those among the ancients who 'carried on this politics of themselves with these notebooks' (ibid).[76] Harkening back to the registers in Chapter 1, the managing of oneself is carried out through these 'new instruments' in a 'permanent relationship' to oneself—'one must manage oneself as a governor manages the

[75] I was reminded of the Foucault's warning in regard to the archive: 'The archive is neither the sum of all texts that a culture preserves nor those institutions that allow for that record's preservation. The archive is rather that "system of statements," those "rules of practice" that shape the specific regularities of what can and what cannot be said' (Foucault 1972:79-134).

[76] Although Foucault cites earlier sources, it was commonplace in literary studies to recognize Montaigne as the 'first great autobiographer' (ibid: 276), who wrote about his fathers' techne of oikos (Chapter 1).

governed, as a head of an enterprise manages his enterprise, a head of household manages his household' (ibid).

Foucault emphasises the shift from the practice of 'knowing thyself' to the ethic of 'taking care' of oneself that developed 'toward definite objectives such as retiring into oneself, reaching oneself, living with oneself, being sufficient to oneself, profiting by and enjoying oneself' (ibid:274). He continues by stating

> the principal work of art which one must take care of, the main area to which one must apply aesthetic values, is oneself, one's life, one's existence… It was a question of making one's life into an object for a sort of knowledge, for a tekhne-for an art (ibid).

Thus the *Technology of the Self* was also a techne for an art, where the question of producing knowledge reappears fused with certain kinds of knowledge about oneself and is related to what Foucault defined as 'truth games'. Whereas Foucault's *Technology of the Self* refers to daily writing, for me it is searching—captured and annotated through the 'mediality' of hupomnemata. Rather than notes, registers, ledgers or paper it is hupomnemata—computers (hardware) and their ubiquitous unseen interfaces (software), with which I searched as a 'critical ethnography of the self', simultaneously citing the works of other researchers.

Referencing Kittler's 'commands, addresses and data' (1999), 'prescriptive' commands play an inherent part in the organisational aspects of this 'critical ethnography of the self'. According to Ursula Franklin, the 'prescriptive' 'is a whole set of social arrangements, characterised by the pervasive mechanisation of mass production, all bundled together in a single social system' (1989). 'Prescriptive' technologies are then a practice, how it organises work and people (1989:04), as it is an 'important social innovation' that commands, 'because they are designs for internal compliance and discipline, order and obedience' (ibid). Reflecting on the ubiquitous technologies of her time, Franklin links the action of typing on a word processor to a workstation (office) with a system where 'assignments can be broken up and timed' and the 'interaction between operators can be monitored' (ibid). Moreover, prescriptive technologies 'become normalised because everyone thinks there is only one-way of doing it' (ibid).

Franklin deems this commandeering as 'a seedbed for orthodoxy' that makes those who participate enter into a milieu of external control where they become acculturated, which is analogous to the hegemony of Google Search today (and personalisation). Constantly receiving and subsequently clicking on links, as a personalised subject I govern myself voluntarily and constantly in a 24/7 field of operations. Because there were no colleagues readily available to answer my technical questions, I often consulted the 'oracle' (Google) on my personalised Apple computer for advice, or read manuals, visited websites or watched 'how to guides' and tutorials. I received recommendations (I allowed this in the Google settings) on YouTube and I incorporated personalised search results into my research, which I will discuss in forthcoming chapters.[77]

[77] Owned by Google since 2009, YouTube is a database, which can be searched and has its own recommendation algorithms that provide suggestions to users in the playlists. I watched the suggested 'Eli the computer guy' videos on YouTube who advised me how to search with Tor and TAILS on the Dark Net for Chapter 6.

1.2 Experiment in Living

In her text on 'Configuration' in *Invented Methods*, Lucy Suchman describes 'methods assemblage' as 'a device for articulating the relation between the "insides" of the socio-technical system and its constitutive "outsides", including all of those things that disappear in the system's figuration as an object' (2012:55). Law contextualises these processes through the lens of Post-structuralist representations of the Other, which is 'hidden, repressed or uninteresting' (Suchman 2012:55). I apply Law's 'hidden Other' to those constantly fluctuating algorithms I could never capture; as an observer I could only guess at how I was personalised or profiled or why I obtained certain results. Drawing on the work of Bateson and Meade (1973), Gordon Pask (1993) articulates the role of the observer in the 'new cybernetics' (or second order cybernetics) as an ecosystem with feedback and this 'organism-plus-environment' as a single circuit.[78] (Figure 28) However, within this socio-technical system of the black box its contents remain invisible to the researcher (myself).

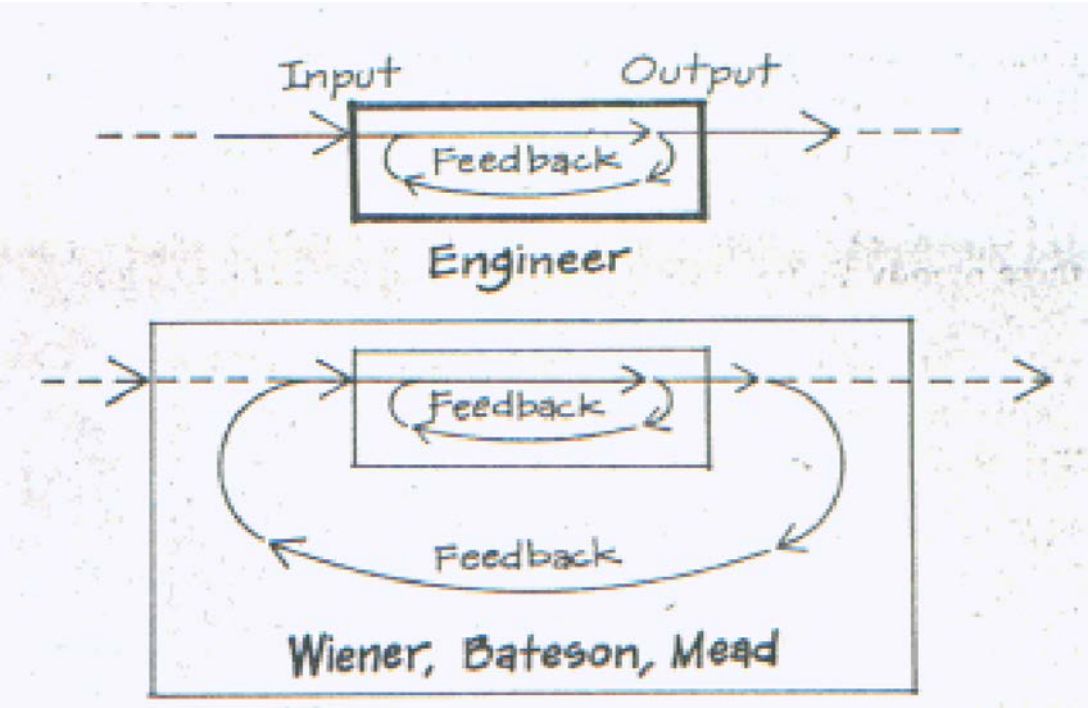


Figure 28: Gregory Bateson, Margaret Mead diagram comparing first and second order cybernetics during an interview (1973). http://oikos.org/forgod.htm (WayBackMachine) Internet Archive.

It was W. Ross Ashby who suggested in 1956 that the observer could never really see what goes on inside the black box. Instead, it is the interaction of the inputs and outputs of the black box as described by the observer which offer explanations, as products of an 'observer's interaction with the 'whatever-it-is' (1962). Echoing Ashby, Rob Kitchin also states that there are 'two openings that enable lines of enquiry: input and output' (2017:24), even though corporations 'remain opaque' about their software. Therefore, an investigation into what kind of data is 'fed into an algorithm' along with what kind of 'output is produced' are the basic ingredients of the

[78] 'Cybernetics means governance, steering or helmsperson and refers to the scientific study of control and communication in the animal and the machine' (Wiener 1948). I will return to cybernetics in Chapters 5 and 9.

'recipe of the algorithm' as well as the criteria (weights, preferences) in this 'reverse engineering' process (ibid). How could I, without a 'skill set' and limited 'tech-savvy', capture and understand what is going on in this black box? In order to engage directly with these socio-technological ecosystems and the constant feedback loop of searching, I needed to invent a method that explored the hidden workings of the black box and opened up a field of dialogue with the machinic (algorithms) as a form of address. (Appendix E).

Besides *addressing* a specific problem, with an inventive method 'its (repeated) use is always oriented towards making a difference' (Lury and Wakeford 2011:11). Following Dewey's 'identification of the problem' as put forth in the Introduction, I gleaned inspiration from the fact that 'a method must rather be made specific and relevant to the problem' (ibid:2-3). When discussing methods 'the problem at hand' is that devices such as tools, instruments and techniques are always part of an 'assemblage, configuration or apparatus' in the Foucauldian sense of the word (ibid:7). As elucidated in the previous section with regard to hupomnemata or memory extensions, these 'techniques and technologies' render everything accountable in everyday 'living experiments', as part of the methodological repertoire of the social sciences (Marres 2012:82). In her contribution to *Inventive Methods*, Noortje Marres shares the development of John Stuart Mill's 'experiment in living' who 'first used it to make the case for the affirmation of social and cultural diversity, as something that is distinctive of liberal societies' (ibid).

> But the "living experiment" presents a special variation on the theme: unlike scientific experiments, this form of experimentation has been explicitly associated with the moral purpose of the improvement of society since its very inception (ibid).

In order to answer my main research question, *how do search engines organise us(er),* I designed an 'experiment in living' that explored this socio-technological ecosystem and enabled me to capture certain aspects of my human computer interaction: inputs and outputs. My office at CBS became the 'site' of data collection where I typed the same keywords into the search box, almost simultaneously, on two computers: one was an Apple using Google Search with which I was assigned an IP (Internet Protocol) address and the other was a PC with the Tor browser, where my IP address was hidden. I collected data on myself as a 'personalised subject' and 'anonymous user', acknowledging that other unseen actants—algorithms—play a large role. I documented my interactions with the machinic at all stages of these processes—from choosing the keywords and instating parameters, along with all of the unexpected 'happenings' en route. These search experiments are detailed in Appendix D and I discuss my results in Chapters 5 and 6.

With my method, 'experiment in living', search engines 'address' the 'personalised subject' and 'anonymous user' through technical standards (TCP/IP), in the Kittlerian sense. As foreshadowed in the Introduction, Kittler draws on the 'Von Neumann architecture of computers' (registers, busses and random access memory) (2009:30) to show how, as previously articulated, 'mediality' has shifted from form and matter to commands (which I just discussed), address (which I am here considering) and data, which I will come to shortly. The manner in which I was addressed or assigned is a type of '*inscription*' or what Kittler deemed 'Aufschreibesystem' that contains protocols and agents, which interact with each other in a distributed process of communication involving 'coding, signals and decoding' (ibid:44). In other words, 'people use the systems according to their own volition—but that the system is subjecting the user to its structure, or in this case: standards' (Parikka 2012:78). With my semi-

machinic role (scraping and collecting data), I attempted to capture the various forms of address by algorithms, which the subject, or user, engages with when querying with search engines. This 'interpellation' (Althusser 1972) produces various subjectivities, which I will explain in more detail in Chapter 7.

Moreover, being addressed by search engines and their organisational effects brings up new conundrums and concerns. Contemporary theorists Alexander R. Galloway and Eugene Thacker state in *The Exploit* that this 'new physics of organization' reflects the internet and the 'protocological nature' of contemporary control mechanisms—aspects of communication that are responsible for hierarchies and navigation (2007:29). In the contemporary genres of social and cultural research, an 'experiment in living' operates as a 'notable device', one that 'provides a format or protocol for exploring and testing forms of life, which is today widely applied across social life (Marres 2012:78).

> Thus, experiments in sustainable living can be said to undertake the modification of habits and habitats according to a fixed procedure: they are a way of implementing changes in everyday routines and living spaces according to a protocol (ibid).

With my 'experiment in living' I defined the parameters or boundaries of the 'out there' in a continuously dynamic environment—'in the wild' (Callon et al. 2001) of the web. Another parameter that surfaced in my 'experiment in living,' was that the field of my office was in constant flux—my router configuration changed, which 'messed' (Law 2004) with my results and protocols. Additionally, researchers such as Ormen have brought to light some of the complexities:

> Search results are however not easy to archive since the exact rankings of websites are re-evaluated by proprietary and inaccessible algorithms (including Google's PageRank) for each query and thereby subject to change constantly (2013:190).

Phrased otherwise, perhaps an 'inventive method' can make a difference to the manner in which it 'makes itself' and 'produces relations beyond itself' in its deployment (Lury and Wakeford 2012:11). The value here lies in its ability to 'embrace the variability of human life' using alternative methods by those 'who take up the challenge' and 'come to terms with particular social and technological changes that are currently affecting social research' (ibid). Behind the search interface on screen, the black-boxed code of corporate algorithms are non-transparent, along with protocols and infrastructure. Yet in a way I was carrying out qualitative 'interviews' with algorithms with my inputs, where the outputs were responses. Viewed in this way, as time-based moments, my 'interviews' with algorithms took place not with words but with code.

1.3 Data Visualisation as Transcription

> *Are you a 1 or a 0?*
> *If you stay you can change the world, you become a yes, 1.*
> (Mr Robot: Season 1. Episode 2. Time code: 25:00)

The capturing of screenshots during my 'experiment in living' does not collect data, only an image of the results. Yet saving 'complete webpages' of my interaction with algorithms captures code that can be processed afterwards. In order to 'make sense' of the small data sets I collected

from my search 'experiment in living' I needed to visualise them. In his text *Are some things Unrepresentable?* media theorist Alexander Galloway addresses the visualisation of information and data through the dilemma of 'unrepresentability'.

> Either *data* offer *zero* help as to how they ought to be aestheticized, or they eclipse all available possibilities under a single way of seeing. One might assign a name to this curious contradiction. One might call it the dilemma of unrepresentability lurking within information aesthetics (Galloway 2011:91 emphasis mine).

Furthermore, Galloway pinpoints the problematics of 'unrepresentability' in 'information aesthetics' [data visualisation] through the lens of discriminating between 0's and 1's by applying a 'dialectical logic'. He first maps out the etymology of data (that which is given) with a double entendre on the word 'données'. Bundled up within this 'gifting' is the 'ontological rawness' of data, as they are not just measurements or recorded facts (ibid:87). Yet there is confusion between the term 'data', which is considered raw and numerical, and 'information'—that which has been 'given' form and he discriminates between them (ibid:81). The first thesis is that 'data have no necessary visual form and are represented by a '0'. The second thesis is that only one visualisation has ever been made of an information network' because all 'visualizations look the same' (ibid:90) and this is represented by a '1'.

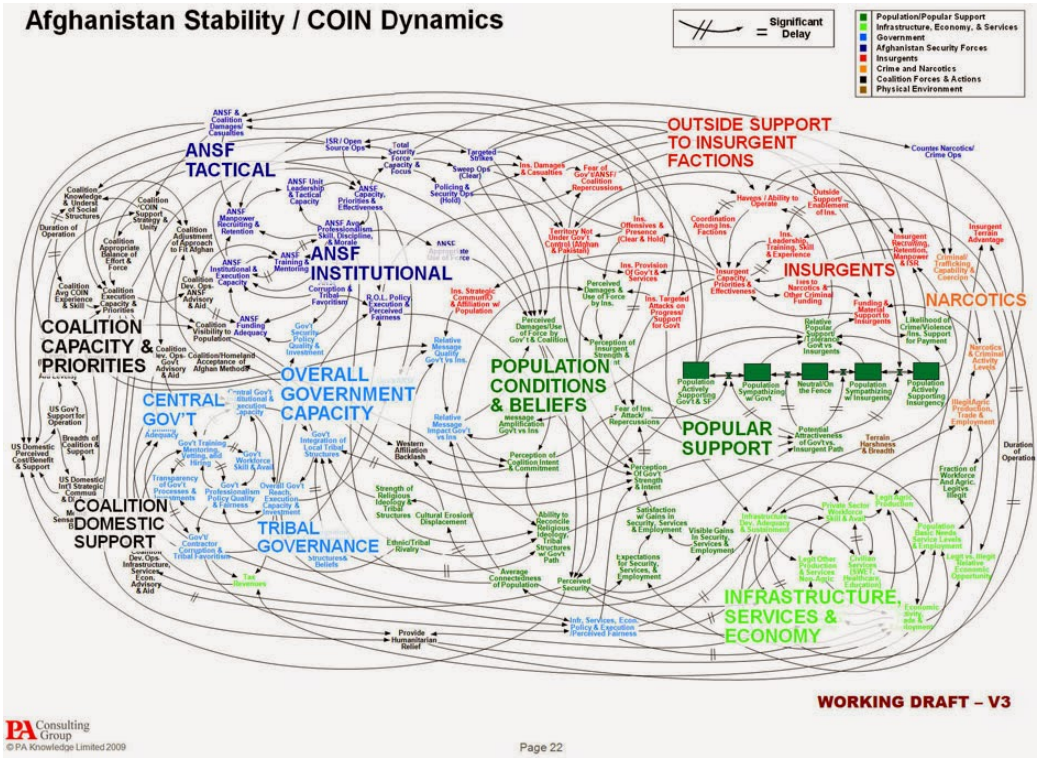


Figure 29: Afghan Stability/COIN Dynamics used by McCrystal (2009).

He employs US General Stanley McCrystal's power point slide (2009) depicting the American military strategy in Afghanistan as an example of his '1'.[79] (Figure 29) Yet according to Galloway '1's are not sufficient; they are actually zeros: 'when there is only *one*, there is

[79] There was critique in the media that Power Point software simplifies information, 'edits ideas' and does not link these ideas or facts to any kind of human narrative (Borger 2010). Or, 'Is there another way to present the information that doesn't look like it has been put together by a kitten with a ball of string?' (Rogers 2010).

nothing. For a representation of the *one*, is, in fact, a representation of *nothing*' (2011:90 emphasis mine). Therefore, he argues that 'information aesthetics' actually facilitates a decline in the transparency of informatics.

> The point of unrepresentability is the point of power. And the point of power today is not the image. The point of power today resides in networks, computers, information, and data (ibid:95).[80]

Galloway then continues by developing an argument between ontology and aesthetics:

> Thus if data open a door into the realm of the empirical and ultimately the ontological (the level of being), information by contrast opens a door into the realm of the aesthetic (Galloway 2011:96).

With aesthetics, the countable alphabet of 0's and 1's becomes 'operable and develops organizational powers' yet the data needs to be processed somehow (Beverungen et al. 2019). 'Any visualization of data must invent an artificial set of translation rules that convert [an] abstract number to semiotic sign' (Galloway 2011:96). Yet certain 'critical' cartographies remain incomplete. In this sense, information aesthetics has not been able to represent all that needs to be imaged and, referencing Gilles Deleuze (1992), Galloway proclaims that 'adequate visualizations of control society have not happened. Representation has not happened. At least not yet' (Galloway 2011:95).

Speculative visualisations, cognitive mappings (Jameson 1988), or diagrams of 'work flows' have already happened in artistic practices, such as the work of Mark Lombardi (2001). Within the realm of 'digital aesthetics' this provocation is precarious and tenuous. 'Many information structures have graphical analogies and can be understood as diagrams that organise the relations of elements within the whole' (Drucker 2009:16). However, part of the reason why these visualisations are inadequate is because algorithmic formulae, the code of proprietary algorithms, is unknown and their interaction with humans has not been able to be depicted visually —what is now often called the 'black box' dilemma or 'unrepresentability'. Galloway's provocation incited me to invent a method of converting my search results, making what is not visible or 'representable' legible by transcribing data into information.

In 1997 Ellen Ullman elucidated how users not only interface with software tools and techniques such as the spreadsheet, which are 'maps' of information, but how they demonstrate agency in the process. Even in an era when searching the web was a nascent technology, Ullman keenly articulated how the transcriptive process unfolds with a spreadsheet of 'coding' data, with the human actor giving data form, thereby 'informing' the tools of technology.

> The user gives data its shape—places it in columns and rows, expresses the complex relationships among those columns and rows—and eventually turns data into more knowledge. It is the end user who creates information, who gives *form* to data, who *informs* the spreadsheet (1997:78).[81]

[80] In his article *The Total Archive*, Andreas Bernhard concurs that making this 'distinction is productive' and that it could be applied as a means to 'isolate the rift that exists between mathematically calculated and visualized knowledge' (2015:22).

[81] Perhaps this is what inspired Galloway's distinction between data and information.

During my 'experiment in living' I carried out search queries with Google and Tor on two computers and compiled 'small' sets of data. I provided the input— keywords, or 'terms of art' in order to test out their "currency" in an era of cognitive capital (Appendix A). The data from these qualitative 'interviews'—my output, URLs and webpages of data—can be seen as 'notes' from my fieldwork that I then 'transcribed.' This entails not only collecting, extrapolating, cleaning and (re)structuring data but 'coding' the data. What used to be transcription in ethnographic coding of data becomes algorithmic data visualisation in the era of digital research methods.

Figure 30: Data visualisation as transcription, *Re:search - Terms of Art* (Performativity, Contemporaneity)

Re:search - Terms of Art, produced together with the interactive graphic designer, Richard Vijgen, is the result of this conversion process (Appendix E). (Figure 30) These visualisations of my small data set show how the outputs (hyperlinks) are ranked, along with the similarities and differences between my two search methods: Google Search and the Tor browser.[82] These 'data visualisations as transcription' are only representations of what was captured in a certain time frame and this presence also incites absence, in that they are incomplete. They offer only a glimpse—a peek into the mysterious black box—with the results as URLs. However, this data visualisation process enabled me to compare these two types of online querying as a base for analysis. Therefore, as much as I endeavoured to answer Galloway's call for 'a poetics as such for this mysterious new machinic space' (2011), whether my data visualisations are adequate representations remains to be seen.

[82] I was invited to take part at the exhibition *Hacking Habitat*, Utrecht, NL where I exhibited these data visualisations made in collaboration with interactive designer Richard Vijgen (see Appendix E).

2.0 What I didn't do or couldn't do

If I were writing a thesis about search I could have typed in 'search engines' as a keyword on Google Search and received links for articles or dissertations that dealt with 'search engines' or had 'search' in the title. I consciously did not do this. Alternatively, I could have used corpora already organised and available to me, such as social science databases that were now 'free' and accessible because I was a PhD student at CBS.[83] Besides being advised that I should write an 'article-based' PhD of around 40000 words, I was entering a system enforced by bibliometrics where my 'success' was to be measured by the amount of articles I published in 'high-ranking' journals. In order to understand bibliometrics, I signed up for courses given by the CBS librarians who kindly took the time to explain to me how to use the institution's authorised databases so that I could see what had previously been written in the field and to build an initial 'literature review' in order to identify the 'knowledge gap'. Their advice was to begin with the *Scopus* and *Web of Science* that proffers to 'search and discover the world's most impactful scholarly journals and scientific research'.[84] *Scopus*, or rather Elsevier's website states that,

> Scopus is the largest abstract and citation database of peer-reviewed literature: scientific journals, books and conference proceedings. Delivering a comprehensive overview of the world's research output in the fields of science, technology, medicine, social sciences, and arts and humanities, Scopus features smart tools to track, analyze and visualize research.[85]

Additionally, the librarians at CBS also instructed me in how to search 'efficiently' to find the most relevant articles or journals based on my search terms and advanced search fields.

Although not explicitly stated, I felt I was being encouraged to join in the [Page]ranking system of academia to compile my literature review. However, my tutorials at the CBS library were extremely helpful, enabling me to better understand the relationship between the Scientific Citation Index and Google Search's PageRank, as explained in Chapter 3. Notably, certain librarians downplayed the importance of Google Scholar while promoting the above-mentioned databases. Yet certain academics explained to me how they used Google Scholar to view their ranking and citations, along with how it was the 'most up to date' on publications of their work. Because the very nature of my thesis is about 'search' and how one could find things serendipitously, I decided not to do what many scholars (and librarians) suggested.[86]

During my 'experiment in living' I received ads in my search results. Although I took a one-day class at CBS and learned about AdWords (now Google Ads) and would have liked to

---

[83] It was the first time in my life that I had access to any of these databases and I have to admit I was a bit overwhelmed by the amount of academic articles I suddenly had access to.

[84] Formerly known as the IP (Intellectual Property) and business of Thomas Reuters, *Clarivate Analytics* is what I obtained as my second search result with the keyword 'Web of Science'. Self-branded as the 'world's most trusted publisher-independent global citation database' and 'guided by the legacy of Dr Eugene Garfield, inventor of the world's first citation index, *Web of Science* is the most powerful research engine, delivering your library with best-in-class publication and citation data for confident discovery, access and assessment'. Available here: https://clarivate.com/webofsciencegroup/solutions/web-of-science/.

[85] https://service.elsevier.com/app/answers/detail/a_id/15534/supporthub/scopus/#tips

[86] I downloaded articles from other pay-walled journals such as *Sage Journals*, *JStor* and *Muse* during the course of my PhD.

experiment further, I decided not to focus on this aspect. When presenting my PhD research in progress over the course of recent years a few academics suggested using Mechanical Turk to carry out my search 'experiment in living', with people in various locations in the world querying the same search term. I was fortunate to attend the Digital Labor conference in New York in 2014 and take part in a 'Turkopticon' workshop run by Lili Irani and Mushon Zer-Aviv. After speaking with actual 'turkers' and discussing their abysmal situation and low payments where they do not necessarily get paid even if they do the work, along with considering the ethical concerns regarding certain methods within academic research, I chose not to do this.

Furthermore, I wished to obtain insight into the invisible workings of proprietary algorithms and understand how they functioned so I contemplated doing fieldwork at Google in Mountain View, California. I first consulted academics who knew present and former Google employees and I spoke to them briefly about their knowledge. They couldn't share details about their programming of search algorithms or what they actually did at Google because all employees sign 'non-disclosure agreements'. My experience was similar to Elizabeth Van Couvering who stated that she 'was unable to secure entry into a [Google] search engine company in the requisite timeframe' (2010:89-90).[87] All of the above helped inform my decision not to do a series of ethnographic interviews with 'search engine' experts. I also attempted to visit a Google Data Centre in Belgium and was denied access. I didn't pursue obtaining the archival documents of Google's public changes to their algorithms, which would have taken me years. Moreover, Google has constantly changed its algorithm over time (see Chapter 5).

> [F]luctuations in search rankings are very difficult to explain, since the number of factors informing the search rankings and the individual weight of each factor is impossible for us mortal researchers (read: not employed by Google) to decipher (Ørmen 2013:192).

I discuss Google Trends in the Introduction and Chapter 9, yet I specifically choose unpopular and eclectic keywords relative to my readings, art projects and new media, which despite carrying specific currency to a certain audience, were not 'trending' on Google. I also didn't try to examine log analysis, 'to gain a clearer understanding of the interactions among searcher, content and Web search engine' (Trevisan 2014). Halavais cogently points out (2009) this could be conducted with 'first generation' search engines but nowadays corporations protect their 'user activity logs' for 'commercial reasons and due to the ethical implications of releasing such data in the public domain' (ibid). As Trevisan notes, there is not enough methodological literature on these issues because of the 'excessive emphasis placed on the need for raw search records has created a self-perpetuating impasse that has hindered methodological innovation' (2014). However, there are increasingly more methods that challenge 'the presumed indispensability of activity log data' (ibid).

Potentially (with a team of researchers) I could have built a web scraper to send out automated queries and in Chapter 5 I reference one empirical study that successfully did so. But I did not have the financial means to hire computer programmers to facilitate the use of proxy servers or to create fake user profiles and programme bots to carry out my search queries, nor the technical skills to do so myself. Although 'Google has previously explicitly banned this option in the Terms of Service and it is generally seen as a 'dirty research method' (Ørmen 2013:191), there are those researchers who think otherwise:

[87] She also writes: 'In fact, as one of the people in charge of relations with outside researchers for Google remarked to me, "There are floors in this building where even we can't go"' (Van Couvering 2010:74)

> And, though (according to the terms of service) you are not supposed "access [the information] using a method other than the interface and the instructions that we [Google] provide", nothing prevents you from being methodologically inventive and re-using this information for purposes other than those which were originally intended (Lury and Wakeford 2012 cited by Venturini et al. 2018:4197).

Using the API and having server-side access would have been an option yet these proprietary 'functionalities' could change at any moment (ibid). 'The official API to search through the entire index of Google Search was discontinued by Google as of November 2010 and replaced by a Custom Search API that offers very limited search options' (Ørmen 2013:191). I also didn't use longitudinal method (various constellations of queries and results at regular intervals over time).

Also, there were so many parameters that changed during the course of my search querying that it was problematic in relation to other 'scientific' studies in regard to its 'repeatability'. This would be considered highly problematic within the field of science, where '[r]eplication is figured as one of the cornerstones of scientific progress and encourages replication of established studies in order to test their validity and reliability' (Blackman 2016:16). I, too, was warned that my work would not be 'empirical' because its results cannot be reproduced. Not only was I not able to reproduce results, the very comparison between my querying with Google and Tor was never carried out exactly at the same moment, as it was conducted by hand. However, as I have hoped to make clear, my 'case' resists '[a]ppeals to empirical verification, to the reduction of complexity into simplicity, to the principles of repeatability and objectivity…[that] have gradually edged out all others' (Galloway 2014:108).

Therefore, I wasn't able to 'reverse engineer'— the 'process of articulating the specifications of a system through a rigorous examination drawing on domain knowledge, observation, and deduction to unearth a model of how that system works' (Diakopoulos 2013:13) — the black box. Furthermore, reverse engineering usually only provides fuzzy glimpses of how an algorithm works in practice but not its actual constitution (ibid). Although reverse engineers can 'give some indication of the factors and conditions embedded into an algorithm, they generally cannot do so with any specificity' (Kitchin 2017:24). Google has its own obfuscation methods that cloaked its processes of filtering during my 'truth games' and the process of data gathering demanded much research into the changing technology, as it was a technical learning curve for me. Because my 'findings' represent a small-scale study and are perhaps not representative of other users, my small data set is without external validity.

3.0 The Cybernetic Hypothesis

To conclude this 'method chapter', I draw on *The Cybernetic Hypothesis* by Alexander Galloway to address the methodologies and tools of research in the digital era. His hypothesis 'refers to a specific epistemological regime in which systems or networks combine both human and nonhuman agents in mutual communication and command' (2014:111). Echoing the research of Franklin and her 'commands' of prescriptive technologies, *The Cybernetic Hypothesis* by the French collective Tiqqun (2001) spelled out this new type of 'social management involving both human and non-human assets' that determines human activity through a computational and media-oriented society (ibid). Galloway claims that this 'digital

universe' involving the 'provisional nature of cybernetics' has shifted the 'foundations of knowledge and culture' with the introduction of digital tools (ibid). The challenge of digital humanities then is that 'we live within a cybernetic universe without necessarily being conscious of it and we use these digital tools without necessarily reflecting on them' (ibid).

In 2007 the notion of 'digital methods' came to the fore 'as a counterpoint to virtual methods', e.g. 'traditional' research methods such as online surveys or online ethnography, and introduced 'the social scientific instrumentarium to digital research' (Rogers 2009 cited by Venturini et al. 2018:4200).[88] Expressed otherwise, 'digital methods' can be defined as the repurposing of the inscriptions generated by digital media for the study of collective phenomena (Rogers 2014). They are however not as straightforward as they seem. If methods are the 'path', or way, derived from the Greek 'hodos', with which we actually learn about human-computer interaction with search engines, then the results return to their Greek origin as data—'a fact given as the basis for calculation in mathematical problems' (Galloway 2014:108). Although digital methods can 'capture' the computational aspect of online platforms, often researchers 'confuse the phenomena that they investigate from the features of the media in which they manifest' (ibid).

> [M]ethod—the learned and applied activities of measuring appearances—has become the world itself, though through refinement its ubiquity is often concealed. Something is not something until it appears on a radar, is caught in a net, is classified as a member, is staked or buried in a plot, is stored and recalled in a search (Beyes et al. 2019:504).

The 'circular' etymology of the word 'search' is represented by the cybernetic system of control and feedback loops, where collected data reflects how 'human and non-human agents are connected in networks' (Galloway 2014:112). According to Galloway, 'digital methods are at best a benign part of the zeitgeist and at worst a promulgation of late twentieth-century computationalism' (ibid). By promoting computer centric research methods, it is assumed that research follows the 'trend of normalization' instead of being opposed to it in regard to 'intellectual endeavours that value deviation over normalization, heterodoxy over orthodoxy' (ibid).[89] Concomitantly these technological devices 'fragment and reorganise social life around specific economic mandates' with the very forms of critique (deviation and heterodoxy) that challenge hegemonic infrastructures, part and parcel of the 'knowledge production' machine (ibid:108). One example is that of artistic critique, as Simon Sheikh explains in the following:

> The knowledge economy as it has been called, is comprised of its own critique, or in the words of Luc Boltanski and Eve Chiapello, the 'artistic critique' of capital has been integrated into capital itself, and virtuosity, creativity, performativity and so on, are all the basis of this production, and knowledge itself a type of commodity' (Sheikh 2009:3).

Drawing on the title of Boltanski's and Chiapello seminal book, Galloway points out that the 'new spirit of capitalism is found in brainwork, self-measurement and self-fashioning, perpetual critique and innovation, data creation and extraction' (2014:110).

---

[88] Digital Methods are computational 'but written for the web and digital culture' and they 'work with natively digital data', whereas Digital Humanities 'often use standard computational methods' and 'work with digitized material using standard computational methods' (Rogers 2014).

[89] Galloway emphasises that it is 'obligatory' to identify the shifting forms of value—either recapitulative or contestational claims—and whether or not the humanities must mirror these societal trends or propose an 'asymmetrical rethinking of those larger trends' with an 'autopositional posture' (2014:112).

Digital research is rife with enquiries into the use of tools to capture and analyse large corpora of data and it is the 'deskilling problem' or the 'minimalization of agency' that lies in the ubiquity of digital tools, with what Galloway characterizes as becoming 'literate in a digital device rather than a literary corpus' (ibid). 'Tooling' means becoming the master of buttons and apps, thereby reinforcing the power constructs instead of contesting the processes and outcomes of how cultural objects are analysed *and* interpreted. Rather than valorising these '[h]ighly coded interfaces [that] reduce the spectrum of possible input to a few keywords of algorithmic parameters', Galloway warns that there needs to be a searchlight directed onto the 'dark side of a dissipated human agency'. His concern is that these 'low-agency scholars are deskilled scholars, proletarianized thinkers denuded of their authority to make claims (at least claims that haven't been culled directly from a measurement device)' (ibid). In this sense a deskilled approach and the fetishisation of tools is juxtaposed against that of human agency in regard to research methodology.

With the increase of habituation to digital technologies there is a decrease in criticality, is there then simultaneously a rise in the power of ideology? (ibid) Galloway's slogan, 'critique is foe to ideology' is a call for criticality and, specifically in regard to my research, the hegemony of Google, where this hidden infrastructure for information discovery is 'emboldened' and becomes naturalised (ibid). It then becomes a question of hegemony regarding 'claims' made about 'knowledge or reality' that are either 'recapitulative' or 'critical' of these hierarchies that exude domination and subordination (ibid). Rather than be maligned with Google Search ideology, I aim to articulate the components that erode not only the edifices of education with seductive tools, such as APIs and SCIs, but human agency itself.

Or, if I transcribe Galloway's argument to search engines, 'can we still use our tools now that the master [Google] has taken them up?'(ibid). The 'modern regime of critical thought' defined above, which has enabled 'criticality', along with my methods assemblage could provide a means to question the 'naturalization of technology' (ibid:127).[90] More importantly, Law's 'heterogeneous engineering' concerns the (re)configuring of these arrangements or 'methods assemblage' in 'technological discourse and practices' (Suchman 2012:49), such as human and non-human actors. Returning to the impossibility of an ever-changing habitat, Lucy Suchman emphasises the boundaries between human-machine interfaces and whether the performative is hidden, along with the configuration of agency. It is perhaps then what Françoise Laurelle deems the 'weak force' —a certain kind of agency in 'generic humanity' that not only provides access to commonalities of history and society but must rise to the challenge of researching with non-human actants (Galloway 2014).

Thus my approach to 'digital methods' is as someone without a background in computing or any definable online skillset and the claims that I make in this thesis are humble ones, based on my methods assemblage that reflects the still, 'quiet methods, slow methods, or modest methods without imperialism' (Law 2004:104-105). Although unassuming, this 'methods assemblage' attempts to 'oppose the intransparency, unrepresentability and incommensurability of algorithmic ordering with a different "understanding" of digital media' (Beyes and Pias 2019:102). In spite of my limited 'tech savvy', I hope that my methodological framework is a small contribution to knowledge directed towards a high agency scholarship instead of a low

[90] 'Method, then, unavoidably produces not only truths and non-truths, realities and non-realities, presences and absences, but also arrangements with political implications' (Law 2004:143).

agency, deskilled scholarship. Either way, *The Cybernetic Hypothesis* (Tiqqun 2001; Galloway 2014) concerning human/machine interaction presents continual challenges to research and, in particular, to (Post)Digital Cultures. (Figure 31)

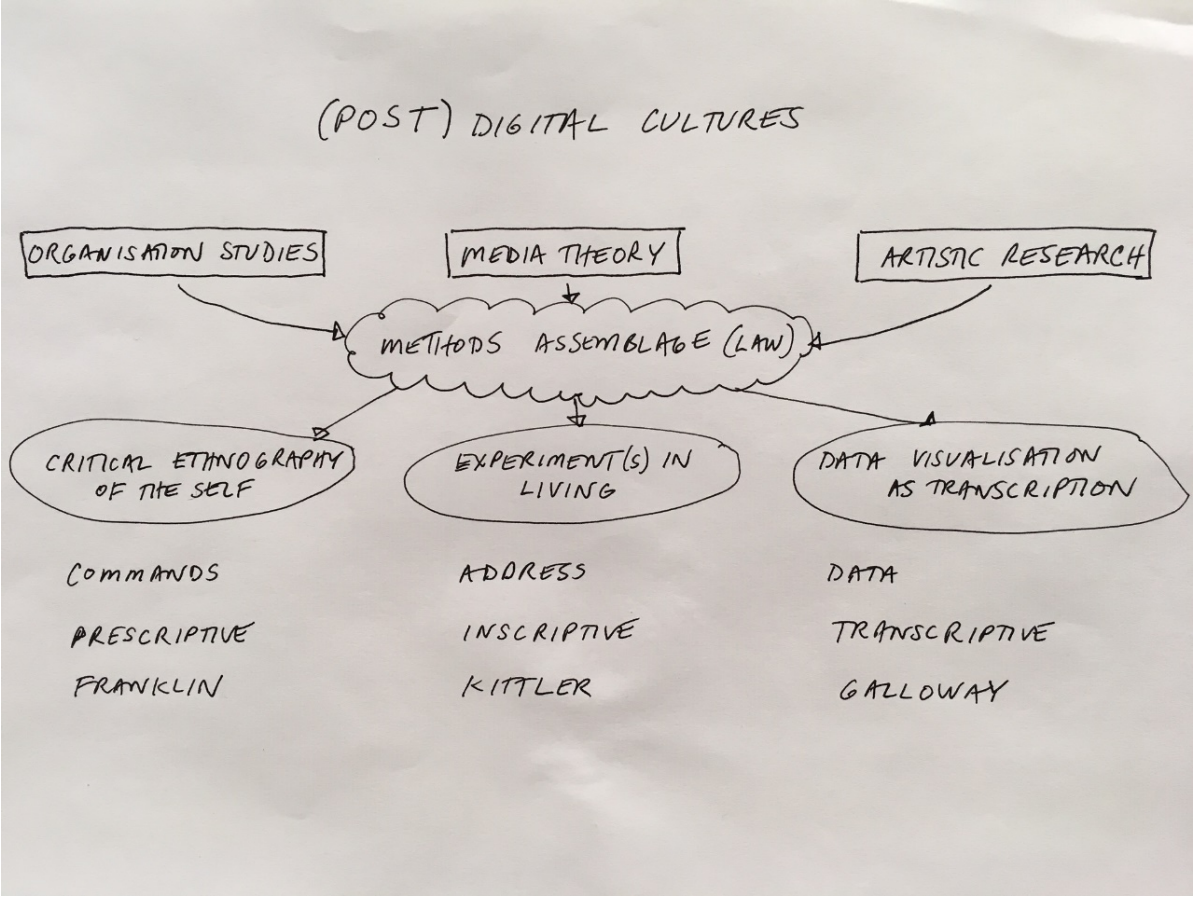


Figure 31: Methodological Framework (Post)Digital Cultures

Information is useless if it cannot be found and it is not a coincidence that a search engine like Google has turned into one of the most significant companies of the new century. These engines are never just practical tools to deal with information overload. Such cognitive technologies embed political philosophy in seemingly neutral code (Becker and Stalder 2009).

Chapter 5:

The Personalised Subject

1.0 Search as Habit

It all begins with words typed into a search box.[91]

As explained in Chapter 2, search engines are navigational [new]media (Van Couvering 2009). In her book *Updating to remain the same: Habitual New Media,* media theorist Wendy Hui Kyong Chun elaborates upon habit through the contradictions and paradoxes of new media: 'its enduring ephemerality, its visible invisibility, its exposing empowerment, its networked individuation, and its obsolescent ubiquity' (2016:43). *A (new)media archaeology of citation* (Chapter 3) demonstrated how Google Search organises information and makes it accessible, with PageRank commanding traffic and user attention through its authority as well as (back)linking structure. In the section 'Always Searching, Never Finding' Chun spells out that '[i]magined connections are habits: they are projected links based on frequent repetition'. In this way, the 'imagined connections and edges' are 'traces of habits' with 'the strength of a friendship—its weight, gauged by how often specific actions occur, such as clicking on links, where 'information *is* habit' (ibid, emphasis mine). Having recourse to Alexander Galloway and Ellen Ullman, in Chapter 4 I articulated that information is data given form, whilst Chun contextualises habit through information as communication and conditioning, how it *forms* and *informs*:

> Habit resonates with two seemingly unrelated meanings of information: one, the archaic definition of information as the formation—the training—of individuals (to inform was to form); and two, Shannon and Weaver's definition of information as communication that lies beneath meaning (ibid).

On the communication highways of today's internet, users apply search engines and chose 'keywords' to help them navigate. Historically, keywords have become the method for tagging and assigning terms to articles, books and projects in order to enable their 'searchability' (Appendix A). After 2000 on all websites and in the web browser of search engines, a 'long-sided-white-rectangle-longer-than-the-length-of-an-average-word' (Halavais 2009:09) appeared that is now commonplace. Determining which keyword to type in the browser's search box, whether it is one word, or two or a phrase, is one of the most common daily activities. Nowadays 'search' has become a habit (Chun 2016).

In this chapter I continue to investigate *How does Google search work?* drawing on results from my methods and structured through three lenses: Advertisement, Authenticity and Authorship.

2.0 Advertisement

In this section I explain how 'Googleconomics' plays a crucial role in the ranking of search results with advertisements. As an intermediary of online marketplaces, this type of 'platform

[91] https://moz.com/beginners-guide-to-seo/keyword-research

capitalism' (Srnicek 2016) enables Google to prioritise its own services (books.google) as it 'intervenes' (Gillespie 2014).

2.1 Googlenomics

As discussed in Chapter 3 with reference to Rieder (2012), if social relations no longer hold authority because they now rest with the measuring instrument, the analysis of not only the technical, but financial logic of PageRank in regard to advertising, is crucial. In 2000, Google introduced its own business model, *AdWords*, where advertisers created targeted advertisements that emerged on search engine result pages (SERPs), based on the auction of keywords. Ads appeared on the right-hand side of the webpage with the innovation that 'all the sidebar slots on the results page were sold off in a single auction, (compare that to an early pioneer of auction-driven search ads, Overture, which held a separate auction for each slot)' (Levy 2009). Not just large companies but all types of enterprises can take part in the auction, with the bids for keywords submitted by advertisers beforehand, who only pay if users clicked on their ads. However, the differential between bids was adjusted in this online auction:

> Googlers decided that the winner of each auction would pay the amount (plus a penny) of the bid from the advertiser with the next-highest offer. (If Joe bids $10, Alice bids $9, and Sue bids $6, Joe gets the top slot and pays $9.01. Alice gets the next slot for $6.01, and so on.) Since competitors didn't have to worry about costly overbidding errors, the paradoxical result was that it encouraged higher bids (Levy 2009).

Moreover, the 'complete calculus' of Google's bidding auction process was obscured to the parties involved behind a frontend interface and 'remain[ed] invisible to the searchers themselves' (Haeselin 2017:7), as the algorithms calculated in real time.

Steve Levy's WIRED article describes how Google's Chief Economist Hal Varian realised that this was no longer traditional advertising and instead compared the enterprising innovation of these 'second-price auctions' to online dating (ibid).[92] Varian references a paper by Herman B. Leonard, 'Elicitation of Honest Preferences for the Assignment of Individuals to Positions', that describes an 'assignment problem' (two-sided matching market) in auctioneering where '[eli]citation' and 'matching' are solved mathematically with a two-step algorithm 'for conducting the allocation procedure and deriving the appropriate prices' (Leonard 1983:478). Obtained through 'incentive compatibility', Leonard draws on the analogy of auctioning off a Rembrandt painting and the 'second-price auction' described first by Vickrey (1961 cited by ibid). 'The second price auction is conducted by soliciting sealed bids for the painting and "selling" it to the highest bidder at the price offered by the second highest bidder' (ibid:473). His conclusion was that this 'competitive market allocation process' could be applied to sales of all commodities.[93]

The other important factor of *AdWords,* besides second-price auction, is the 'quality score' that is keyword based and derived from past performance data. It has three determining aspects: 'Ad

[92] Additionally, a 'Dutch auction' and the 'physics of clicks' (Susan Wojcicki) are important factors and the Keyword Pricing Index (Diane Tang) is a 'reality check'. For a detailed explanation please see Levy's article.
[93] However, 'the procedure described here has the distinct disadvantage that it relies for 'market discipline' on an external medium of exchange. In a society in which one such medium, money, is increasingly unpopular as a basis for allocation for some commodities and positions, this may be a serious shortcoming' (Leonard 1983:478).

relevance'—the relationship between the keyword and the ad, the 'Landing page' experience—the quality of the page the ad is linked to and 'Expected click through rate'— the percentage of the times users click on ads. Reflected by the habitual actions of users clicking on ads, the value assigned to each keyword in *Adwords* can be broken down even further into microtransactions of data. Hal Varian promulgated this type of valuation as 'Googlenomics', based on the theory of Ken Arrow (1987) who describes 'Atomic Economics' in the following:

> Googlenomics is the economics of items at its finest granularity. Pre-internet, ads were optimized for an entire audience, for an example everybody would see the same front page ad on a newspage, but not on the internet. Ads can now be optimized for each ad impression. So it is Atomic (2016).

The CTR (Click-Through-Rate) is the most important feature when determining the quality score and emotionally charged ads receive even higher CTRs. *AdWords* then became the combination of results 'related to what people were looking for at that very moment' by logging the 'clicks that registered interest by users' as a means of 'miraculous effectiveness' (Levy 2009:223 cited by Haeselin 2017:7). In June 2003 Google launched *AdSense,* which syndicated CPC (cost per click) ads to partners automatically. AdSense is seemingly a subtler means to advertise; banner ads (Google display advertising) are placed with Javascript code on non-Google websites relative to the content of the page.[94]

> Such a technique provides great potential for advertisers, as users spend most of their time on the web on content pages, as opposed to search engine result pages. The advertiser pays a certain amount for every click on the ad, and the revenue is shared between the publisher and the ad-network (Esteve 2017:40).

Contextual matching increased and as of February 2010, AdSense started using search history to offer more relevant ads.

Already 'large scale', in March 2003 Google had over 100,000 advertisers and, in 2005, Google's syndicated advertising comprised 44% of its advertising revenue (Google 2006a cited by Van Couvering 2009:116). 'According to Google's website, by 2006 it had "the largest online advertising network available, reaching over 80% of 30-day US internet users"' (ibid). In April 2007, a year after its acquisition of YouTube for 1,65 billion dollars, Google purchased the web advertising company DoubleClick for 3,1 billion dollars in cash. 'The acquisition also gave Google access to DoubleClick's user metrics and allowed it to track users on any site on which Google advertising appears' (Fuchs 2011; Kang and McAllister 2011 cited by Hillis et al. 2013:17). Additionally, the takeover expanded Google's relationship beyond algorithm-drive ad auctions, incorporating DoubleClick's network of web publishers, advertising agencies and advertisement software, which I will return to in Chapter 8. In 2010 the EU launched an investigation regarding the merger of Google's online advertising and DoubleClick's display ad serving technology (*AdSense*) yet Google was eventually cleared of 'anti-trust' practices (Davilla 2017:377), however further antitrust lawsuits continue at the moment of writing.

[94] When the user visits the page the code uses JSON to display content from Google servers and the advertiser pays CPM (cost-per-thousand impressions) for website advertisements. 'The Google *AdSense* program differs in that it delivers Google *AdWords* ads to individuals' websites. Google then pays web publishers for the ads displayed on their site based on user clicks on ads or on ad impressions, depending on the type of ad' (Google 2016).

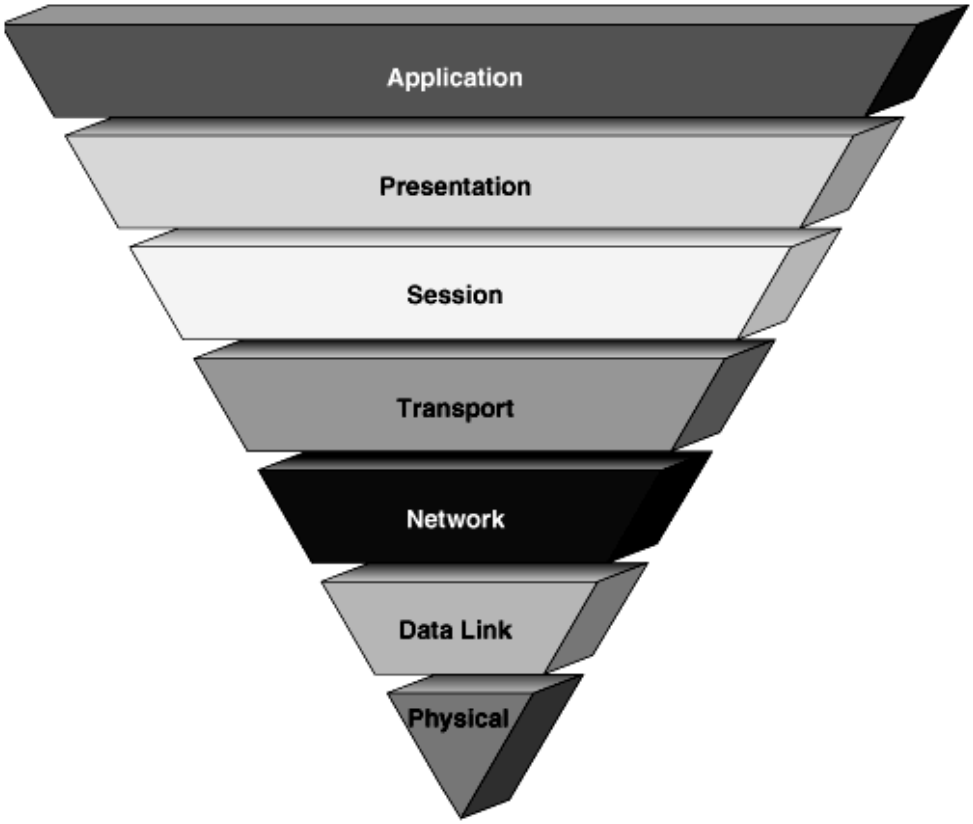


Figure 32: OSI model. Image credit: Critical Engineering Course at Weise 7, Berlin

2.2 Platform Capitalism

Analogous to the offices in Chapter 1, which put people in touch with each other to exchange goods, objects and services through the register and advertisements in journals, Google Search solves the distance and coordination problem between parties by connecting users with what they are seeking through a platform interface. Nowadays located as the top layer on the Open System Interconnection (OCI) diagram, an API (Application Programming Interface) enables the user access to the search engine though its workings (criteria, second-price auction, data collection) that determine results whilst remaining invisible. (Figure 32) Supported by networking a participatory culture of users who contribute and share resources between each other, this type of mediation involves interaction with multi-sided markets. The intermediary of these online marketplaces—the platform—'brings together at least two distinct groups of end-users' (Rieder and Sire 2013:5). In this way the Google search platform reflects the consequences of the fragmentation of an original monolithic market into multiple differentiated markets and users and, according to Rieder and Sire, it is a three-sided market.

> On one side, Internet users query the engine to find information, entertainment, and so on. On a second side, Google indexes 'content providers' that want users to reach their websites. On the third side, advertisers are trying to attract visitors *beyond* the traffic received from 'organic' results. Google subsidizes two of the three sides and charges the third: Internet users search the Web for free; content providers are charged neither for getting indexed nor for the traffic they receive from organic results; advertisers, however, pay for every click and thereby finance the platform (ibid).

As elucidated in Chapter 3, not only users but other parties advertise with Google Search because they dominate the market, having the greatest flow of traffic and because other people use them more people join in, as shown by 'network effects'.

> Integral to framing 'marketisation' and, in particular, to structuring 'market encounters' in digital space (Calişkan and Callon 2010:14-16), platform intermediation is distinctive because it attempts both to make the 'connections' of multi-sided markets and to coordinate the network effects of 'connectivity' (Van Dijck, 2013 cited by Langley and Leyshon 2016:3).

As a socio-technical intermediary, coordinated network effects invite the co-creation of value by users as they search online, for information, shopping or knowledge acquisition. Since the second decade of this millenium, 'Platform Capitalism' has been used as a term to describe how companies such as Google connect users with markets to 'achieve a quasi-monopoly position' with their 'competitive advantage from network effects, aggregating large numbers of participants' (Srnicek 2016:31).[95] Concurrently usurping the middleperson, the platform is 'a newly predominant type of business model premised upon bringing different groups together' (ibid:32)—customers, advertisers, service providers, producers, suppliers and even physical objects are structured by advertising.

> The increase of digital platforms, including search engines such as Google, are firms, or enterprises that regulate the flow of capital and obtain profit through an advertising business model, which extract information on users, undertake a labour of analysis, and then use the products of that process to sell ad space (ibid:21).

Serving up ads that influence the user's experience by creating detours in their path to information, the platform acts as a digital infrastructure (Srnicek 2016). 'Platforms are particular comings together of code and commerce: when infrastructures of participation and connectivity are designed and data is realised and acted upon, this is the intermediation of digital economic circulation in action' (Langley and Leyshon 2016:9).

As Plantin et al. point out, Google is now an example of 'this combined infrastructuralization of platforms and the platformization of infrastructures' (2017:301). Yet here I wish to emphasise that ingrained in the 'social and historical production and organization processes' (Noble 2018:148), advertising dictates commercial search that not only influences 'link culture' and 'click culture' but determines the ranking of search results. Instead of just interpreting them as 'results', users should be 'seeing them as the result of strategic actors selecting and assembling user content into a particular composite', what Gillespie enounces as 'platforms intervene, and the public culture that emerges from them is, in important ways, the outcome' (2015:2).

2.3 Ranking:Ads

With Google's 'platform capitalism', the 'trusted user's' habit of search queries is mediated by an exchange: users obtain 'free' search results in exchange for their data. Facilitated by data collection, the expression of users, through their search queries, is tied to an economic logic and the platform incorporates user interaction with search results that rank and recommend. Whilst

[95] Originally the term 'Platform Capitalism' was coined by Sacha Lobo in a Der Spiegel article from 2014.

carrying out my ‘critical ethnography of the self’ as a ‘trusted user’, I realised I was receiving not only organic search results but also advertisements in my SERPs (Search Engine Result Pages) for both Google and Tor. Google separates the so-called ‘organic’ or ‘not-paid-for-results’ from ‘paid ads’, which are defined as

> the commercial portion of message content for which an advertiser has or will pay when a searcher sees their content after submitting a query in a search engine or Web site search box, which will typically take a searcher to another Web page’ (J. Jansen 2011:232 cited by Lewandowski 2017:7).

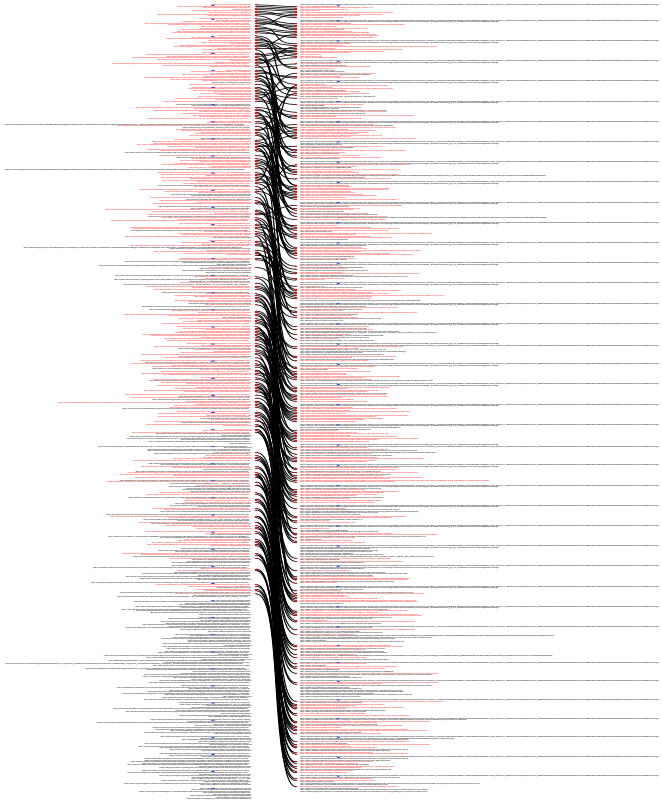


Figure 33: Comparison of Google Search ‘personalized’ and Tor Browser ‘anonymized’ search results with keyword ‘postinternet’. The visualisation shows full URLs, where the longest ones are Google Books (left) and searchdisconnect ads with Tor (right).

Displayed together on the same page, search-based ads are similar to organic search results in their appearance because they both have a title and a short description and use the same colours for the heading but most notably because they both display an URL (ibid). The size and scale of the results also matters, with users looking at displays that have slick graphics or images that capture users’ attention more often, what has been called ‘screen real estate’ (Nicholson et al. 2006 cited by ibid:9).

> Users tend to focus (and click) on results that appear above the fold. It is essential to bear in mind whether a search engine result appears above the so-called “fold,” i.e., in the area of the search engine results pages that is immediately visible without the need for scrolling (ibid:8).

However, I did look beyond the fold, where I noticed that Google also returns ads which are very long, which is shown in my visualisation including the full URL. (Figure 33). I received links for books.google.dk that reflects the promotion of their own service of scanning and uploading books, which they do not own yet users can read sections of these books for 'free'. (Figure 34) This echoes EU lawsuits and fines of recent years in regard to Google's 'anti-trust' behaviour whereby it prioritises its own search services above competitors, which I will discuss in Chapter 9. Additionally, I postulate that the reason these ads are the longest URLs of all my results, as shown in the data visualisations, is because they not only give the exact page of where the keyword appears but contain all of the tracking data that has been collated by Google about the user (me) via my IP address.[96] (Figure 35)

Figure 34: Keyword: **anthropocene**. Google results. Pages 22 and 23

Figure 35: Keyword: **postinternet**. Google results. Pages 6-11.

[96] As shown by the URL, all kinds of data is being collected on me.

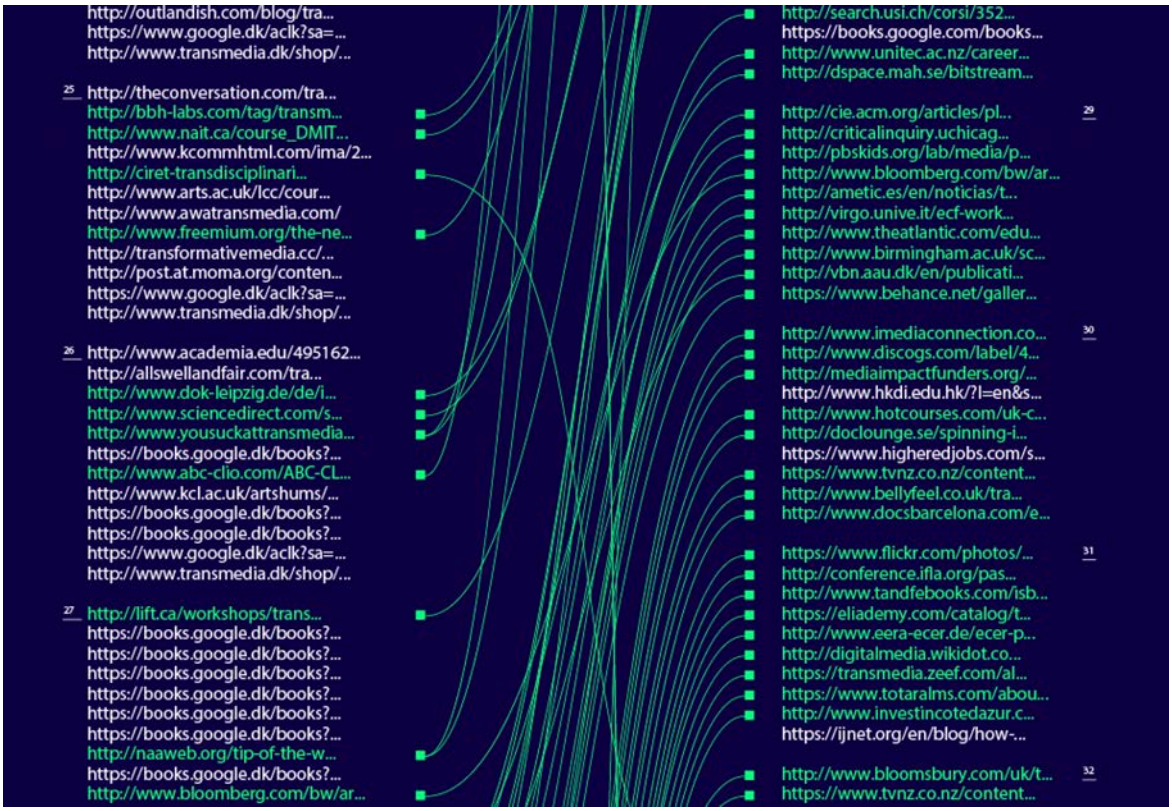


Figure 36: Keyword: **transmedia**. Google results. Pages 26 and 27

Moreover, books.google.com appear with the keyword **transmedia** on page 27 (Figure 36). As I continued to look beyond the fold my data revealed a difference in ranking there as well. Google and Tor often returned SERPs from Wikipedia as a first answer yet the total ranking results were never identical. With the Tor Browser, the keyword **postinternet** delivers (sakshippost.com), a link for an Indian newspaper discussing the fact that 620 million surf the internet via a mobile phone and which was ranked on the 37th page, whilst the same link was on the 28th page of Google Search. In a further example I received a hyperlink expressing concern about the 'global public internet splintering into a series of bordered cyberspace segments, according to a white paper on 'internet fragmentation published by the World Economic Forum (WEF)' from the Pakistan Post, which appeared on the 28th page of Google Search whilst on the 38th page with the Tor Browser.[97] (Figure 37)

Figure 37: Keyword **postinternet**. Google results page 28.

[97] http://www.pakistanpost.pk/2016/01/restrictions-haunt-internet-operation.html

Another discrepancy (25th on Google and 30th on Tor) was the result of a page from the Occupy Wall Street forum discussing Mark Zuckerberg's (Facebook) proposal for his Internet.org, exemplifying 'balkanisation' through Facebook's attempt to provide 'free internet' to less wealthy users in India in exchange for their data. (Figure 38) These results reflect the 'zeitgeist' of January 2016 when I collected this specific data.

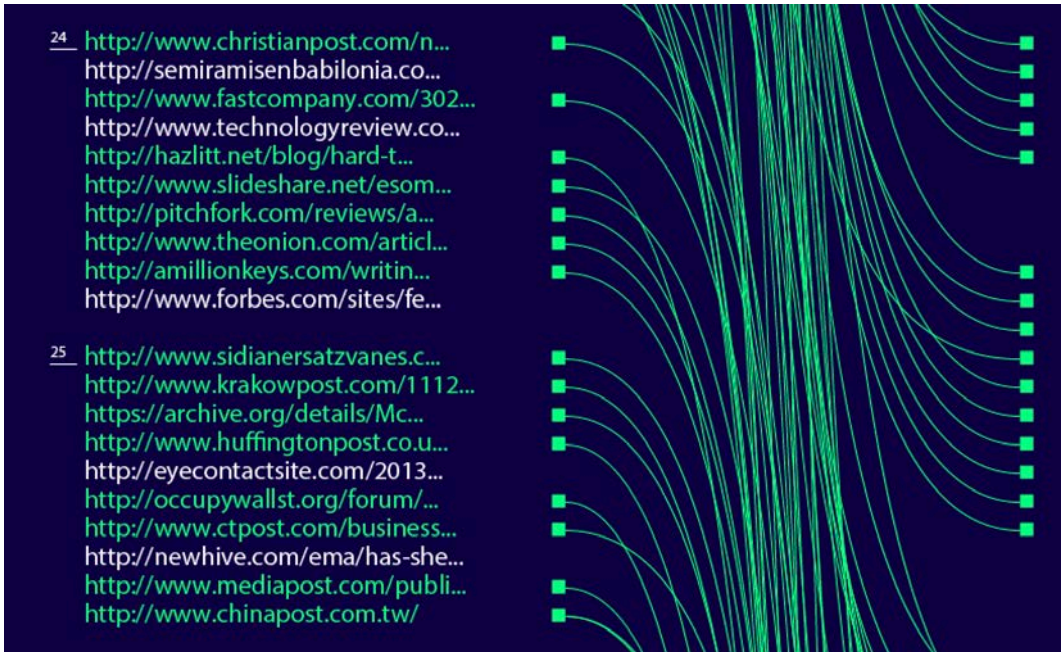


Figure 38: Keyword **postinternet.** Google results page 24-25.

3.0 Authenticity

In the forthcoming section I show how targeting searching subjects with identity markers such as IP addresses and the 'trusted user's' search history facilitates 'personalisation' in an era of Semantic Capitalism. I also question the authenticity of this by postulating how users are categorised into groups of others 'like them' through collaborative filtering and their 'YOUs value' (Chun 2016), where 'you' is always both singular and plural.

3.1 Ranking: Personalisation

Analogous to the 'pre-history of search engines' described in Chapter 1, where the 'office' collected the addresses of citizens, in contemporary digital society, the IP (Internet Protocol) address for communication is a unique identifier or numeric label assigned to any type of device that is connected to a network. [98] The IP address is thereby comparable to the physical address of a neighbourhood, such as the name of the country, city, street, house number and floor/ apartment. The IP address usually consists of groups of numbers (4), with a network part and a host part, where one can find the ISP, the country, the region and state, the city and the location (longitude and latitude). It is read from right to left, gradually increasing precision as it reaches the final numbers to pinpoint the location of the device (and thereby the user) but it does not reveal a person's name, the exact address, phone number or email address. The IP is the third or

[98] As shown in Chapter 1 with the history of the postal services and later the telegraph and telephone cables, the present-day internet builds upon these infrastructures, through private and public means.

'network' layer that determines the path of the data packet and address, transmitting data through channels, which are occupied during transmission and can be used for other traffic afterwards.[99] (Figure 32)

To return to my Google search habit, Chun offers the reader a key insight about the 'rehabituation' of individuals through the changing nature of the IP address:

> When the Internet was first conceived, IP addresses, even when fixed, were not viewed as permanently tethered to a computer, let alone a user. With the advent of changes to IP addressing, and more importantly the emergence of cross-platform logins, cookies, and other means of tracking through 'unique identifiers,' it has become easier to tie users to their actions. This traceability has entailed the massive rehabituation of individuals into authenticated users through the expansion (2016:57).

When users type in keywords, Google Search connects multiple parties through data trails comprised of 'signals' or markers that identify users, such as the IP address.[100] The IP address is the key signal that affected my results from my 'experiment in living'. Acting as a tracking device (as well as GPS) and serving as a location marker, the IP address determines the device and identity of the user online. The IP address also facilitates data transfer, with search engines creating a database on each user containing their respective queries and search histories, which enables the construction of 'the personalised subject'.

At the end of Chapter 3, Brin and Page first mention personalisation in their seminal text with their 'trusted user' who would 'provide feedback and improve the quality of the search experience' (1998). I explained how the more people used Google search, the larger its proprietary 'Database of Intentions' became and that this 'information represents the real-time history of post-Web culture', as 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 2005:7).

One of these 'ends' is 'personalisation', which Google publicly announced on December 4, 2009. Capturing the searching subject's IP address as well as maintaining a log of previous queries, personalisation adapts them into real-time search results, even if one is not signed into a Google account. With personalisation, anticipatory searches are based on previous search histories and users remain within what Eli Pariser describes as the 'filter bubble' —where they receive positions, opinions and news that they already know and support (2011). Pariser's critique is that this 21st century zeitgeist creates a sense of deprivation, leading to the 'distortion effect', one of the challenges posed by personalised filters.

---

[99] In 1974 Vint Cerf and Bob Kahn applied another layer on top, the TCP (Transmission Control Protocol), which enables the sharing of transmission resources that use packet switching (connectionless datagrams) among the nodes and is connection-oriented. It is the DNS (Domain Name System) protocol, located in part of the Application layer (7), that translates the (human readable) domain names into a numerical network-specific address (IP) that enables it to map it to the Data Link layer (2) with the help of the ARP (Address Resolution Protocol), a communication protocol.

[100] 'The standard search engines (Google, Bing, Yahoo, etc.) track and record everything you search for. Your typical search engine records the following information any time you use it: Your IP address, User agent, Unique identifier (stored in browser cookies) and Search terms' (Taylor 2018).

> Like a lens, the filter bubble invisibly transforms the world we experience by controlling what we see and don't see. It interferes with the interplay between our mental processes and our external environment (ibid:82-83).

In *The Filter Bubble* Pariser also shows how different people, on different computers, in diverse locations in the world, receive different search results using Google Search (2011). This is echoed by Halavais who elucidates that personalisation is 'an ongoing response to users that focuses on their individual needs' (2009:51), which include not only their search queries and history but context.[101] I received locative data, a known signal based on country and language that triggers a 'response' to Google's personalisation algorithms. My results for *Re:search - Terms of Art* show that although certain URLs were the same in the dataset (Appendix E), the ranking was different. Therefore, I propose that the main difference between Google Search and Tor (Disconnect search) is that Tor obscures the IP address without pinpointing location. Disconnect states that it does 'detect non-personally identifiable geo-location information to optimize our services, but [unlike Google] we definitely don't collect your precise geo-location or associate geo-location information with a particular user' (Disconnect Search 2016).

Figure 39: Keyword **postmedia**. Google results page 2.

With the keywords containing the prefix 'post' ('**postdigital**', '**postmedia**', '**postinternet**' and '**posthumanism**'), I received Google results reflecting my search histories and location thereby showing user personalisation. Google Search results for '**postinternet**' were often relegated to the Netherlands, where I live most of the time even though I was searching and collecting the data in Denmark.[102] Besides '**postinternet**' returning Danish postal results and information regarding governmental e-post and internet that shows locative data was being collected, I obtained a few results for mail delivery with the keyword '**postmedia**', as well as URLs referring to Danish media outlets. Besides geolocation, I postulate that the 'semantic' interpretation of the keyword determined the divergent ranking of my results. My 'keywords' were not 'trending' in the commercial sense or terms of financial speculation, nor would they be included in the high bidding wars of AdWords, as I explain in Appendix E. Instead, because I used terms from contemporary art beginning with the word 'post', the algorithmic interpretation

[101] In exchange for their data, users receive free search and ostensibly 'tailored' advertising, turning themselves into commodities for advertisers if they don't delete 'cookies' or installing adblocking plug-ins that would inhibit it. This personalisation is then a currency, with data correlated through algorithmic technologies and acquired by marketers, or third parties (Ridgway 2015).
[102] Postinternet reflects the way that the internet as a tool can produce art, but is not specifically situated in it (as Net Art would be) and its influence of the internet on society and culture, both online and offline.

of the word 'post' varies (postal services, 'post' as a common naming convention for newspapers, as a widely used contraction of 'posting', as in blogging or commentary, and as a term for histories and intellectual movements). Appendix F provides more examples. (Figure 39)

As shown above and described in the previous section, language and its interpretation become important criteria along with the monetisation of certain keywords that reflects users' thoughts as queries, giving these keywords the power of a certain 'semantic governmentality'.

> Not only is it the case that every word in every language now has its price that fluctuates according to the laws of the market, but additionally, both search results and the corresponding advertisements shown are now optimised according to their potential market value based on pre-emptively calculated individual 'user relevancy' (Feuz et al. 2011).

With personalisation, Google claims that every search and individual user is unique, therefore it seeks to deliver customised results for each person that satisfies their interest. As demonstrated above, if personalisation does exist is very difficult to capture because it is hard to carry out an experiment and measure the results, as my 'experiment in living' attempted to do. Going beyond these claims, one particular study was able to show the effects and degrees of personalisation on the one hand and on the other, 'deindividualised' results.

3.2 Semantic Capitalism: Diagnosing the mechanisms of personalisation

> We don't want to know everything about you. What we want to do is to try to help to connect you with the peoples, ideas, and things you are looking for. You decide which information you decide to give to us. It is a utility that improves if you decide to share information. (Google spokesperson 2010)

Google states above that the more users share information, the higher the relevance of the search results they will obtain. In order to test this statement, Martin Feuz, Matthew Fuller and Felix Stalder designed the empirical study, *Personal Web Searching in the age of Semantic Capitalism: Diagnosing the mechanisms of personalisation*.[103] The study began with the premise that not all users are looking for the same information when they type in a keyword and therefore the quality of search results is decreasing. To combat this problem Google had been working on ways of obtaining better search results for the user— personalisation. In the study, Feuz, Fuller and Stalder first assigned identities for Immanuel Kant, Friedrich Nietzsche and Michel Foucault, who represented the 18th, 19th and 20th centuries respectively, creating one Gmail email account per user, then selected their likely 'keywords' as a vocabulary and subsequently programmed thousands of search requests from the same server in London.

Key to this empirical enquiry (and in determining personalisation) is that they tested three ways of profiling: 'the knowledge person', 'the social person' and 'the embodied person', as it was assumed Google does something similar in order to produce personalised search results. The first looked at what people are interested in, based on search histories. The second looked at networks and who the person is connected to using email, social networks and communication

[103] Published on the *First Monday* blog in February 2011, the research was conducted with great difficulty in the preceding years because Google interfered with the testing by blocking IP addresses and adding personalisation whilst it was being carried out.

technology. The last looked at the environs of where the person is located and their bodily state in context. By merging the three profiles, relevant search results would be delivered for each individual user, with algorithms interpreting the user's behaviour and deciding what is relevant for the user.

> Together, these three profiles promise to provide a detailed, comprehensive and up-to-date context for each search query, with the potential to deliver precise results, that reflect not just the information "out-there", but also the unique interest a user has at any given moment (ibid).

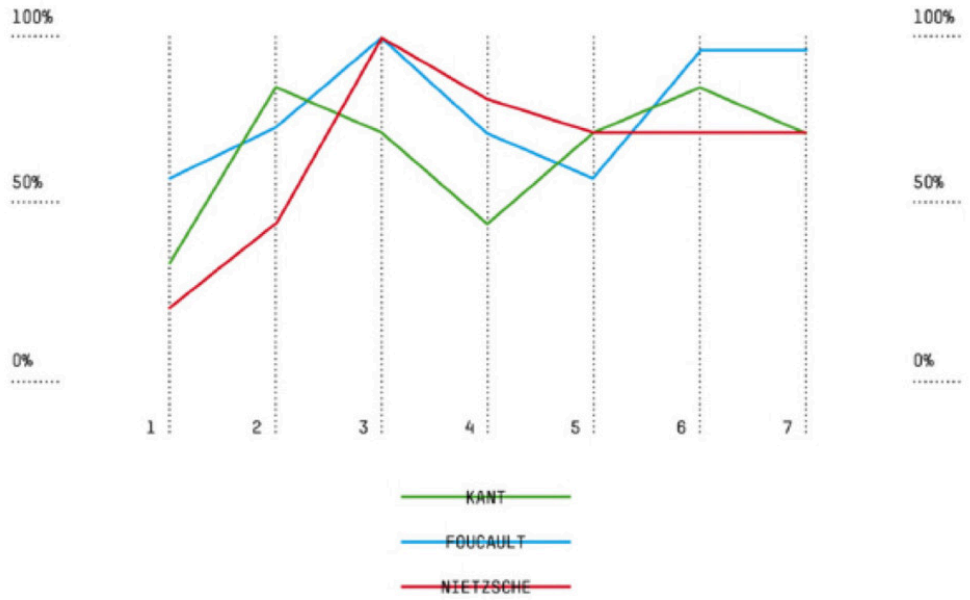


Figure 40: Feuz et al.'s Fig. 3. Infograph Hypothesis 3 (2011).

Based on the fact that each philosopher received personalised results for some queries, 'even if there was no relationship between the search history and the test query', *they refute their third hypothesis, Personalisation reflects only an individual user's past search and Web interests* (ibid). Displayed by the upward trend of each session (1-7) of queries, their data visualisation brings to light that Google applies personalisation to search queries outside the user's domain of recorded search and web history' (ibid). (Figure 40) They deduced that 'Google does not only rely on a user's personal semantic history, but that it extrapolates from what it knows about a person to his or her association with statistical group profiles that Google has built up over time' (ibid). Instead of divulging what might be of interest with an entered keyword, the searcher is presented with a 'preselected image of the world based on what kind of group the search engine associates us with' and that the 'result of such group patterning in the background, unseen and undetectable to the user affected by it, could be an *inversion of the promise of personalisation'* (ibid emphasis mine). One of their conclusions enounces the 'subtle homogenisation' that leads to an 'adherence to a preselected world', which 'rather than increasing diversity', […] 'becomes a self–fulfilling prophecy' (ibid).

Their findings also suggest that Google's personalised search, 'does not fully provide the much-touted benefits for its search users. More likely, it seems to serve the interest of advertisers in providing more relevant audiences to them' (ibid). Thus Google has sold the 'audience

commodity', to return to Smythe's terminology in Chapter 3. Yet Google does not stream the same ad to its billions of users and users do not type in the same query. Instead it collects user data, providing search results to users for free and connecting users to targeted advertising.

> In order to produce this context, vast amounts of personal information need to be collected, organised and made actionable. Within the fast receding limitations of storage space and computing power, profiles can never be too comprehensive, too detailed, or too up-to-date (ibid).

Moreover, behavioural targeting schemes apply analogous technology by collating user data to define 'audience segments', dependent on other users with similar profiles. 'A query is now evaluated in the context of a user's search history and other data compiled into a personal profile and associated with statistical groups' (ibid). Based on buying habits, search histories and so on, the user is first classified and assigned according to demographics, yet not as an individual—rather with mass customisation. Therefore, this study produced the first evidence that:

> Google is actively matching people to groups, which are produced statistically, thus giving people not only the results they want (based on what Google knows about them for a fact), but also generating results that Google thinks might be relevant for users (or advertisers) thus more or less subtly pushing users to see the world according to criteria pre-defined by Google (ibid).

3.3 Collaborative Filtering

Referencing Chapter 3 and Chun's treatise on habits described above, users from 'a diverse array of countries train Google's algorithms', receiving as a response to their search queries 'localized results' that are 'relevant' (Stone 2018:138). Because Google delivers results that 'paired the history of the users' searches with her up-to-the-moment interests' (Haeselin 2017:7), users assume they receive personalised links and information. However, their data is used to shape a personality profile or to present them with criteria that is not necessarily their own, which is somewhat paradoxical. 'We are presented with the picture of the world made up of what someone else, based on proprietary knowledge, determines to be suitable to one's individual subjectivity' (Stadler and Mayer 2009:99). Expressed differently, personalisation is being advertised (and subsequently sold) to the general public on the basis that people will obtain customised results that are directly connected to them, yet there is a strategic organisation—profiling—within Google databases, which I will address in Chapter 8.

As Google gathered user data over recent decades, it became increasingly powerful by learning users' preferences, matching individual users to groups, based on statistical data (Feuz et al. 2011). This *Semantic Capitalism* created a 'second index'—an opaque second layer drawing on the searching subject's reaction to submitted keywords (Stadler and Mayer 2009:99).

> Sites hope to anticipate the user at the moment the algorithm is called on, which requires knowledge of that user gleaned at that instant, knowledge of that user already gathered, and knowledge of users estimated to be statistically and demographically like them (Beer 2009)—drawing together what Stalder and Mayer (2009) call the second index (Gillespie 2014:173).

It is the orchestration of these 'socio-technical arrangements', which informs the logics of informational and affective organisation that comprises the 'second index' a.k.a. personalisation, in this way becoming 'device cultures' (Rogers et al. 2013:161 cited in Weltevrede et al. 2014). These device cultures interact with 'how users are imagined and prescribed into the interface through notions of affordances and grammars' (Agre 1994 cited by Weltevrede 2016:14) such as Google's personalisation, which is fed back to group-assigned users as search results, or 'neighbourhood recommendations'.

> When considering Google's algorithms, a user can only enter input in the form of a search query and receive output in the form of result listings, but what remains hidden is how the recommendation engine works (ibid:105).

Building upon the previous study's conclusion, where Google's personalisation assigns users to pre-defined criteria, I propose that categories are constructed not through individuation but by 'collaborative filtering'—a technique used in recommendation systems, which sorts through and analyses data for patterns. Wendy Chun's chapter in *Habitual New Media*, 'Searching, Never Finding', elucidates how collaborative filtering algorithms move from analysis to prediction by building 'neighbourhoods', based on strong similarities (and differences) between the reactions of users to movies (2016:103).[104] 'User to user collaborative filtering' is the term for a group of other users whose likes and dislikes are similar to that of a user, called 'the neighbourhood'. Automatic 'predictions' or 'filtering' is needed in order to define the range of similarity values of the neighbourhood yet 'user to user' prediction is complicated, as there are many characteristics and it is often difficult to find the 'nearest neighbour'.

These recommendation systems work with a 'rating vector', which attempts to quantify the intuition that users with similar tastes will have higher similarity than users with dissimilar tastes, shown by using a 'centred cosine similarity' (also known as the Pearson Correlation) 'that measures the "mean" and the distance, thereby determining which users are similar with each other' (Leskovec et al. 2016). Following media critic Douglas Rushkoff, it is not about the individual user but the demographic category to which they are assigned:

> Recommendation engines measure what people like me would do and telling me what that is, so I can then find out what people like me do, so I can become much more like a person like me. By telling me what people like me do, and encouraging me to be more like a person like me, they help me to become more typically one of my kind of person. And the more like one of my kind of person I become, the less me I am, and the more I am a demographic type (2014).

3.4 YOUs VALUE

As described in Chapter 1, the human crawlers of the bureau d'adresse are analogous to the spiders of search engines, drawing on personal contacts for information and acting as recommendation systems that filtered answers back to those querying. They also found matches and put like in touch with like. Chun draws on the research of Bourdieu to accentuate the reinforcement of former 'offline' groups: 'if the practices of the members of the same group or class are more and better harmonized than the agents know or wish, it is because, as Leibniz puts it, following only [his or her] own laws, each nonetheless agrees with the other' (1977;

[104] However, very few users take the time to actually rate the films.

2016:14). In this way Pierre Bourdieu's work on 'habitus' is both 'linked to class formation' as a 'generative principle of regulated improvisations' in which members of the same class purport harmony, as well as to the unconscious (ibid:7). In the 19th century, Gustave Le Bon's work on how contagion forms the 'psychological crowd' that creates the 'collective unconsciousness' (1895) and Gabriel Tarde's research on the crowd understood through imitation (1898), contended that the social bond is greater than individual subjectivity (Borch 2016), which has been the premise for many studies on mass psychology.[105]

'Whereas, for Pierre Bourdieu, the habit is the unconscious, or productive unconscious, for Gabriel Tarde, modern man is a somnambulist who is linked to others through habit' (Chun 2016:7). Is it not unimaginable then that the contemporary online 'habitus' is exemplified through the personalised subject sleepwalking in the dark by using Google Search. Thus not only habits have been altered but habitus, as shown by 'collective unconsciousness' of crowd contagion a.k.a. mimetic behaviour. In the late 20th century, the anthropological philosopher René Girard articulated in his book, *Things Hidden since the Foundation of the World* that 'imitation is at the root of all behaviour' and this 'mimetic desire' is not autonomous, as it is based on other people's desires and in this manner propagates and spreads. In other words, users desire what other users desire and in turn copy that desire.[106]

As demonstrated in the section on 'Platform Capitalism', network effects are a positive externality where not only those with similar interests come together, but also because everyone else is doing it. 'Mimetic desire' for some, 'network effect' for others—users' habits have contributed to Google's hegemony and monopoly on search and the perception that they are receiving 'relevant' and 'customised' search results specially manufactured and crafted for them. Over the past 15 years Google has promoted the idea that it delivers individualised search results based on a unique user, yet it is not 'authenticity' that is actuated.

With seemingly a more complete profile composed of the 'personalised subject', the searcher is actually presented with a 'pre-selected image of the world' based on associations with others who have similar interests, categorised into demographic groups as demonstrated above (Feuz et al. 2011). This 'subtle homogenisation' reflects an *inversion of the promise of personalisation* (ibid) dominated by what Chun deems 'YOUS value' in regard to the collation and correlation of user data (2016:64).

> That is, if our world is data rich, it is not simply because we provide content for free, but also because every interaction is made to leave a trace, which is then incorporated with other traces and used to understand you, where you is always both singular and plural. Whether any particular YOU is aware of it or not, as YOU we constitute a latent resource (ibid:251).

---

[105] In the article *Market Sociality: Mirowski, Shiller and the Tension between Mimetic and Anti-mimetic Market Features*, Christian Borch and Ann-Christine Lange address Nobel Prize winner Robert J. Shiller's claim that 'mass psychology may well be the prominent cause of movements in the price of the aggregate stock market' (2017). According to Schiller's study, group pressure (though there were only two), influences people to fads and fashions of 'social movements'. 'Put differently, people are essentially mimetically constituted, and this characteristic, argues Shiller, also applies to financial markets, in which investors could mimic the behaviours and assessments of others' (Borch 2016).

[106] A great follower of René Girard, Silicon Valley entrepreneur Peter Thiel in 2014 postulated that Facebook is based on mimetic desire [which is why he is part owner?]

As I have proposed in the sections on personalisation, collaborative filtering and mimetic desire, users keep producing multiple selves as 'YOUs value' (Chun 2016), reminiscent of Nietzsche's 'phantom of the ego' in regard to the power of the mimetic faculty in the formation of subjectivity (Lawtoo 2013).[107]

With the habit of search and clicking on links like other like-minded searching subjects, it is not only the authority of PageRank (Chapter 3) and how 'platforms intervene' (Gillespie 2015), which is evoked. Collaborative filtering and Chun's 'YOUs value' challenge the authentic subjectivity of search — personalisation. If personalised subjects are not individualised but collaboratively filtered and assigned to groups, would these subjects not start to behave similarly to each other, ostensibly because they receive the same 'nudges' from the algorithms? Thus, I postulate that the desires of personalised subjects are actually the desires of others as there is no self that is not originally identified with the other, as the research of Le Bon (1895), Tarde (1898), Girard (1961), Lawtoo (2013), Rushkoff (2014), Chun (2016), Borch (2016), Borch and Lange (2017) has shown.

4.0 Authorship

This section addresses my 'unique results' through the lens of authorship in regard to IP, intellectual property rights and the (in)visibility management of PageRank cum RankBrain's black-boxed secrets through 'algorithmic capitalism' (Bilic 2017).

4.1 Signals: Unique Results

The Black Box discussion has gained momentum in recent years but the decision-making aspects of algorithms remain obscured, whether for assigning loans based on credit, recommendations for just about any product or service or, more specifically, search results. With my 'methods assemblage' (Law 2004) I attempted to investigate Google's 'black boxed secrets' regarding the PageRank algorithm. As my 'data visualisations as transcription' demonstrate, the same URLs are represented by green whilst the white URLs exhibit 'unique results' (Appendix E). These appear in a specific configuration—with 'personalisation' (Google) or with 'anonymisation' (Tor) and I set out to discover why there are unique results for each keyword in each browser. My humble 'experiment in living' that attempted to decipher what is going on inside the 'black box' is however, not new. (Appendix G) As a reaction to PageRank results, during the past two decades an entire industry, Search Engine Optimisation (SEO) has developed around attempting to reverse engineer Google's proprietary algorithms. Often referred to as the grey art of 'fixing' the ranking, 'some companies' entire business consists of helping others improve their position in Google search results' (Plantin et al. 2018:305). Although

> Google (to its credit) goes to considerable lengths to prevent the gaming of its search algorithms—for instance, Google asserts that no one can pay to increase a web page's rank, except via "sponsored links" clearly labeled as such—many strategies exist for raising that rank' (ibid).

[107] Perhaps Le Bon's 'crowd contagion' could also be applied to the present predicament of online filter bubbles, though instead of physical interaction between people, computational processes within 'datascapes' determine like copying like.

In order to ‘game’ PageRank, content creators employ tricks and methods, ‘like large amounts of text made invisible to everyone but the search engine (by making it the same color as the background, for example)’ (Halavais 2008:71). Sometimes called ‘stuffing’, keywords trending at that time would be inserted in order to receive attention, affecting the content as spam, called ‘spamdexing’ in SEO jargon. PageRank counterfeiting occurs through this type of ‘link spam,’ where sites only exist for hits and are void of content. ‘Google bowling involves sabotaging competitors’ sites by making them appear to be engaging in obvious link farming and other practices used by search spammers’ (Pedone 2005 cited by ibid:75). ‘Scrapersites’ or ‘AdSense farms’ gather information automatically in order to be highly ranked on SERPS and these ‘techniques are generally invisible to the user, but prey on the indexing process of the search engine to subvert its function’ (Nathanson 1998 cited by ibid:71). These ‘symbionts’ or ‘parasites’ are considered ‘click fraud’ and penalised by Google when discovered, as often competitors continually click on rival sites to drive up the cost (ibid:80). ‘If Google catches them they can knock them down in rankings, manually, going to the seventh page or so. This is called going to Google Jail’ (Evans and Schmalensee 2016:145).

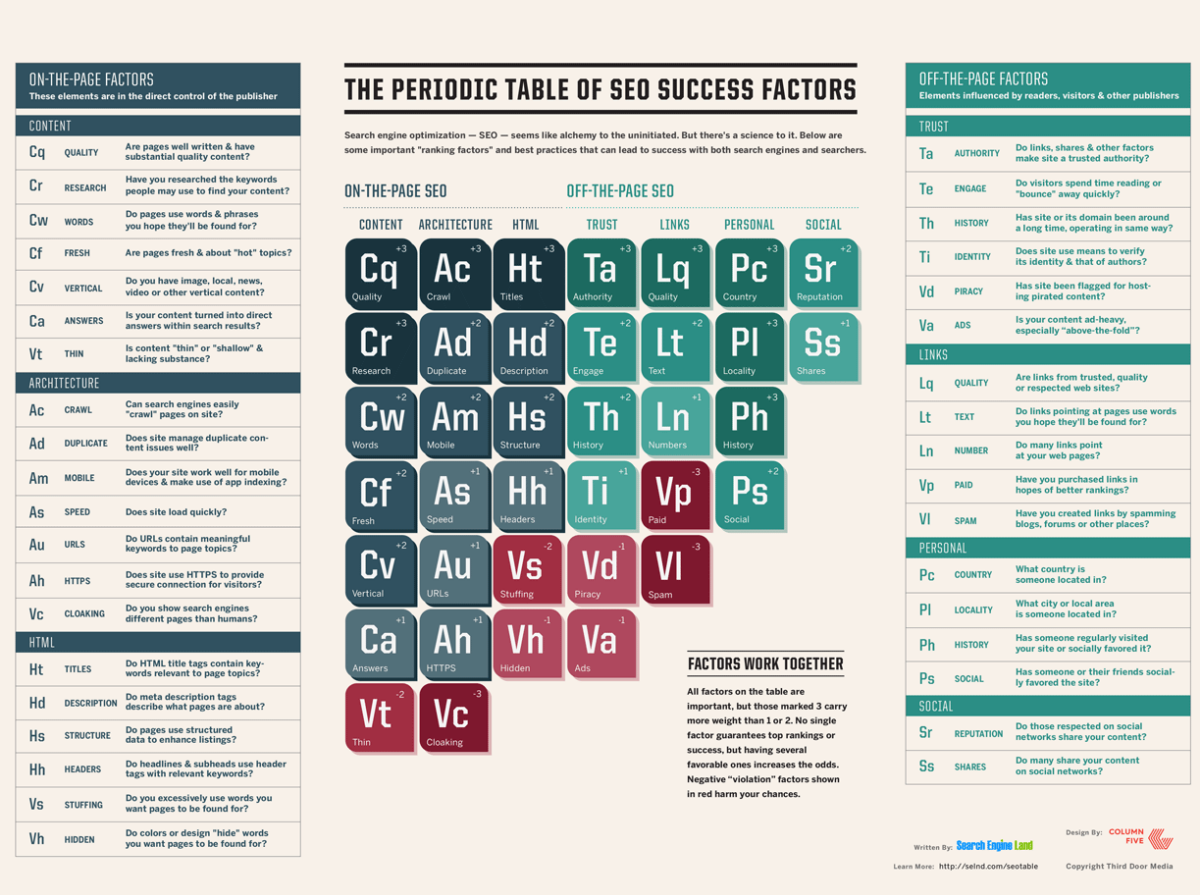


Figure 41: Replaced elements from the Periodic Table with SEO factors, online and offline, which determine ranking, offering insight into the cryptic ranking criteria. Visualisation: SearchEngineLand (2016).

It is often because of these practices that the reputation of the SEO industry has been associated with spamming, as many tactics are not ethical and instead practice deception. However, SEO ‘games’ the PageRank algorithm by ‘reverse engineering’ processes in order to ‘identify their own set of signals that seem to affect search engines directly’ (Fishkin and Pollard 2007 cited by Halavais 2009:83). Signals are typically factors that are tied to content, anything regarding text

such as the words on a page, headings and font weights, which determine the ranking of an individual page. (Figure 41) As explained in Chapter 3, ranking signals can also be structural, which 'are more concerned about the linked structure of pages' (Hugill et al. 2013:246) —the links pointing at a page, outbound link quality and whether a page is on a secure server and so on (Sullivan 2010). Besides links and content, keyword density, words in bold, duplicate content and domain registration duration are some other examples of factors, or 'clues'. Signals also need to be 'repeated' so that they can be 'transmitted'—[s]ignals that are not repeated or repeatable 'die' (Chun 2016:52).

Another criteria measuring the participation by users can 'contribute to ranking algorithms such as the clickstream' that 'could be seen as direct user-relevance feedback' (Hugill et al. 2013:246). In this way signals can also be tied to a user, such as where a searcher is located or their search and browsing history' (Sullivan 2016), as discussed above, with recommendations based on collaborative filtering.

> [A]lthough certain details are known, others, such as how specific criteria are measured, weighed against each other, and which criteria override one another remain obfuscated (Weltevrede 2016:105).

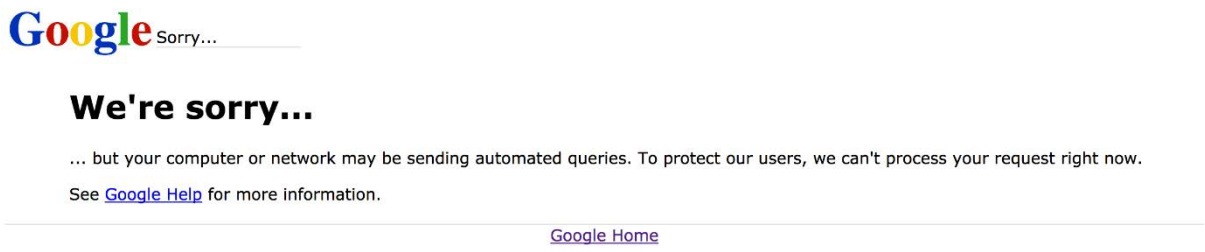


Figure 42: 'Google sorry', November 2015.

Unlike the SEO industry that adjusts websites in order to have them appear higher in search results for clients, 'black hat hackers' reroute and redirect search results so as to game Google and make money, which I mentioned in the Prologue. With my method 'experiment in living' I entered keywords into the search interface and saved every web page to gather the data—a repetitive 'digital labour' process analogous to 'mechanical turking'.[108] My data collecting

[108] Mechanical Turk is a crowdsourced marketplace where people can be subcontracted virtually to carry out usually menial tasks, although they are often labour intensive, for very little remuneration. Please see Irani and

process was also interrupted. Drawing on the research of Seaver (2013), Rob Kitchin explains that 'many proprietary systems are aware that many people are seeking to determine and game their algorithm, and thus seek to identify and block bot users' (2017:24). This is exactly what happened to me when I was carrying out my search queries—Google perceived that I was a bot because I was only clicking on and collecting data from the 1st, 10th, 20th, 30th, 40th, etc. of SERPs. At a certain moment I received a Google 'We're sorry' message. (Figure 40) Although Google 'intervened' and I had to start over, I was eventually able to collect my small data set.

4.2 From PageRank to RankBrain

> We are building the ultimate search engine, which would understand everything. It would understand exactly what you wanted and would give you the right thing back. It's an artificial intelligence that can answer any question. If you look at your programming, your DNA, your whole operating system, it's about 600 megabytes compressed. So it's smaller than any operating system. Smaller than Linux, Windows, or anything like that. Your program algorithms aren't that complicated. It's probably more about the overall computation. (Larry Page, February 24, 2007 cited by Degoutin and Wagon 2015)

The quote above reflects the ambitions of Google in 2007 and their intentions to model their search engine on the human brain, when PageRank dominated SEO's understanding of ranking. 'PageRank is still in use today, but it is now a part of a much larger system' (Google 2008) and based on Google's public divulgences, Hillis et al. state that 'the suite of algorithms that constitute PageRank remains the core of Google's technological advantage within the field of search' (2013:18). Not only was it known that 'Google's famous PageRank algorithm measures the relevance of webpages by valuing hyperlinks', as explained in Chapter 3, but increasingly more information has come to light about adding 'other signals in their algorithm' (Weltevrede et al. 2014:14).[109] A timeline visualisation provides a comprehensive overview of minor and major changes from 2002-2015, though it 'is by no means exhaustive' because Google changes its algorithm 500-600 times a year (Weltevrede 2016:120). (Figure 43)

> Google made more than 400 changes to PageRank in 2010 alone, and the entire apparatus has achieved such non-deterministic and stochastic complexity that it is no longer possible to know exactly how any given change affects the algorithmic matrix as a whole (Martinez 2011 cited by Hillis et al. 2013:18).

Therefore, it is been increasingly harder to 'game' Google's ranking system and the signals that determine them, which remain the 'unknowns' of the black box. 'It is virtually impossible, moreover, to reverse engineer search algorithms because both they and the Web are constantly changing entities' (ibid).

In the course of 20 years, SEO blogs have been constantly publishing what is known about PageRank signals and updates. In August 2009 the 'Caffeine' update was a change to the indexing architecture and this web ecosystem facilitated the searching of content immediately after it is crawled, providing a 50% fresher index. The overhaul also incorporated social networking into search and increased speed, hence its name. 'Panda' was a search filter update in 2011 that prevented sites with 'poor content' and downranked sites, which were considered

---

Silberman's *Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk* paper (2013).

[109] Esther Weltevrede compiled the known changes of PageRank between 2010-2016.

lower quality, enabling higher quality pages to rise. In April 2012 Google launched the 'Penguin' update that attempted to catch sites that were 'spamming', e.g. buying or obtaining links through networks and boosting Google rankings and as of September 30, 2016, updates in real time are part of the core algorithm. Since June 2013 'Payday' fights spam such as 'payday loans', 'Pigeon' improves local results, 'Top Heavy' demotes ad-heavy pages, 'Mobile Friendly' rewards mobile users and 'Pirate' fights copyright infringement.

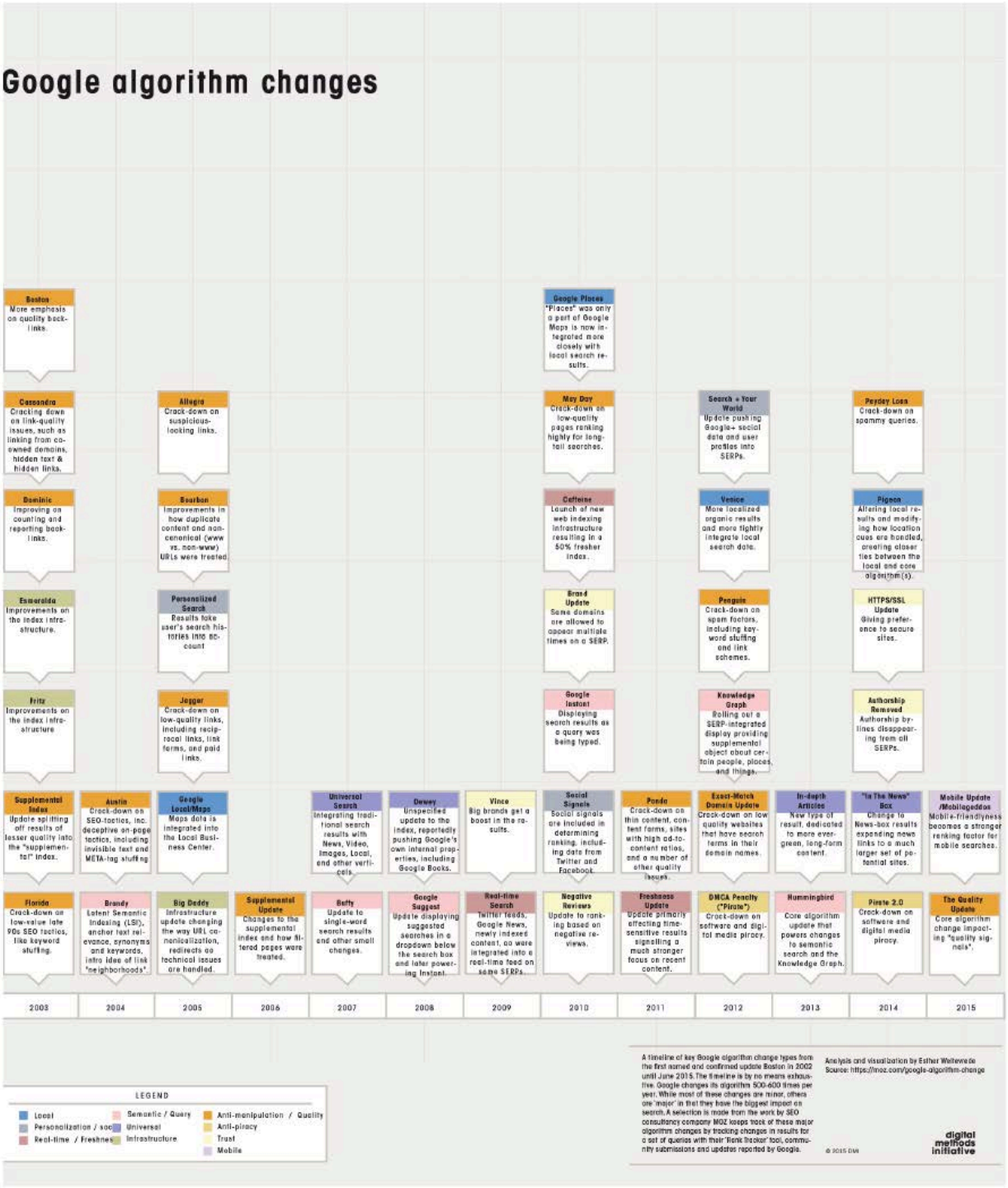


Figure 43: Weltevrede's timeline of key Google algorithm changes from 2002-2015 (2016)

Analogous to the components of a car engine that has had it parts replaced, where Penguin and Panda might be the oil filter and gas pump respectively, the launch of 'Hummingbird' in August 2013 was Google's largest overhaul since 2001. With the introduction of a brand-new search engine the emphasis shifted to the contextual—it became less now about the keyword and more about the intention behind it—the semantic capabilities. Whereas previously certain keywords were the focus, now the other words in the sentence and their meaning gained importance. Moreover, the complexity level of the queries went up, resulting in an improvement in indexing web documents. According to David Amerland, author of *Google Semantic Search,* the 'relationality linking search queries and web documents' comes together with the *Knowledge Graph*, along with 'conversational search' that incorporated voice activated enquiries (Gesenhues 2013). (Figure 44).

Figure 44: Amy Gesenhues's Twitter feed from September 27, 2013.

Publicly declared as Google's third most important ranking 'signal', after links and content (words), PageRank infers the use of a keyword by applying synonyms or stemming lists. User's queries have also changed and are not only keywords but also multi-words, phrases and sentences that could be deemed 'long-tail' queries. To a certain degree these need to be translated from 'ambiguous to specific' or from 'uncommon to common,' in order to be processed and analysed (Sullivan 2016). Hummingbird has a more sophisticated understanding of search queries, even long-tail searches, by applying synonyms and semantics. As mentioned in Appendix D, my small dataset (*Re:search -Terms of Art*) reflects divergent results in ranking when searching with two browsers, Google and Tor (Disconnect Search), with Google's results based on ads and locative data, as shown above. Additionally, my keywords were neither trending nor popular—most are eclectic terms and 'long-tail' queries. I postulate that when I entered two words or phrases in the search box which are not usually phrased together, the search engine separates the words (**artistic** + **research** or **new** + **aesthetic**, for example), it was less likely there would be a match in ranking between the two search browsers, resulting in greater variability between the two results and therefore delivering more unique results.[110] (Figure 45)

[110] I propose this based on my experimentation with other search engines such as YaCy that works with two stages of ranking.

If Hummingbird is the new Google engine from 2013, the latest replacement part is then 'RankBrain' a 'machine-learning artificial intelligence system'. Launched around early 2015, RankBrain ostensibly 'interprets' what people are searching for, even though they may have not entered the exact keywords. Around October 2015, RankBrain was handling 'a very large fraction' of billions of daily searches that it had never seen before (Metz 2016). Thus, acknowledging that my 'experiment in living' was conducted from October 2014 to January 2016, RankBrain might have been used to answer some of my search queries. I entered new queries to Google Search and at that time, RankBrain was answering around 15% of new searches. I propose that the reason I obtained 'unique results', where certain results only appear with Google, might have been due not only to ads and personalisation but the fact that Google was already applying machine learning algorithms when I was carrying out my study.

Figure 45: Keyword: **artistic research.** Unique results (white links)

As of June 2016, RankBrain is being implemented for every Google Search request because it is a 'query interpretation algorithm' optimised for 'meaning/parsing', which enables it to understand the meaning and intent in a specific context and to determine 'correct retrieval of information from the index' (Fiorelli 2016).

> That is the real reason why Semantics, in the sense of structured data, good architecture and topical research, hubs and closeness are so important IMHO, as well as being directly or potentially relevant for the personal search history of the searchers' (ibid).

Furthermore, the SEO industry speculates that RankBrain is summarising the page's content. The murmur is that the algorithm is adapting, or 'learning' from people's mistakes and its

surroundings. It does this by applying 'deep neural networks' that are modelled after the human brain. By combining hardware and software in an attempt to copy the human web of neurons, development is carried out through trial and error, analysing the results and then adjusting the math, then repeating the steps with new data.

Previously computers were not fast enough, or the data sets were too small, to carry out this type of testing. Now there is enough computational power at Google's data centres to handle much more data and it enables the pace of the research to quicken. RankBrain is continuously fed vast amounts of data to train the deep-learning neural networks, splitting computing tasks across machines.[111] In this way the algorithms are 'trained' and they 'learn' but it is 'difficult to directly tweak a machine learning-based system to boost the importance of certain signals over others' (Lau 2017). In the past, humans—programmers—wrote the code and then tweaked the results, now with RankBrain the models are machine-readable and therefore less human-readable. At the moment, no one is quite sure why neural nets behave the way they do.

> Neural networks are really just math—linear algebra—and engineers can certainly trace how the numbers behave inside these multi-layered creations. The trouble is that it's hard to understand why a neural net classifies a photo or spoken word or snippet of natural language in a certain way (Metz 2016).

These machine-readable models are less human-readable and it is extremely difficult to determine why priority is given to certain results (higher ranking) over other 'unique results'. Nowadays with its 'learning process' of deep neural networks replacing written rules and code, 'one of the benefits of Google is the ability to scale' (Giannandrea 2017) and measure user interaction.

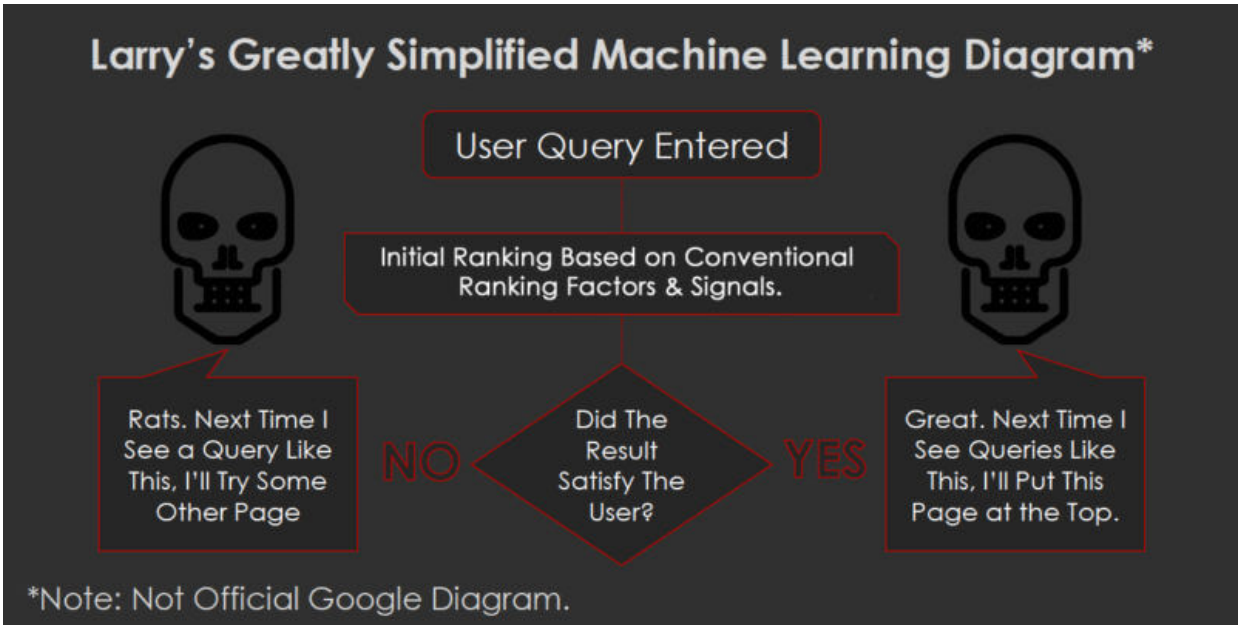


Figure 46: Search Engine Land's mockup by Larry Kim (2017)

[111] This progress in technology facilitates a constellation or coming together of different capabilities from various sources, through models and parameters. According to Google the algorithm first learns offline, being fed historical batched searches (or photos or spoken commands) from which it makes predictions. Eventually the subject, or learner, in this case the algorithm, is able to make predictions through the constant repetition of this cycle. If the predictions are correct, the latest versions of RankBrain go live (Sullivan 2016).

As of 2017 what seems to matter with the ranking is engagement, with high ranking now based on user interaction, or 'traffic', clicking on ads and creating 'network surplus value', as elucidated in Chapter 3. Users have been constantly clicking on the links but now RankBrain is placing greater importance on these user signals, as shown in the SEO mock-up diagram below (Figure 46). RankBrain now ostensibly deranks sites that may have good content if the user doesn't click on the results (where before the signals were measuring keywords, relative to content). Moreover, RankBrain is combined with the amount of time a user spends on the page, or 'dwell time' and only Google can measure this. Once again, clicking is the measurement that determines the value of the web pages returned, constantly reflecting the cycles of user engagement. Traffic, another important factor, diminishes over time if there is no user interaction and

> [m]achine learning then becomes a "layer" on top of this. It becomes the final arbiter of rank—quality control, if you will (Kim 2017).

In 2016 Google admitted that 'ranking systems are made up of not one, but a whole series of algorithms'. With constant tweaking to its proprietary algorithm, in 2017 there were more than 2400 changes and in 2018 more than 3200 changes (Grind et al. 2019), there are now reportedly 'more than the 200 signals that Google uses to rank results' (Weltevrede 2016:117; Sullivan 2010). Over the past twenty years PageRank cum RankBrain has been mythologised, fetishised and commodified because of the undisclosed 'signals' that determine ranking, yet its code still remains a corporate secret (Pasquale 2015).

4.3 Algorithmic Capitalism

According to the sociologist George Simmel, secrecy is a social formation. Secrecy in this sense—i.e., secrecy which is effective through negative or positive means of concealment—is one of the 'greatest accomplishments of humanity' because certain elements of life cannot happen publically (Simmel 1999:406 cited in Beyes and Pias 2019:88). In their article *The Media Arcane*, Beyes and Pias accentuate the concept of secrecy as an object of digital cultural analysis through its genealogy: 'to write the history of secrecy is thus to trace the development of society: a sequence of revealed things that have become secret and of secret things that have been revealed' (ibid). The role of secrecy also operates within state security operations and by controlling the flow of information through technological affordances, with Brighenti observing that 'secrecy is technically and technologically managed, and increasingly so' (2010:66 cited by Flyverbom 2016:104). Maintaining its ever-increasing intellectual property rights (IPR) through patents (Pasquale 2015:98), corporate secrets such as Google's proprietary algorithms make search technology incredibly valuable and 'these algorithms are guarded, like missile codes' (Noble 2018b).

As described in Chapter 3 and above, developers defined the rules of the algorithm—comprised of code written by a programmer deemed 'machine readable'—which in turn was processed by computers. Eventually, the abstraction of languages and the types of code depended on its purpose and the machines that could execute predefined steps. Code is then a conversion system, where language is codified and users and machines alike need to learn the symbols in

order to communicate.[112] In *Speaking Code*, the software studies theorist Geoff Cox argues that the importance of code, along with speech, should not be underestimated in its centrality to economics and culture as both an aesthetic and political expression. Programme code 'mirrors the instability inherent in the relationship of speech to language', with its performative aspects rendering it 'interpretable in the context of its distribution and network of operations' (Cox 2012).[113] In this way code is understood not only as script but performance and, in this sense, resembles spoken language in that it is always ready for action (ibid).

This performativity in the execution of a code is the calculation of a function for algorithmic procedures that produce a value, which reflects not only the relation between their inputs and outputs but facilitates the reification of the code.

> The contemporary dominating rationality in technical artefacts such as software, code, algorithms, devices and gadgets is presented as a rational universal whose production process and commodity nature are reified behind smooth technical designs (Bilic 2017:6).

It is this 'technological rationality' that brings code and calculation to the forefront of contemporary capitalist production, promoting efficiency and competition that prioritises commodity exchange and, along with this, new business models. 'On the one hand, it [technics] is a product of human society and social conditions. On the other hand, its objectified existence exerts a specific form of influence over behaviour and consciousness of humans' (ibid:7).

This type of 'algorithmic capitalism' (Bilic 2017) then alters the framework of social relations, incorporating humans and things with 'technics'; concomitantly it has a favourable role in commodity exchange by producing high profits for those who own it along with control and domination. As Friedrich Kittler proclaimed in his seminal text, *There is no software*, 'copyright claims for algorithms' have already occurred, '[p]recisely because software does not exist as a machine-independent faculty, software as a commercial or American medium insists all the more' (1993). He further goes on to explain that 'under these tragic conditions', the German criminal law defined software as a 'material thing' instead of upholding 'software as a mental property' (ibid). In turn, the IP (Intellectual Property) becomes the defining feature of decision-making through the mathematical work carried out by the algorithm—the processing and ranking of information to make it accessible. In this way capitalism no longer only innovates with the living labour of humans.

As explained in Chapter 3, the 'general intellect' encompasses the social knowledge of workers––the living labour of users—who are searching online, yet the end products (user data) of these processes are privatised. The subsumption of labour to capital, Marx's third stage of the division of labour, takes form through human words, however the social processing of this configuration is that of humans and the machinic combined. The social force in production is the human labour of the many creating knowledge, which now is often replaced by machines—what Marx

[112] According to Kittler, everyone should be literate in at least one human language and one programming language (1999).

[113] The performativity of algorithms has been much discussed, primarily in the publications of Michael Callon and Donald MacKenzie. John Law and others have written on performative 'states' in regard to research methods. In linguistics, John Austin discussed performativity as 'speech acts', later built upon by Jacque Derrida and Judith Butler in regard to gender theory and the construction of the subject, or 'self-making'. However, this is beyond the scope of my thesis.

envisioned in the Grundrisse with his text *Fragment on Machines*. Thus, the machine 'does not produce surplus value but serves to accumulate and augment surplus value based on the exploitation of the general intellect' (Pasquinelli 2015 cited by Bilic 2017:13). Google's proprietary algorithm has already processed trillions of users' search queries and interactions, yet the inputs and outputs of the black box are not transparent. 'Code is not only invisible but also largely imperceptible in terms of its complex relationship with the economy and political agenda of giant software systems like Google' (Parikka 2010:118 cited Soon 2016:73). With algorithmic capitalism Google's key business strategy remains a trade secret comprised of patents (Bilic 2017:8) and along with it the 'underlying logic of technological rationality' (Marcuse 1941; 1964 cited by ibid:6), enacted through practices of 'visibility management' (Flyverbom et al. 2016).

4.4 (In)visibility Management

As relayed in the beginning of this chapter, AdWords, or Googleconomics, are the backbone of the company's business model, yet that income is derived from 'supporting intellectual property rights laws (IPR)' (Munro 2016:567).

> Transparency matters. And yet many companies go out of their way to hide results of their models or even their existence. One common justification is that the algorithm constitutes a "secret sauce" crucial to their business. It's intellectual property, and it must be defended, if need be, with legions of lawyers and lobbyists. In the case of web giants like Google, Amazon, Facebook, these precisely tailored algorithms alone are worth hundreds of billions of dollars (O'Neil 2016:29).

These companies further promote 'the need to withhold information and protect[ing] strategic positions prevails' (e.g., Sproull and Kiesler 1995 cited by Flyverbom et al. 2017:392) by not being willing to share exactly how this technology works because it constitutes their competitive edge (Noble 2018b).

Under pressure from lawmakers concerning fair use relative to commercial interest, 'information providers often contend that their algorithms are trade secrets that must not be divulged in a public venue' (Gillespie 2014:185):

> Our patents, trademarks, trade secrets, copyrights, and other intellectual property rights are important assets for us. Various events outside of our control pose a threat to our intellectual property rights, as well as to our products, services and technologies (Google market report cited by Bilic 2017:10).

Since buying Double Click in 2007, there have been a series of mergers and acquisitions that have increased Google Search's market dominance by incorporating the IP rights of other companies. Besides providing enormous amounts of revenue, Google's patents play a crucial role in withholding knowledge from competitors and the general public. In Chapter 3, Rieder mentioned that originally two patents were filed (1998, 2001) in regard to the coveted PageRank. As of July 2017, 15.073 patents were ascribed to Google (Bilic 2017:10). The protection of the company's IP corresponds to controlling information and ensuring the scarcity of search services in the market (ibid). This protection of IP has produced not just proprietary software, but perhaps the most revenue-generating corporate secret (patent) of all time.

When Brin and Page were academics they started out with a utopian vision of 'organising all the world's information and making it accessible' and criticised the effects of advertising on search results (1998). In their white paper Brin and Page denounced their competitors as guilty of commercialisation and a lack of transparency (ibid). Although Google proclaims itself to be a 'transparent company' that improves the lives of the people who use its products, if they share their data, the information that it decides to disclose is regulated and organised.

> Thus, if we want to understand how contemporary organizations operate, we need to investigate how they "manage visibilities"; that is, how they make things transparent, keep some things hidden and seek ways to monitor others (Flyverbom et al. 2016:98-99).

Which information is kept hidden by Google is a response to how insight and scrutiny are controlled, 'keeping certain types of information *out of* the open while demanding other information be *out in* the open is what […] management of visibilities bring to the fore' (ibid:107).

This visibility management reflects Google's concealment of IP and patents but also the power dynamics inside the corporation. Despite organisational decisions, such as forcing employees to sign non-disclosure agreements before starting their jobs, 'Google, for example, states that it seeks to "share everything, and trust Googlers to keep the information confidential"'(Flyverbom et al. 2019:397).[114] Although Google 'manages visibilities' by hiding the workings of its algorithms, the SEO industry and researchers (Feuz et al. 2011, Pasquale 2015, O'Neil 2016, Chun 2016, Weltevrede 2016, Noble 2018) have increasingly played an important role in 'visualising' the interior workings of the black box. The focus then might need to shift to visualising secrecy, as power begins with a 'move from a politics of knowing to a politics of seeing' (Flyverbom et al. 2016:104). Keeping in mind, however, that what one sees is not always what one knows.

Regarding the notion of secrecy, George Simmel speculated that 'as the affairs of people at large become more and more public, those of individuals become more and more secret' (Simmel 1906:468 cited in Beyes and Pias 2019:88).[115] Ostensibly the organising of secrecy is reversed– – nowadays individuals fight to maintain their secrets and privacy. Yet the habit of searching could alternatively become a new form of (in)visibility management. In lieu of personalised subjects subjected to Google Search and supplying their data, they could hide, control or even delete it and therefore need not give data away in exchange for free service. Rather than a Google 'trusted user' they could embody agency, evincing a hacker ethic with the goal of being off the radar and able to decide, just as interfaces do, what to show and what not to. Similar to the proprietary corporate search algorithms of Google, the evaluative criteria and code of which are concealed, they instead could find ways to obfuscate their online presence. Hidden from the proprietary algorithms that are designed to be obscure and that facilitate obscurity, the 'trusted user' could become much more like the algorithms, stealthy and arcane, shrouded in the (onion) layers of the Tor Browser instead of the filter bubble of Google Search.

[114] More on Google's 'perks' (2007): https://www.youtube.com/watch?v=XyVDF6BiKtQ. Google also keeps track of all the data it collects on its employees, through 'living labs' and 'nudges' (2013): https://www.youtube.com/watch?v=9ANgEo40VSE

JToasty93 2 months ago

True cyber-geeks know the proper steps required to conceal their identity online. Knowledge is power.

Reply • 7

Hide replies

wellsandlava 1 month ago

You're right, just gotta be smart.

Reply • 1

chupacabra20 3 days ago

+JToasty93 Would you be so kind and share the knowledge?

Reply •

N1KOD95 4 hours ago

+chupacabra20 Tails, throw away laptop, public wifi, VPNs (multiple ones)

Reply •

Chapter 6:

Worlds and Technologies of Anonymity

In Chapters 3 and 5, I explained how Google Search works, concluding with how it manages (in)visibility through proprietary secrets—algorithms. Although Google states that it respects the privacy of their users, users are not by default 'allowed' to be anonymous online. In the next chapter I will attempt to answer the sub-research question, *How can search be reimagined?* but in order to do so, I first need to explain how Tor works. In this chapter I describe the development of technologies that facilitate users' ability to be anonymous online, such as encryption, Tor and TAILS and the 'non-indexed worlds' in which they are situated.

1.0 Cyberspace Anonymity

David Chaum was a computer programmer working at Berkeley in the late 70s when he came up with the idea of using cryptography techniques for encrypting email correspondence, even though only academics had email accounts at this time.[116] His seminal paper *Untraceable Electronic Mail, Return addresses and Digital Pseudonyms* predicted the growth of electronic mail that was not only focused on keeping the content of emails hidden but the transmission of metadata concerning where and when they were sent, as well as the patterns in peoples' emails known as 'traffic analysis' (1981). This was an important realisation concerning networks—who users talk to and when could ultimately pinpoint users. Because of his conviction that users should not have their traffic analysed, Chaum felt the need to create a secure network that obtained strength in numbers. His solution was to apply public key cryptography where '[o]ne correspondent can remain anonymous to a second, while allowing the second to respond via an untraceable return address' (Chaum 1981:1). By developing what he called a 'mix' that processed each item of the mail and the 'use of a "cascade", or series of mixes', it 'offer[ed] the advantage that any single constituent mix is able to provide the secrecy of the correspondence between the inputs and the outputs of the entire cascade' (ibid:3). With communication patterns disguised through cryptography, the user's identity would be harder to distinguish from other users.

Drawing on Chaum's research, during the late 1980s the 'cypherpunks'—hackers and programmers—were united by their proactive usage of cryptography and encryption services that promoted privacy and security with the transmission of messages on the internet. These 'cypherpunks' attempted to envision an online utopia where this territory was off limits to government interference and corporate commercialisation.

> The cypherpunks have developed many tools in order to democratize and promulgate privacy by 'building anonymous systems… defending our privacy with cryptography, with anonymous mail forwarding systems, with digital signatures, and with electronic money' (Hughes1993:2 cited by Munro 2016:19).[117]

[116] Cryptography, from the Ancient Greek 'kryptos' means 'hidden' or 'secret', whilst 'graphein' means 'to write' and was used for centuries all of the world in various methods and applications.
[117] Nowadays WikiLeaks 'explicitly allies itself with the 'cypherpunk' hacker movement which holds that secrecy and privacy ought not to be the privilege of the powerful (Assange et al., 2012; Hughes, 1993 cited by Munro 2016:19).

Cypherpunks realised that encryption wasn't the default setting and created a movement of internet discussion groups that promoted applying anonymity online and privacy as a means to circumvent traditional packet switching. Tim May contributed regularly to the cyberpunk mailing list and most notably, his *Crypto Anarchist Manifesto* published in 1992 describes the 'specter of crypto anarchy, haunting the modern world' (1992).

> The technology for this revolution—and it surely will be both a social and economic revolution—has existed in theory for the past decade. The methods are based upon public-key encryption, zero-knowledge interactive proof systems, and various software protocols for interaction, authentication, and verification (ibid).

In 1991, based on Chaum's research, Phil Zimmerman developed PGP (pretty good privacy) an encryption programme for signing, encrypting, sending and de-encrypting texts to increase the security of users online—the most widely used worldwide. 'The idea is to make such data communications immune to wiretaps, electronic eavesdropping and theft by scrambling the transmissions with a secret code—a security technique known as data encryption' (Lewis 1994).

Furthermore, the 'Internet Shopping Network' as it was called, 'relied on Secure [X] Mosaic' for searching (browsing) the burgeoning global network, which was 'a browser that required users to grasp the concepts of public key encryption technology, a system for securing electronic transactions and managing digital signatures' (Gilbert 2004). Early web browsers such as Mosaic, mentioned in Chapter 2, became at that time (1994) the technical means to navigate, convey information and communicate between parties whilst even commercial transactions were fostered by PGP (Lewis 1994).

In 1995 Daniel Bernstein questioned the export control of 'munitions' by designing secure emails and DNS, with his decision to sue the US government. *Bernstein vs. United States* concerned not the Fourth Amendment but addressed the First, with Bernstein represented by the Electronic Frontier Foundation (EFF) and later on defending himself. He eventually won the case by showing that software was protected speech. By overturning the bans on encryption software, the cypherpunks won the 'cryptowars', which subsequently led to changes in US law with encryption being decriminalised.

> But, in response, the NSA and others simply changed tack – and developed ever more sophisticated systems of spying, some of which would eventually turn into the subject of Edward Snowden's revelations (Bartlett:2015).

2.0 The Privacy Turn

In response to the US government's creation of the Telecommunications Reform Act, *A Declaration on the Independence of Cyberspace,* published in 1996 by John Perry Barlow, applied 'founding father' rhetoric to notions of censorship and freedom on the internet. However, critiques of Barlow's declaration debunk it for its virtual citizenship of 'electronic bodies' that replace real ones, as if the internet is a 'gender-neutral space' (Nakamura 2014:1). In her short *Afterword*, Lisa Nakamura states in response to Barlow that

> [e]arly Internet utopians claimed the Internet would give everyone the power to surveil, to see and not be seen, to become a body-less and thus unseeable user. Instead we have become more visible and trackable than ever (ibid:3).

Due to the 'rise of both the public eye (the eye of the citizen, and of the reporter) and the private eye (the eye of the detective)' (Lepore 2013), self-surveillance, along with the monitoring of others by the press, companies and eventually corporations, became commonplace. Contained within the complexities of privacy, surveillance by outside parties and individuals themselves lead to the creation of pseudonyms when participating online. With social media profiles, comments and blog posts, users constructed numerous identities and insisted on anonymity. In her essay *The Politics of Real Names: Power, Context and Control in Networked Publics*, danna boyd addresses some of the conundrums encircling 'real name' policies in designing socio-technological systems. 'The 'nymwars' as they were called, triggered a passionate debate among bloggers and journalists about the very essence of anonymity and pseudonymity' (Dash 2011; Fake 2011; Skud 2011 cited in ibid 2012:29).

The 'nymwars' turned into a battle between users who desired to have self-determination regarding their online identities and the corporations, who wanted to have the user data assigned to one 'real' person. 'Social norms drove the "real names" culture of Facebook but Google's approach was purely driven by the market and reinforced by corporate policies and technology' (ibid). Boyd's argument concerned the 'furor' surrounding the nymwars, stating that 'real names policies aren't empowering; they're an authoritarian assertion of power over vulnerable people' (ibid cited by Van der Nagel 2015:5). She also articulated the loss of control users experienced in these wars and the shift in power, including those who are querying:

> When people are expected to lead with their names, their power to control a social situation is undermined. Power shifts. The observer, armed with a *search engine* and identifiable information, has greater control over the social situation than the person presenting information about themselves. The loss of control is precisely why such situations feel so public. Yet, ironically, the sites that promise privacy and control are often those that demand users to reveal their names (boyd 2012:31 emphasis mine).

As revealed by Edward Snowden, the divulgence of citizen's private matters since the tapping of telephones and the spying of the NSA (and the Five Eyes) offers a chronological lineage: '[T]he relationship between secrecy and privacy can be stated in an axiom: the defence of privacy follows, and never precedes, the emergence of new technologies for the exposure of secrets'(Lepore 2013). In other words, legislation for privacy is always after the fact whether that be in relation to government surveillance directed at the citizen body or, nowadays, the spying on users by corporations. Although governments state that they need to maintain such secrecy for the safety of their citizens which justifies their surveillance, corporations have attempted to appear as if they are on the side of users. Concomitantly the public has become more aware:

> Edward Snowden's revelations about government surveillance mechanisms have sensitized people to vulnerabilities when they use communications technologies and called this the 'Snowden effect' (Forte et al. 2017:2).

With leaks of classified information and increasing public knowledge about governmental surveillance activities, this 'Snowden effect' has led many users to seek out privacy and anonymising technologies. In the past decade (2010-2020) more choices have become available for the user to inhibit the capture of their identity with their IP address and other meta data. Denoted by the increasing number of email services featuring built-in security, or PGP (Pretty

Good Privacy) encryption, there is a revival of encryption technologies. Technologies such as Chaum's 'mixed networks' or routing protocols intervene between request and destination using a chain of proxy servers with layers of encryption layers. Based on these encryption techniques developed in the 1980s, over the past decade the Tor browser has gained popularity as it became easier to implement, with around 2,5 million daily users worldwide.

3.0 Tor (The Onion Router)

'Tor is a low-latency anonymity-preserving network that enables its users to protect their privacy online' by applying encryption through 'onion routing technlogy'(AlSabah et al. 2012:1). Onion routing was first developed and designed in the 1990s by the US Naval research laboratory, which faciliated encrypted communication between parties in order to secure online intelligence activities. Tor is a relatively decentralised mesh of proxy servers (p2p network) where the data is bounced through relays, or nodes.[118] A node is a computer running Tor software, which takes a request, with data sent through a proxy configuration adding a layer of encryption at every node and whilst still encrypted, sends it to the next random node. In a nutshell this means that the data which is sent over the network is first packed in multiple layers of encryption, which are peeled off one by one by each relay on the randomly selected route the package travels (Spitters et al. 2014:1).

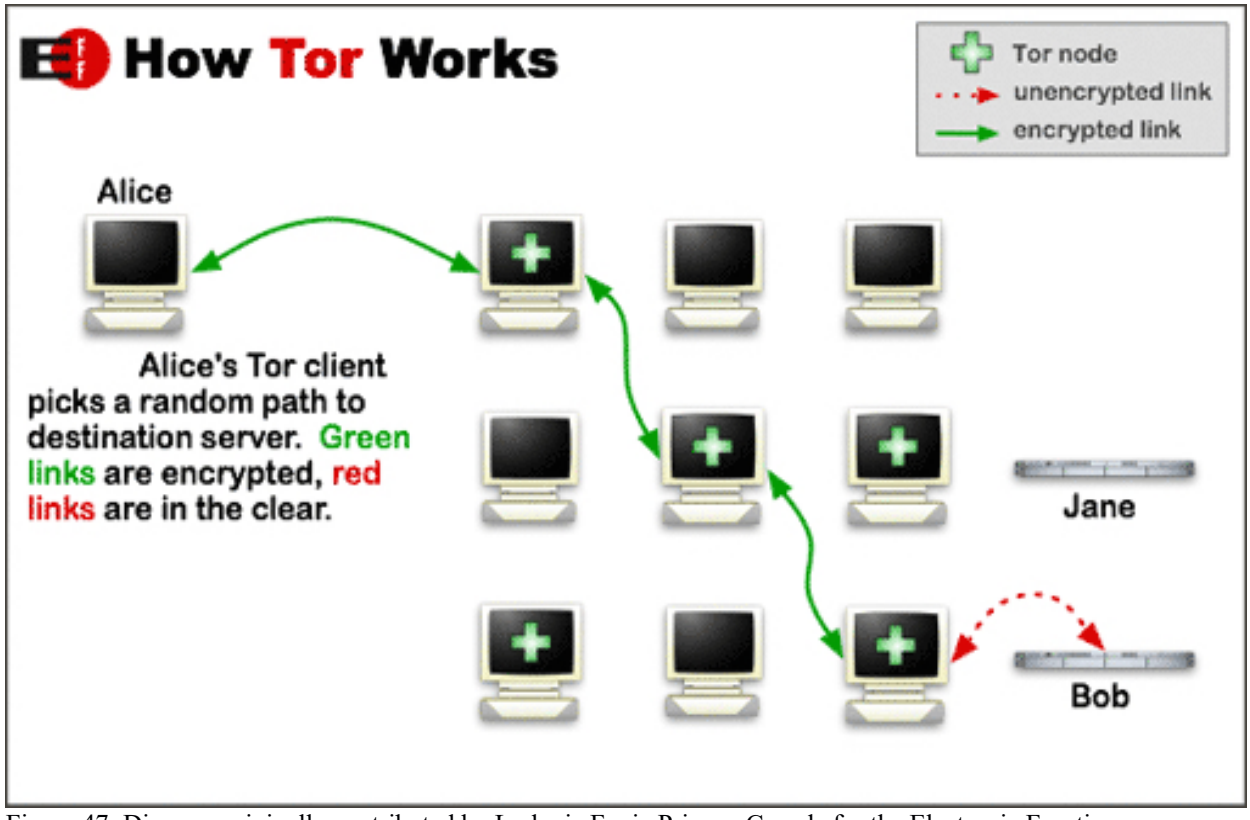


Figure 47: Diagram originally contributed by Ludovic F. via Privacy Canada for the Electronic Frontier Foundations December edition 2011

[118] There is a central structure that is monitored by The Tor Project.

More than 6000 relays facilitate the transferral of data, applying 'onion routing' as a tactic for anonymity (ibid). With 'three hops is anonymity' (Winter et al. 2014:6), Tor is structured by 3 relays (entry, middle, exit) that transmit the communication through a system of circuits, thereby not divulging the IP address of the user and 'hiding' their identity.[119]

> By default, circuits are composed of three ORs [onion router], usually nicknamed the entry guard, middle and exit OR, depending on their position on the circuit. Of the three ORs, only the entry guard knows and communicates directly with the client, and only the exit knows the Internet destination that the client is communicating with, but no OR can link a client to a destination; this is how a client's privacy is maintained in Tor (AlSabah et al. 2012:74-75).

Whilst decrypting the data at every 'hop' and forwarding it to the next onion router, the data packet exits the closed system and only then does the user's IP address become 'transparent'. (Figure 47) When it finally reaches its destination, whatever website one is trying to reach, the response comes back, with everything happening in reverse until the content is displayed in the Tor browser. At the end of a browsing session user history is deleted along with the HTTP cookie. 'This complex process means that if a message is intercepted at any point it is almost impossible to identify its origin, its content or the intended recipient, even if you have the resources of the NSA' (Glenny 2015:2).

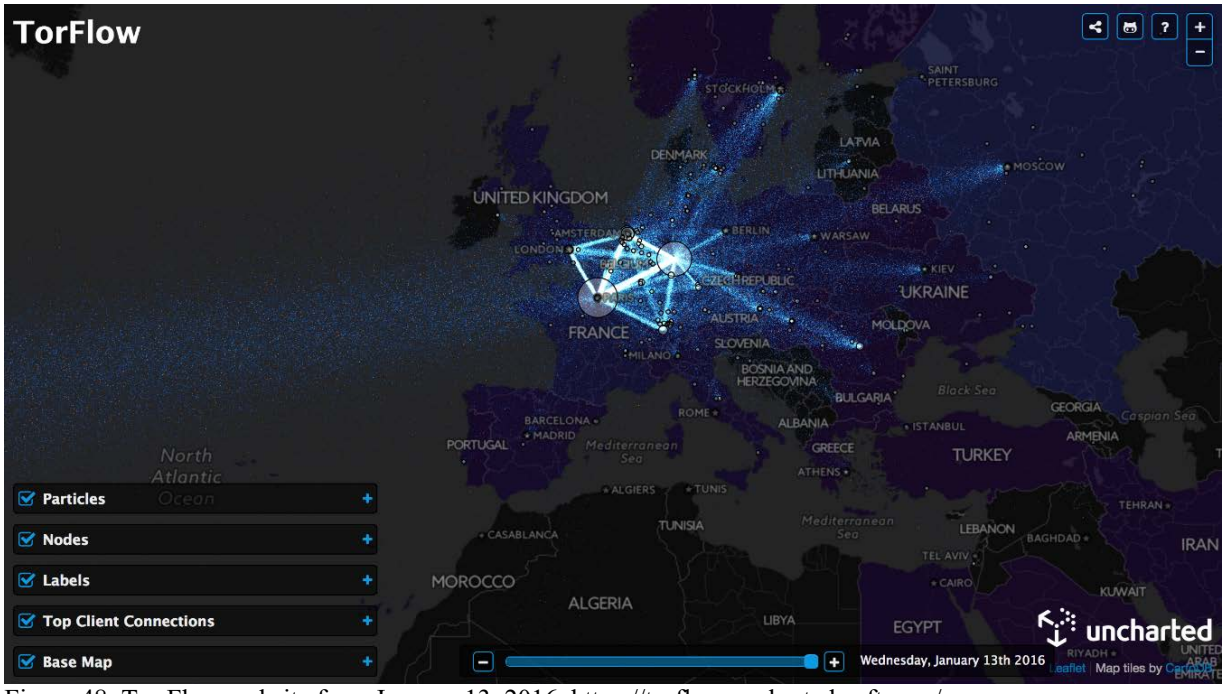


Figure 48: Tor Flow website from January 13, 2016. https://torflow.uncharted.software/

Released to the public in 2002, prior to becoming a not-for-profit in 2006, Tor is a browser that is downloadable for free by anyone with an internet connection. Unlike a search engine which builds its own index or aggregates, the Tor Browser Bundle has an encoded (default) search

[119] 'In general, the complete connection between client and onion service consists of 6 relays: 3 of them were picked by the client with the third being the rendezvous point and the other 3 were picked by the onion service' (Tor Project). Available here: https://www.torproject.org/docs/onion-services.html.en

engine that incorporates Mozilla's Firefox browser. Additionally, the Tor network primarily consists of volunteers from around the world who provide servers and act as relays, facilitating the Tor traffic to flow. (Figure 48) Since 2013, more and more people are using Tor and, as of writing, there are estimated to be between 2,000,000 and 3,000,000 users at any given time worldwide. 'For privacy purposes, this is a positive development; more Tor users means more traffic and thus more obfuscation of user identities' (Gehl 2014:5 citing Dingledine et al. 2014). By increasing the amount of Tor users, the stronger the anonymity becomes because with more users, Tor is harder to crack.

> Losing users has a significant impact on the anonymity provided to users, since reducing the user base results in decreasing the size of the network's anonymity set (AlSabah et al. 2012:73).

However, there is contention surrounding the Tor network. On the one hand it provides anonymity to users by protecting them from being surveilled by governmental entities, yet Tor is now an open source non-profit 501c3 that receives a considerable amount of support from the US government. 'It has received funding over the years from governments, NGOs, foundations and companies, as well as thousands of personal donations' (Tarasov 2018). Moreover, controversy exists in the media mostly in regard to the so-called 'Dark Web' or Tor 'hidden services', ranging from the selling of illegal drugs, weapons and child pornography to sites of anarchism, hacktivism and politics (Spitters et al. 2014:1), which I will address shortly.

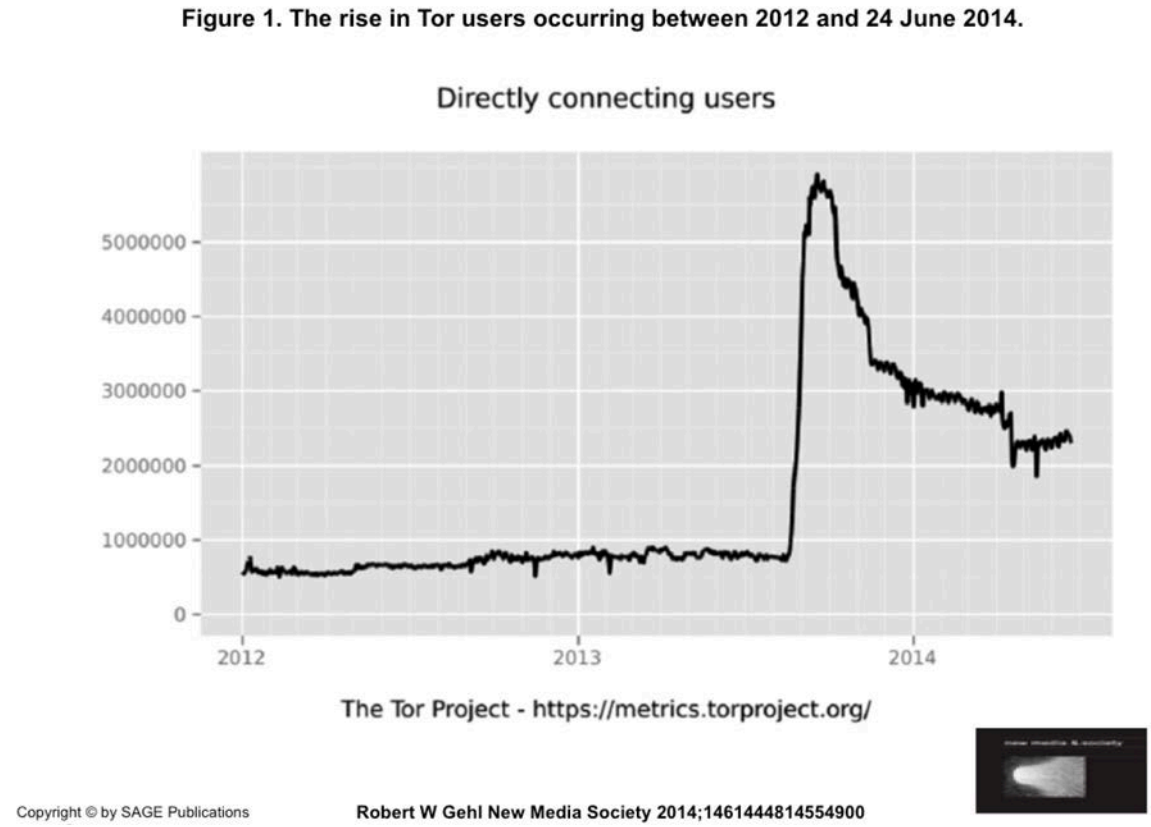


Figure 49: Spike in Tor users (Gehl 2014).[120]

[120] Although Tors users increased, the spike in September 2013 was due to a bot. Available here: https://blog.torproject.org/how-handle-millions-new-tor-clients

In 2014 certain members of the UK government suggested banning Tor and other online anonymity systems and the Chinese government attempted to block and forbid it, whilst others apply it as a life-saving and anonymizing technology. As the US National Public Radio (NPR) reported,

> Tor's executive director is working with victims of domestic abuse who need to communicate without being tracked by their abusers. Tor is also used by Chinese dissidents who can't access sites like Twitter. And it became a valuable tool during the Arab Spring' (Gehl 2014:5 citing Rath 2014).

Therefore the risk involved in using Tor has become more pronounced, as Tor has often been accredited the past few years in protecting the anonymity of the user in areas of protest and freedom of speech, often with the addition of bridges.[121]

> Tor today is an influential anticensorship technology that allows people in oppressive regimes to access information without the fear of being blocked, tracked or monitored. The importance and success of Tor is evident from recent global uprisings where the usage of Tor spiked as people used it as a revolutionary force to help them fight their social and political realities (AlSabah et al. 2012:1). (Figure 49)

Political activists and citizens in dangerous areas of the world, such as war zones and dictatorships, use Tor to protect their online communication, transfering 'human rights activities into other identities through the Tor network' (Forte et al. 2017:6). Dissidents also use Tor to post content and download sensitive material, yet the technologies need to be in place beforehand because 'Tor is one of those things where you want to have it before you need it, for obvious reasons, because if you're being censored, it's very hard to get' (ibid:7).
Though much quicker than in the last decade, Tor is still slightly slower than other browsers, 'but if you value privacy or if you would like to find a way to circumvent the online tracking or if you would like to become a more informed, more active Internet user' (Emerson 2016), then Tor is an alternative. It is crucial to continuously improve the performance and usability of Tor to enhance the anonymity it provides (AlSabah et al. 2012:73). Moreover, Tor enables the user to access regions of the internet that are not indexed by 'clear net' search engines.

4.0 Non-indexed worlds

Returning to Chapter 2, there were various search engines available for querying before and during the dotcom bubble burst of 2000. For instance, Maxine, the fraud investigator protagonist of Thomas Pynchon's novel *Bleeding Edge*, names LexisNexis, HotBot and AltaVista, prominent search engines at that time (2001). However, Google is missing from the list, 'making its absence all the more relevant' (Haeselin 2017:4).

> "Nah, this is one of the dotcoms that *didn't* go under last year in the tech crash. No old software', half a decibel too quiet, "and maybe no statute of limitations either."

[121] 'A bridge is a middleman Tor node that is not listed in the main public Tor directory, and so is possibly useful in countries where the public relays are blocked. Unlike the case of exit nodes, IP addresses of bridge nodes never appear in server log files and never pass through monitoring nodes in a way that can be connected with circumvention' *The CryptoParty Handbook*, Version: 2013-08-21, p. 333.

> Uh-oh. " 'Cause see, if all you want's an asset search, you don't need a forensic person really, just go on the Internet, LexisNexis, HotBot, AltaVista, if you can keep a trade secret, don't rule out Yellow Pages—"
> "What I am really looking for," solemn more than inpatient, "probably won't be any place any search engine can get to."
> "Because… what you're looking for…"
> "Just normal company records—daybooks, ledgers, logs, tax sheets. But try to have a look, and that's when it gets weird, everything stashed away far far beyond the reach of LexisNexis."
> "How's that?"
> "Deep Web? No way for surface crawlers to get there, not to mention the encryption and the strange redirects—"
> "Oh. Maybe you need more of an IT type to look at this case?' cause I'm not really—" (Pynchon 2013: 9-10)

In this excerpt, Maxine explains that mainstream search engines already carry out 'asset search' to her video producer friend Reg who informs her that the Deep Web, or 'non-indexed world', is not where these search engines go, as much of it is not interesting to search engines because of the content.

> There is the unsearchable, and then there is the unsearchable. The latter is the domain of the Deep Web, a space built by programmers who do not comply with the robots exclusion standards simply because commercial search engines see little profit in crawling that data in the first place (Haeslin 2017:5)

As outlined in the Introduction, in 2000, when Introna and Nissenbaum wrote their seminal text, *Shaping the Web:Why the Politics of Search Engines Matters*, many portions of the web they referred to were largely not indexed, that is, search engine spiders had not yet registered them in their databases. Although Google and its anticipatory searches create uniformity, it ostensibly has still only indexed between 5 and 13% of the Internet, or 'surface web'. This percentage varies—depending on where one obtains data—because it hasn't been indexed it's quite difficult to estimate, so the percentage is not fixed. The 'vanilla internet' or 'clear net' as it is called is quite small compared to the 'unknown' of this Deep Web, which is online but the general public cannot access it, or at least it is more difficult to find.

> The Defense Advanced Research Projects Agency (DARPA) explains the Deep Web as simply "the parts of the web not indexed by standard commercial search engines," including unexceptional places such as private commercial forums, sites protected by passwords, as well as sites with no inbound traffic that can be reached only anonymously by proxy (" Memex")(Haeselin 2017:1).

People often use the terms Deep Web, Dark Web or Dark Net interchangeably but in this thesis I refer to the 'Deep Web' as the hidden part of the internet where online material sometimes appears for brief time intervals, or has specific login barriers (such as universities, military, companies, etc.) and can be accessed by certain users but is not (yet) indexed by commercial search engines. Others describe the Deep Web as just unstructured code or data. 'It is those sites not indexed by conventional search engines: an unknowable realm of password-protected pages, unlinked websites, and hidden content accessible only to those in the know, sometimes referred to as the "deep web"'(Bartlett 2014:5).

Inside, or rather alongside, exists what is nowadays called the 'Dark Web' or 'Dark Net'. The Dark Net or Nets were terms coined in the 1970s to designate networks isolated from the ARPANET (Advanced Research Projects Agency Network), a US governmental entity that eventually evolved into the present-day internet. According to the *Way Back Machine*, Internet Archive's search engine pre-eminence, '[t]he root of the name is believed to be related to the term black box, which meant a system or device whose contents were unknown.' Coincidentally, the term Dark Net was also used to refer to the internet back in the 1990s, when 'preGoogle' search engines would find all kinds of 'irrelevant' websites that would hinder searching for relevant information.[122] Later on the term gained traction in a 2002 publication, *The Darknet and the Future of Content Distribution* (Biddle et al. 2003), that denoted the problems of 'copying objects' in the early days of the internet and their distribution, where friend to friend networks or filesharing, as it is now known, was considered illegal in regard to DRM (Digital Rights Management) technologies and copyright laws.

Although in the past the Dark Net seemed to embody a decentralised model of privacy sharing networks, often peer-to-peer, effectively the 'darknet-genie will not be put back in the bottle' (Biddle et al. 2002). Nowadays they are dynamic pages with content that mirrors the 'clear net' and it is accessible by more and more users worldwide. However, there is no exact space as such, rather the 'Dark Net' is

> an idea more than a particular place: internet underworlds set apart yet connected to the internet we inhabit, worlds of freedom and anonymity, where users say and do as they like, often uncensored, unregulated, and outside society's norms (Bartlett 2014:5).

The Dark Web, Othernet, Alternet or Dark Net then exists within/alongside the Deep Web, intentionally hidden, where one needs special software (Tor) to enter and navigate this part of the internet, cloaked in anonymity. The general media and public opinion portray the Dark Web or Dark Net as a territory where buyers can purchase any type of drug, rent hackers or hire hitmen. For the *Christian Post*, it is a place for terrorists, as expressed by their headline: 'ISIS should be banned from the internet to the Dark Web' (2015). Yet others conceive the Dark Net as a parallel universe, or 'dark' mirror-image of the surface web with social networking sites, bulletin boards and companies delivering commodities and services, where people communicate securely and secretly. In other words, the content is public on these sites but off the radar for commercial search engines. Media theorist Jamie Bartlett describes the Dark Net as the

> encrypted world of Tor Hidden Services, where users cannot be traced, and cannot be identified. It has also become a catchall term for the myriad shocking, disturbing, and controversial corners of the net – the realm of imagined criminals and lurking predators (2014:5).

5.0 Tor Hidden Services

The Tor browser enables access to its 'hidden services', a melange of markets with goods to buy ranging from drugs, weapons and pornography to forums on human rights, freedom of speech, anarchy and obscure websites. 'Today [2014] there are thought to be b/w 40000 and 60000 Tor Hidden Services sites in operation (due to its encryption system it's very difficult to measure

[122] According to my interpretation of Pynchon's *Bleeding Edge*, this is where the search engine 'DeepArcher' operates.

accurately)'(ibid:136). Following Biryukov et al., 'Tor hidden services are a feature which was introduced in 2004 to add responder anonymity to Tor' (2013:80). Not indexed by commercial search engines, these 'hidden services' publicize their existence through Tor using a public encryption key, which is then indexed in Tor's directory servers. The system bypasses the direct connection to a server and uses what is called an 'identifier', which is an automatically-generated service name of 16 characters derived from the service's public key (such as *xyz.onion*) that makes them findable. Unlike the nomenclature of URL's, where meaning and usage is intertwined with Google Search, onion addresses are less memorable and the URLs are 'a string of meaningless numbers and letters that end in .onion' (Bartlett 2014b:4).

> Moreover, those who run dark websites that end in .onion are able to hide their identities and locations from most, if not all, Internet users. The idea being is that in most cases, a visitor to an .onion site will not know the identity of the host, nor will the host know the identity of the visitor (Gehl 2014:2).

Client and server meet at a rendezvous point, which are distributed within these hidden services. The user is not exposed to the 'clear net', unless one leaves through an exit node of the Tor network. 'The main goals behind hidden services are access-control protection, robustness of servers and hiding the true identities of hidden service administrators' (Çalışkan et al. 2015:50).

According to an empirical study *Trawling for tor hidden services: Detection, measurement, deanonymization*, 'the number of hidden services with illegal content or devoted to illegal activities and the number of other hidden services (devoted to human rights, freedom of speech, anonymity, security, etc.) is almost the same; among Tor hidden services one can even find a chess server' (Biryukov et al. 2013:84).

> The evidence on this is pretty clear. Although it is vital for free expression, according to researchers at the University of Luxembourg, 44 percent of Tor Hidden Services are given up to criminality (mainly anonymous market places and illegal pornography)' (Bartlett 2014b:9).

A 2014 study, *Towards a Comprehensive Insight into the Thematic Organization of the Tor Hidden Services*, points out that the goal of these types of Dark Nets was originally to propagate freedom of speech and provide anonymity to dissidents. Only later did 'malicious actors' move in to establish illegal markets. The researchers from this study mapped these 'hidden services' around various themes:

> We applied classification and topic model-based text mining techniques to the content of over a thousand Tor hidden services in order to model their thematic organization and linguistic diversity. As far as we are aware, this paper presents the most comprehensive content-based analysis of Tor to date (Spitters et al. 2014:220).

They designed their crawler to collect data from a wide range of hidden services and discovered around 2000 new hidden services. Subsequently they applied classification text-mining to reveal topical content and linguistic diversity. Although they were able to collate many of the hidden services, they also add that content is very dynamic and that its lifespan is unpredictable; many sites are not available or disappear, only to resurface at a later date. They conclude by stating that their study 'infer[red] a topic taxonomy, which indicates that most hidden services—at least in our data set—exhibit illegal or controversial content' (ibid: 223).

Table II, reproduced here, shows the inferred keywords made from the above-mentioned methodology, extrapolated from their collected corpus of around 250 words. (Figure 50)

TABLE II. THE TOP WORDS FOR SOME SAMPLED TOPICS OF OUR 250-TOPIC LDA MODEL. THE TOPIC LABELS WERE MANUALLY ADDED.

Topic label	Most likely topic words
Trading	product quality packaging order good bitcoin btc buy account money usd payment wallet cash paypal coins currency price cart worldwide
Drugs	mdma cocaine quality lsd pills coke speed ketamine netherlands crystal usa methylone high drugs tested meth gram xtc lab heroin results
Anarchism	anarchist social anarchism power movement class workers revolutionary struggle political revolution anti capitalism left groups society action
Hacking	hack ddos hacker channel irc rat bot anonymous attack anon exploit virus website botnet high remote zeus lulzsec quality hackerspace project
Child Pornography	video boy vids download child sexual kids age women pedo adult young girl abuse nude pedophilia jailbait pics society consent pedoempire
Weapons	9mm rifles guns price automatic ammo glock cart payment big 22lr smoke handguns 45acp shotguns special pistols gauge 380acp redstar ball

Figure 50: Table II (Spitters et al. 2014)

What is also noteworthy is that besides building a taxonomy, they added labels manually. 'The result of this manual annotation step is twofold: first of all it gives us a labelled topic model which can be used for classifying unseen pages, and second, it provides us with a list of topic labels that can serve as the basis for our Tor topic taxonomy' (ibid:222). Afterwards they constructed a list of the current examples, mostly consisting of sites well-known by reputation in the left column and then assigned keyword topics to them (Table III). (Figure 51)

TABLE III. EXAMPLES OF TOR HIDDEN SERVICES IN OUR DATABASE AND THEIR TOPICS.

Title	Topics
The Hidden Wiki	Trading, Counterfeit, Child Pornography, Technology, Security, Services, Pornography, Software, Anonymity
Silk Road market place	Trading, Drugs, BTC, Security
Code:Green - hacktivism	Hacking, Security, Software, Anonymity
Pedo Support Community	Child pornography, Communities
HackBB forum	Credit Cards, Trading, Hacking, Paypal, Financial, Bank Accounts, Exploits, Security
Fight back the state	Anarchism, Politics, Law Enforcement
EuroGuns	Weapons, Trading, BTC
Rent-A-Hacker	Hacking, DDoS, Exploits, Services, BTC, Trading

Figure 51: Table III (Spitters et al. 2014)

Finally, they constructed a taxonomy graph using the frequent keywords and topics in order to make a data visualisation of hidden services, with the larger nodes (vertices) reflecting how many sites existed for that particular topic (degree of presence) and the edges, or links giving weight to the amount of times these sites where connected, exemplifying their 'relatedness'. (Figure 50) The correspondence between node size and actual percentage is as follows: Trading: 59%; Software, Security: 20-25%; Child pornography, Drugs: 10-15%; Weapons, Anarchy, Doxing: 2-5% (ibid:223). The amount of hidden services fluctuates constantly and at the end of June 2020, there were around 200,000 unique onion addresses registered but it is unknown exactly how many are accessible at any given time.[123]

[123] See statistics here: https://metrics.torproject.org/hidserv-dir-onions-seen.html

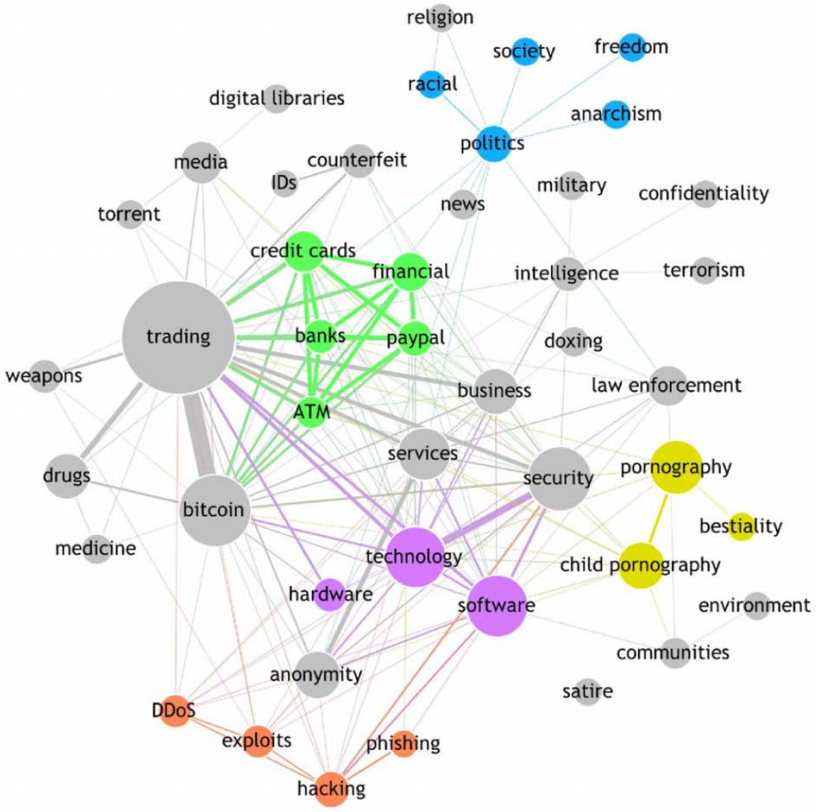


Fig. 3. Topic taxonomy, visualized as a graph, in which the size of a node expresses the degree of presence of the topic on the hidden services, and the weight of an edge expresses the relatedness of the nodes it connects.

Figure 52: Figure 3 (Spitters et al. 2014)

6.0 And then you disappear (TAILS)

With this study in mind I decided to see if could find the lesser-known topics from Table III such as 'Fight back the state' or 'art' by using the various Tor hidden services search engines. Instead of being spied upon by the government, targeted by corporate ads and receiving filtered information that attempts to alter my behaviour when 'surfing the clear web' or even when carrying out experimental research, I felt it necessary to attempt to become anonymous online and to analyse the effects. In this 'critical ethnography of the self' I set out to (re)search the Dark Net or Tor hidden services. However, when searching the Dark Net in particular, you need Tor *and* a VPN (in 2015 I tried to do without a VPN and didn't get anywhere just using Tor).

Instead of a VPN and Tor an anonymous hacker suggested I just use TAILS, an acronym for The Amnesic Incognito Live System where the Tor browser is configured within the TAILS software. Designed to enable privacy and anonymity for the user by applying encryption, TAILS can be installed on a USB stick that boots from the computer. At the end of the browsing session, the data is completely deleted, unless the user has changed the configuration. In order to save complete webpages and screenshots I didn't follow the default settings, where upon shutting down my computer all data is eradicated. Instead, I kept my data by inserting another USB stick into the computer so that I could save my results, which I draw upon in the following section.

A place without limits, a place to push boundaries, a place to express ideas without censorship, a place to sate our curiosities and desires, whatever they may be. All dangerous, magnificent, and uniquely human qualities (Bartlett 2014b:5).

Chapter 7:

The Anonymous User

1.0 Anonymous Users

Drawing on my method, 'a critical ethnography of the self', *The Anonymous User* answers my research subquestion: *how can search be reimagined?* By utilising Dark Net search engines with the Tor (The Onion Router) browser configured with TAILS, which I explained in Chapter 6, in this chapter I elucidate my results interwoven with ethnographical research (Gehl 2014; Bartlett 2014), Gary Marx's 'Rationales for anonymity' as well as an empirical study on Platform Criminalism. My alternative results reflect a range of activities by (other) anonymous users on Tor Hidden Services, which include an interview from an anonymous Dark Net search engine developer, hactivism and activism articles. However, I am just a 'Torist'. I then describe one particular 'deanonymised user', addressing 'expectations of privacy' and the differential treatment of those adopting Tor.

1.1 Escaping Google Search

Before I attempted my methods, I (re)searched by searching with Google on the clear net for articles about the Dark Web. I found Robert Gehl's ethnographic study of the Dark Web Social Network, or DWSN for short. Located with Tor's Hidden Services, DWSN shares many features of social networking sites and forums—discussion groups that focus on the site itself.[124] The three focal points of Gehl's DWSN ethnography concern 'anonymous/social networking, the prohibition against CP (Child Pornography) and the productive aspects of techno-elitism' that are implicit in order to reach the Dark Web (Gehl 2014:1224), and which I focus on here. After reading Gehl's article I attempted to find the DWSN using Dark Net search engines and Wikis without success but also by just 'googling it'. However, the social network attempts to be off the radar from major search engines, as the community does not wish to be indexed. Early in the site's history it was discovered that Google was indexing DWSN, meaning that anyone conducting a Google search could see content from the DWSN, even without a Tor-equipped browser. This is illustrated by the reaction to 'clear web' intrusion into DWSN when an anonymous admin called for a vote on whether to allow Google to continue indexing the site. The admin somewhat shaped the discussion by noting,

> even if [DWSN]'s objective is to provide a non-government, non-profit, anonymous dark web social network for citizens worldwide to share fruitful ideas, the fact that we can be seen [on Google search results] kinda sucks (Gehl 2014:1229).

Many users weighed in on the matter, arguing that Google should not be able to index anything on the DWSN. One member proclaimed,

> I am here because I don't want to contribute to the likes of google and facebook. Allowing major entities to index the content here … will continue to have a significant

[124] During the ten months of his research (2013), Gehl remarks that the number of accounts grew from 3000 to 24000 and that the number of blog posts reached tens of thousands of micro-blog posts (Gehl 2014:1222).

> chilling effect on what users will feel safe sharing here. Yes, [DWSN] is a 'social network' but that doesn't mean we shouldn't value user privacy above all else … (ibid).

In order to escape the commercialisation of the 'clear net' and Google Search, with its personalisation, profiling and targeted AdWords as elucidated in Chapter 5, I decided to reimagine search and explore the potentials of using the Tor Browser and its default search engines. I investigated if I could be anonymous, if so to which degree and what kinds of results would I find?

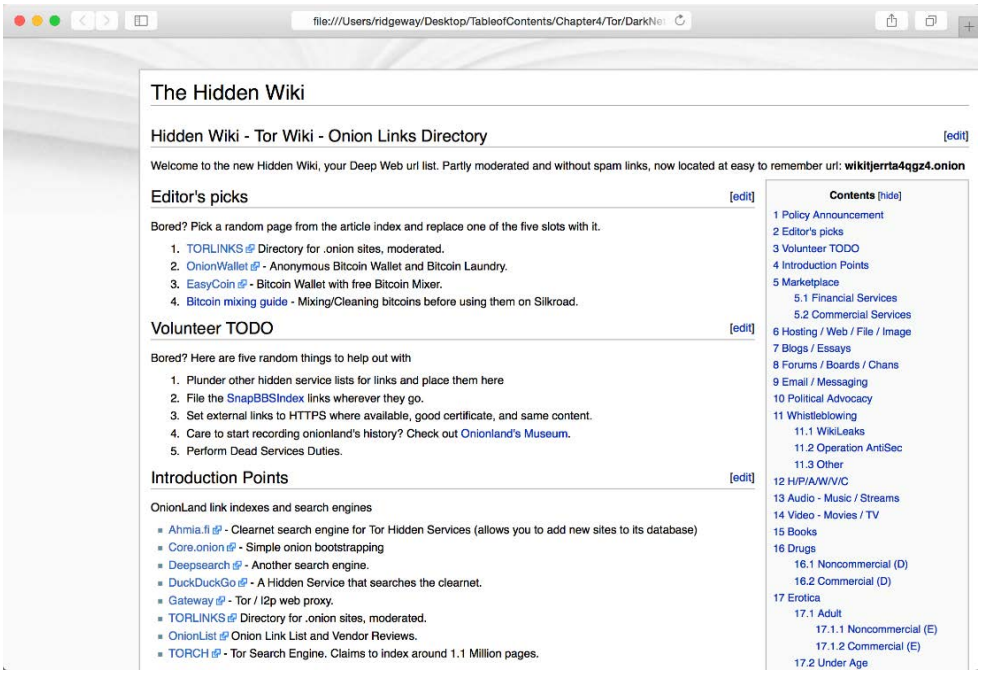


Figure 54: The Hidden Wiki: Editor's picks, Volunteer TODO, Introduction Points 03.10.2015

1.2 Guided To(u)r

"She got busted for all kinds of weird stuff. She's under investigation for a half-dozen things, all kinds of ethical violations. They found everything on her computer, a hundred weird searches, downloads—some very creepy stuff" (Eggers 2013:205-206).

Armed with TAILS, I began my exploration of the 'Dark Net' or 'Tor's Hidden Services' and, analogous to traveling to unknown destinations offline, I needed a 'map' and background information. 'The Hidden Wiki' acts as a guidebook but is not necessarily hyperlinked and, as with many sites on Hidden Services, there are also many 'mirrors' and 'forks'. Dynamic in nature, where many sites are down or offline, the Hidden Wiki also provides a list of search engines available at that moment, under the rubric 'Introduction Points' (Figure 54). Because the URL's onion addresses (random numbers and letters) are hidden to users, there is no connection to the name of the keyword, as with clearnet search. This is one of the main reasons why search engines are of great importance in navigating Tor Hidden Services. If you do not know the specific URL, as with the clear web, seemingly the easiest way to find sites is to use search engines that have been designed for the Dark Net. When I started my exploration in 2015, there was a range of search engines for different types of users. One is 'Torch', along with 'Grams' and 'Ahmia', with its founder, Juha Nurmi explaining three differences between the search engines of the Dark Net and clearnet:

> '"First, the linking between onion sites"—hidden sites on the Tor network with the top-level domain suffix.onion—"is thin; as a result, algorithms using the backlinks aren't working very well," he said. "Second, it takes time to crawl everything because Tor is slow. Lastly, onion sites are replacing their addresses all the time"'(Frediani 2014).

New 'search engines' are also added constantly and I attempted to locate a few but could not access others at my moment of search.[125] First I decided to test out my eclectic set of keywords *Re:search - Terms of Art* (Appendix D), choosing **artistic research** on Grams. I received publications on how to do research on the Dark Web with Tor (2014) and a book about how to search with Google (2012), which I could have found on Amazon as well. (Figure 55)

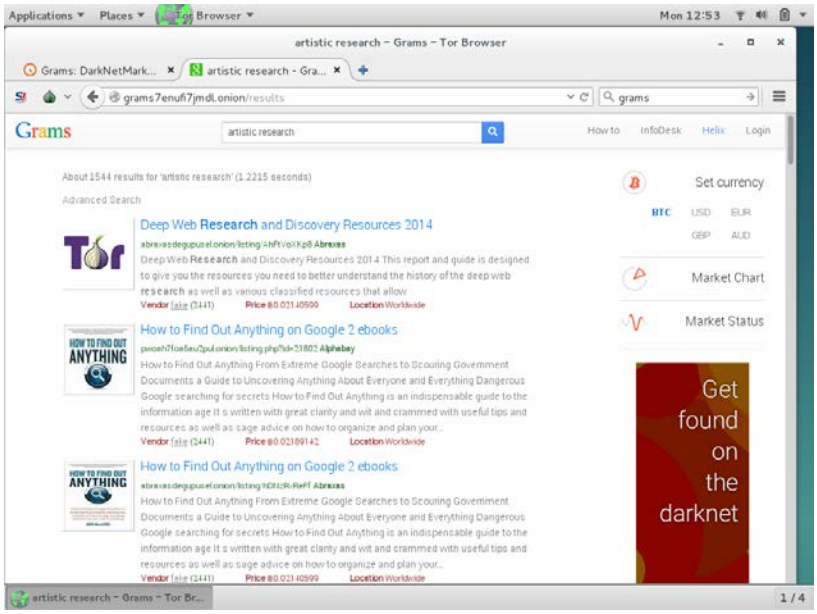


Figure 55: keyword **artistic research** with Grams

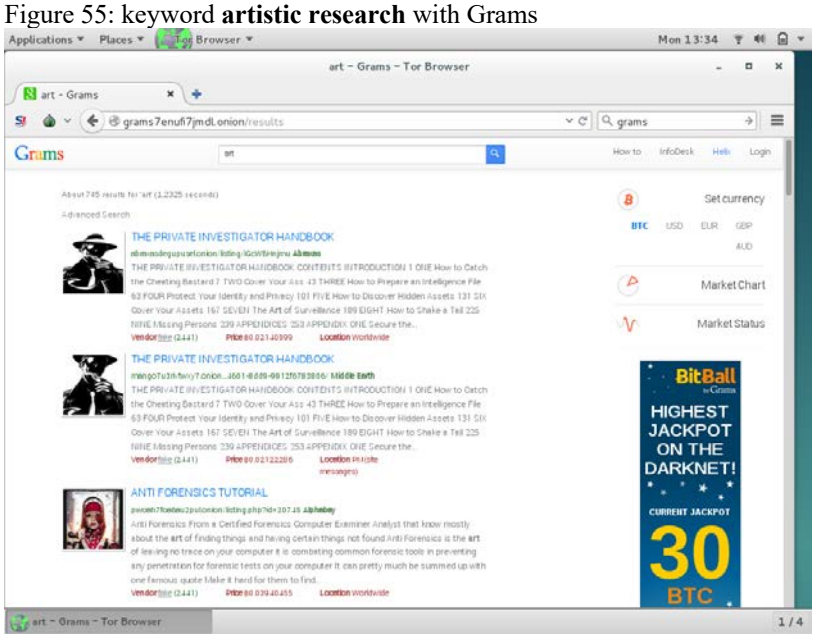


Figure 56: keyword art with Grams

[125] As mentioned in Chapter 2, Memex is the search engine designed by DARPA (Defence Advanced Research Project Agency). It is used by them to search for 'criminals' (Zetter 2015).

I then searched with the keyword ‘art’ and found a handbook on how to be a private investigator, based on the phrase the ‘Art of Surveillance’ and an anti-forensics tutorial on the ‘art of finding things’ and ‘having certain things not found’. (Figure 56)

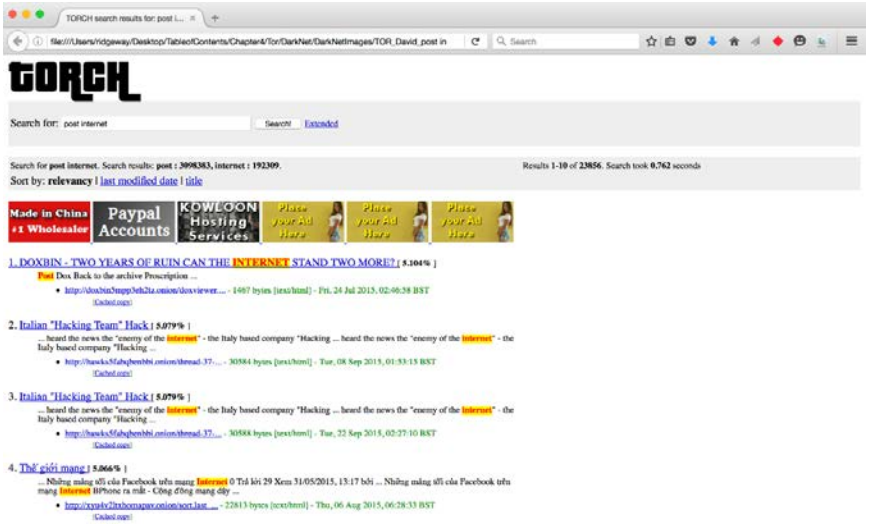


Figure 57: keyword **postinternet** with Torch

With **postinternet** using the Torch search engine I found articles about ‘Hacking Team’, which I will discuss in the next section. (Figure 57) My conclusion was that ‘art’ vocabularies do not apply to the content, which has been linked with other keywords, because this part of the Deep Web is not necessarily so ‘deep’ as to contain these terms. After searching a while on Hidden Services I felt a bit exasperated by the fact that the Dark Net is so hard to navigate—I wandered, got lost and had to return to the Hidden Wiki, or one of them, where many links didn’t work. What I found in the following is based on what I was able to access at that moment, with other sites constantly replacing takedowns.

> Tor Hidden Services act as something of a recycling plant: people upload illegal material to a central hub – and then hundreds download onto their own servers. When it’s taken down, someone simply uploads it again (Bartlett 2014b:11-12).

Having failed at my attempt to find enough search results with my ‘Terms of Art’, I decided to search using keywords from the topical analysis by Spitters et al. (2014) from the previous section, such as ‘anonymity’ and ‘anarchy’.

1.3 Anonymia

The sociologist Gary’s Marx’s text, *What’s in a Name? Some Reflections on the Sociology of Anonymity*, begins by citing Mark Twain’s infamous paradoxical predication: ‘reports of either the recent death or coming dominance of anonymity have been greatly exaggerated’ (1999:99). Remaining contrarian, he argues for and against anonymity as well as cogently pointing out that ‘[i]ronically, anonymity is fundamentally social. Anonymity requires an audience of at least one person’ (ibid:100). Written at the dawn of the millennium, around the time of PageRank’s first implementation and Introna and Nissenbaum’s seminal text on the politics of search engines (2000), Marx draws mostly on the issues surrounding anonymity in contemporary life in the ‘meat world’, before the online era of personalised (Google) search, and maps out the ‘conceptual landscape’ in various contexts of privacy and anonymity.

TABLE 1
Types of identity knowledge

1. Legal name
2. Locatability
3. Pseudonyms linked to name or location
4. Pseudonyms that are not linked to name or location
   a. For policy reasons
   b. Audience does not realize it's a pseudonym
5. Pattern knowledge
6. Social categorization
7. Symbols of eligibility/noneligibility

Figure 58: Types of identity knowledge (Marx 1999)

Marx begins by defining his contexts with what he calls 'Types of identity knowledge' that function 'as an aspect of informational privacy', which 'involves the expectation that individuals should be able to control information about themselves' (Marx 1999:100). (Figure 58) According to Marx, in order to achieve anonymity, there are seven dimensions of identity knowledge that would need to be subverted,

> legal name, location, behavior patterns, social group membership, identifying personal characteristics, pseudonyms that can be linked with other forms of identity knowledge, or pseudonyms that cannot be linked and serve as alternate identities' (ibid; Forte et al. 2017:3).

Concentrating on the first two 'rationales' of Marx's Table 1, a person's legal name answers the 'who' question and contains social and biological information [usually paternally organised]. The second is 'locatability' with which the person is identified with an 'address' that enables what Marx defines as 'reachability'. This harks back to the bureau d'adresse with its collections of house addresses described in Chapter 1. Later on governmental methods include using numbers to identify citizens and, within private homes, the telephone became the identity marker. Marx's elucidates how, with early telephone usage, connections had to first pass through an operator before automated switching was introduced, not for privacy but for efficiency. As stated in Chapter 5, people are now identified online by their IP address. Furthermore ISPs (Internet Service Providers) control users' access to the internet and can view all the websites visited by the client (user) even if they use Marx's third rationale, pseudonyms, which are not linked to a name or a location. Everyone who uses Tor Hidden Services applies a pseudonym (user name) connected to an email when communicating and interacting on the Dark Net. Moreover, onion addresses serve as pseudonyms for websites.

I began searching Grams with the term 'anonymity' and first found how to 'disappear and live free forever with an 'Anonymity Guide'. (Figure 59)

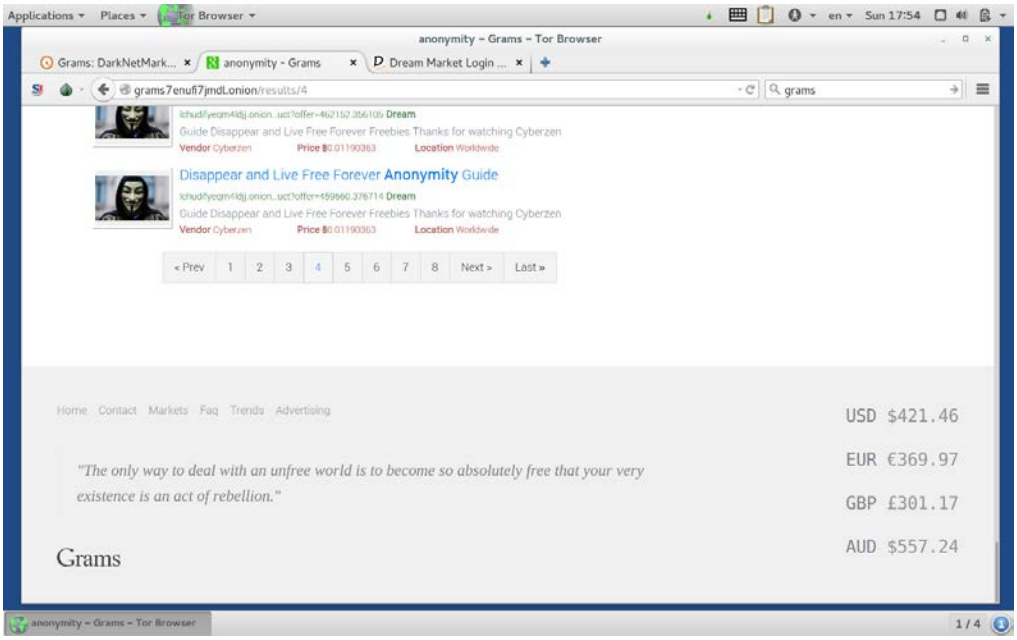


Figure 59: keyword 'anonymity' with Grams

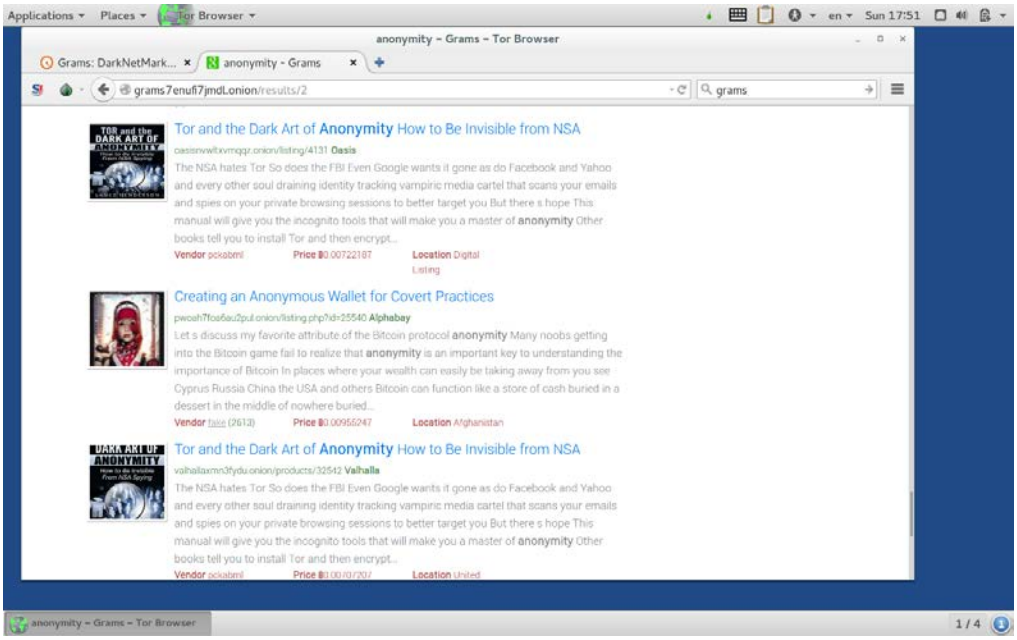


Figure 60: keyword 'anonymity' with Grams

Furthermore, my results support the dissemination of literature, 'how to guides', geared to how to be anonymous online and guidebooks where it was unclear who compiled them.[126] The keyword 'anonymity' delivered information on 'Complete Anonymity', or how to create an anonymous wallet to use Bitcoin and a guidebook, 'Tor and the Dark Art of Anonymity', or even 'How to be Invisible from the NSA'. (Figure 60)

[126] According the US Supreme Court ruling *Talley* v. *California* (1960), 'anonymous pamphlets, leaflets, brochures and even books have played an important role in the progress of mankind' *Talley* v. *California*, 362 U.S. 60, 64 (1960). http://www.law.cornell.edu/supct/html/93-986.ZO.html

Marx's Table 2 structures the usage of anonymity in various contexts and situations, what he deems 'Rationales for anonymity'. (Figure 61) Here I draw attention to the context of anonymous authorship, using pen names or a nom de plume in regard to publishing to protect 'reputation and assets'. If one refers to the Greek etymology of the word, anonymia, it is that without a name, both human and object, whilst the Latin term, *Anonymus* is usually reserved for scholarly documents where the writer is unknown. However, the contributions of indigenous peoples, who often practice the oral tradition with knowledge sharing, to culture, linguistics, botany, mathematics and numerous other disciplines, have often been described as 'anonymous' in the 'Western canon'. Enslaved people and women throughout history have published books or contributed to them, unbeknownst or under pseudonyms. Many define the reason as 'social constraints (modesty)' but they didn't have the same rights as white, upper-class men.

TABLE 2
Rationales for anonymity

1. To facilitate the flow of information
2. To obtain personal information for research
3. To encourage attention to the content of the message
4. To encourage reporting, information seeking, and self-help
5. To obtain a resource or encourage action involving illegality
6. To protect donors or those taking controversial but socially useful action
7. To protect strategic economic interests
8. To protect one's time, space, and person
9. To aid judgments based on specified criteria
10. To protect reputation and assets
11. To avoid persecution
12. To enhance rituals, games, play, and celebrations
13. To encourage experimentation and risk-taking
14. To protect personhood
15. Traditional expectations

Figure 61: Rationales for anonymity (Marx 1999)

Upper class authors, those who had the time to write and were literate, also published anonymously to protect their reputation.[127] Or the inverse, which harkens back to academic citation and the 'impact factor' from Chapter 3, with Foucault's remark that most authors wouldn't publish anonymously because they need their reputation to succeed:

> Why did I suggest that we use anonymity? Out of nostalgia for a time when, being quite unknown, what I said had some chance of being heard. With the potential reader, the surface of contact was unrippled. The effects of the book might land in unexpected places and form shapes that I had never thought of. A name makes reading too easy.
> I shall propose a game: that of the 'year without a name.' For a year, books would be published without their authors' names. The critics would have to cope with a mass of entirely anonymous books. But, now that I come to think of it, it's possible they would have nothing to do: all the authors would wait until the following year before publishing their books ... (1994:321-322).

[127] According to book historian James Raven, between 1750 and 1790, over 80 percent of all British novels were published anonymously.

This also ties into the fusion of 'search' and 'research' mentioned in the Introduction in relation to the peer review system in academia, where an academic reviews another colleague's text and provides critical feedback in the form of a written appraisal, comments and suggestions, which often determine the acceptance of the text (or its rejection) by a journal.

1.4 Pamphleteers, Cryptoanarchists & Whistleblowers

As mentioned in Chapter 6, 'while not a perfect means of anonymity, [Tor] has become widely used by journalists communicating with sources, human rights activists, and dissidents in war-torn areas such as Syria' (Borland 2013 cited by Gehl 2014:1223). Robert Gehl defines two conflicting factors regarding the Dark Web. On the one hand, there is the existing 'media ideology' —with its conduciveness to illegal activites. On the other, anonymous users—— citizens, dissidents and activists worldwide.

> First, there is the conception of the dark web as entirely composed of illegal or taboo activities and in need of policing. Second, there is the idea that the dark web can preserve a valued liberal freedom: freedom of speech (2014:1222).

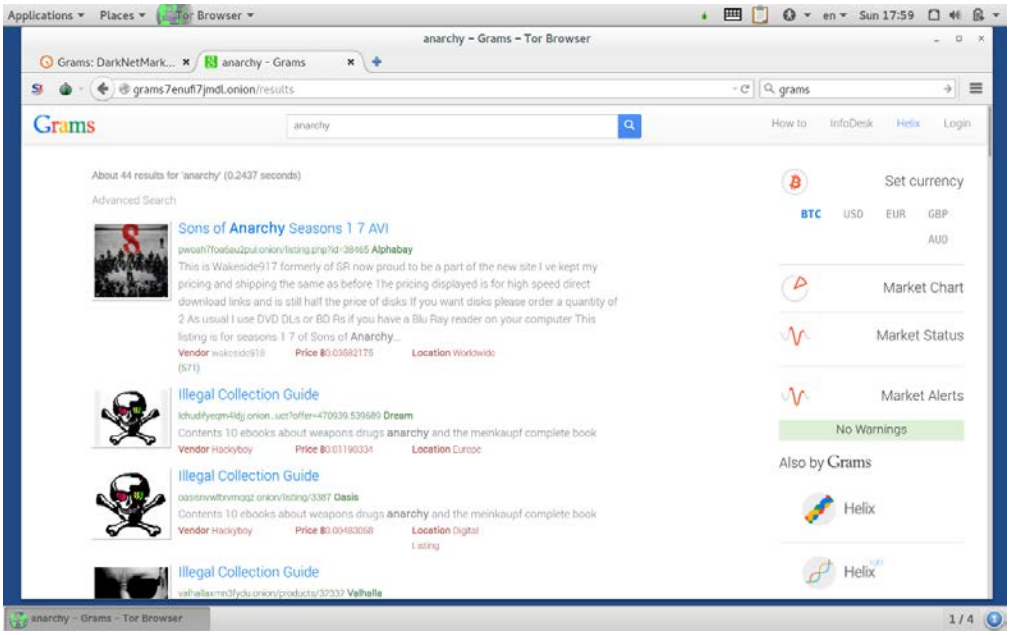


Figure 62: keyword 'anarchy' with Grams

An article from 2002, *To Speak or Not to Speak: Developing Legal Standards for Anonymous Speech on the Internet* contextualises the US First Amendment regarding 'Freedom of Speech' by questioning whether inflammatory speech by anonymous users online should be de-anonymised. It cites the US Supreme Court in Reno versus the ACLU ruling that 'observed that the principles of free speech apply to the internet and extends to protect those who use the Internet as a "soapbox" in the modern version of a "pamphleteer"'(Lipinski 2002:942). When searching with the term 'anarchy' on the Dark Net I found what could be called 'Crypto Anarchism'. My results delivered a range of URLs—the series 'Sons of Anarchy', a fictional portrayal of an outlaw motorcycle club in California and the 'illegal collection guide' that provides information about weapons, drugs and anarchy, along with a complete version of 'Mein Kampf'. (Figure 62) The next page was the traditional and expected 'Anarchy Cookbook' that included services for counterfeiting techniques, picking locks and explosives. This literature

was perhaps a bit surprising—'The People's Scrum: Agile Ideas for Revolutionary Transformation' (Figure 63), the book 'Acid Dreams: The Complete Social History of LSD', a guide for 'carding'— how to carry out credit card/payment fraud, phishing, etc. and other forms of hacking techniques. (Figure 64)

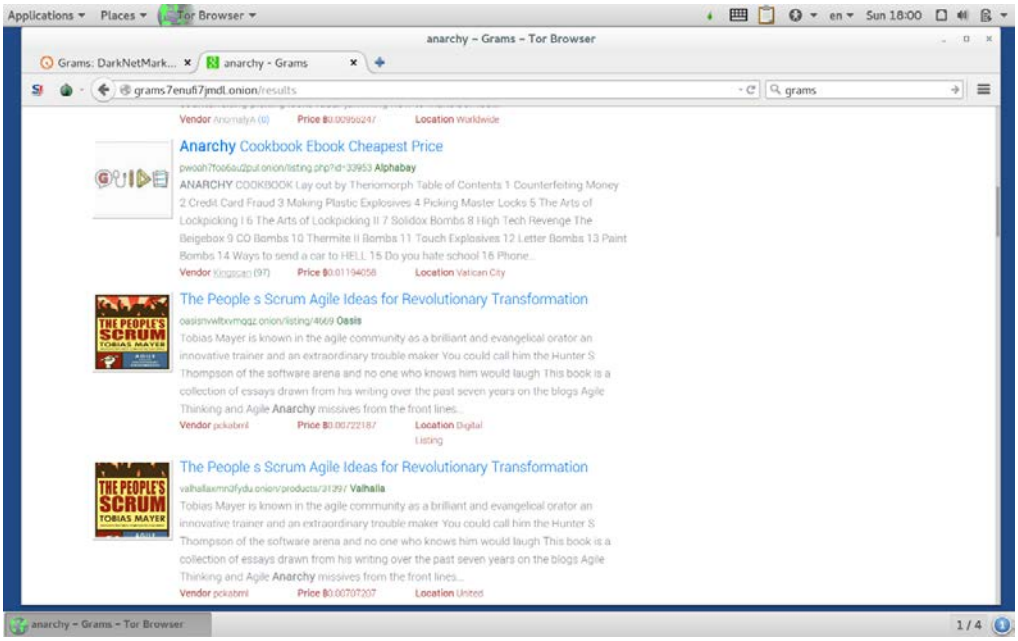


Figure 63: keyword 'anarchy' with Grams

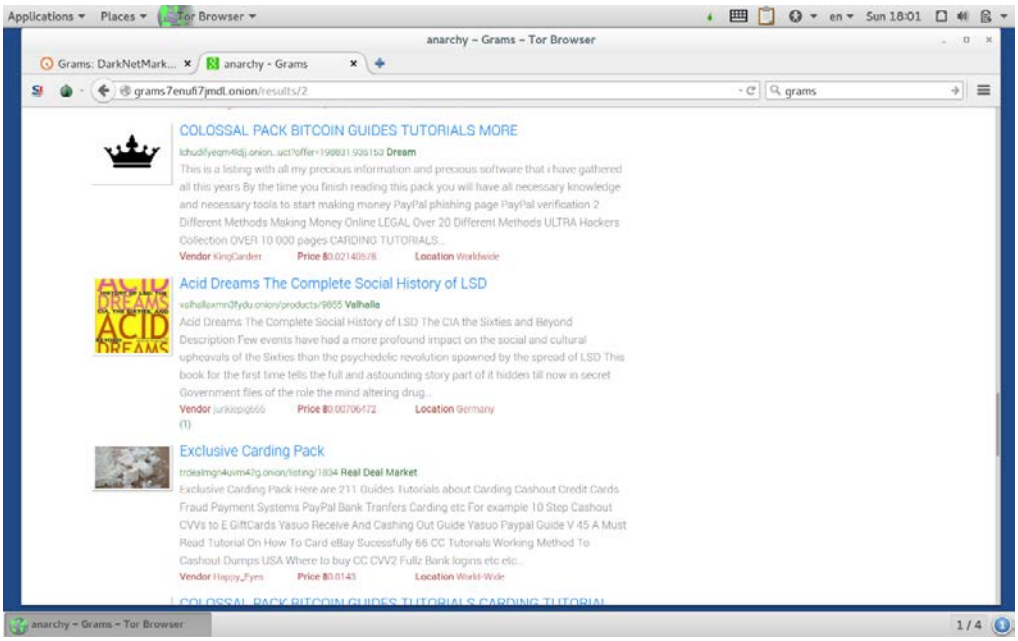


Figure 64: keyword 'anarchy' with Grams

Moreover, I found the 'Code Green' wiki, where the visitor can join in protests surrounding 'ethical hactivism for a better world', providing a range of sites that enabled further participation as a 'sympathizer, enabler, whistleblower, coder, artist or hactivist'. (Figure 65) I postulate these results show that there are anonymous users who view the Dark Net/Web as a political safe haven and who facilitate the flow of information and its dissemination within Tor Hidden Services as contemporary pamphleteers.

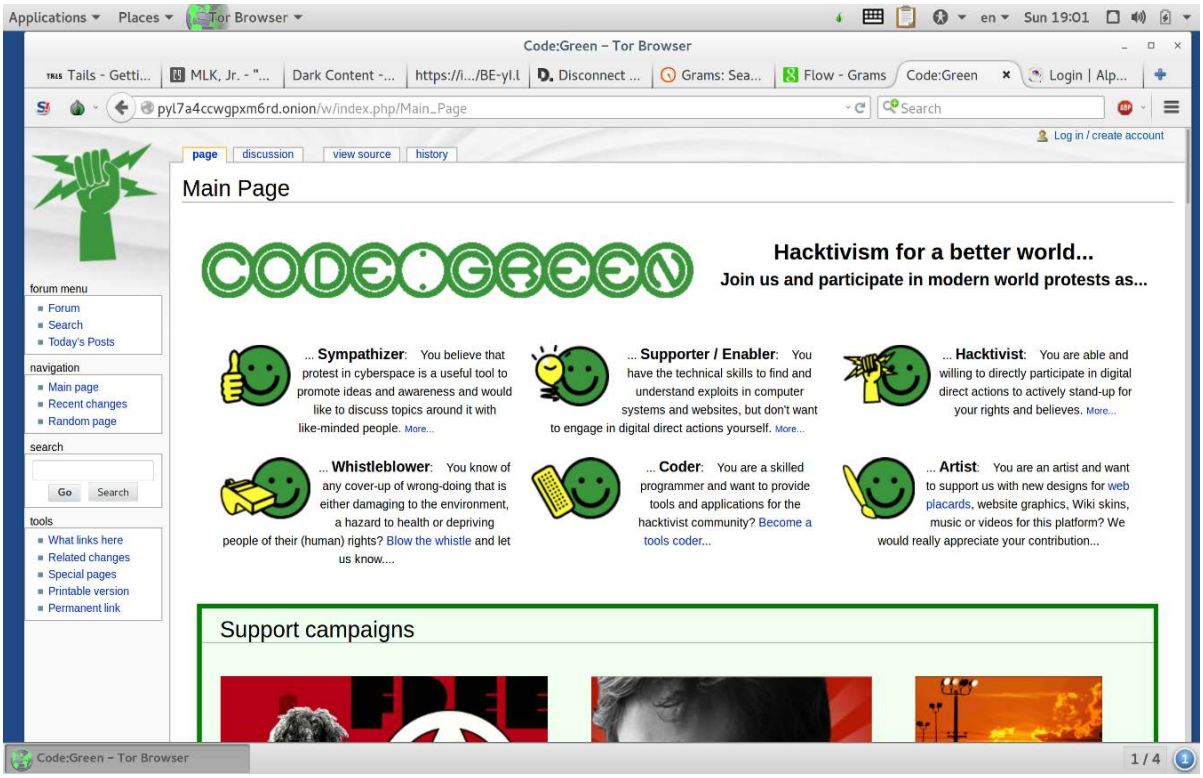


Figure 65: keyword 'anarchy' with Grams

> This little known parallel internet is a natural home for an uncensored drugs marketplace, as it is for whistleblower websites and political dissidents who also rely on its powers of obfuscation (one reason why Tor is widely considered vital for freedom of speech around the world) (Bartlett 2014b:4).

Besides those using Tor to protect their anonymity there are also whistleblower sites on the Dark Web, SecureDrop and Global Leaks and the Hidden Wiki also contains an 'entire section dedicated to whistleblowing, while almost every other resource will send you in loops looking for answers' (O'Neill 2020). (Figure 66). Wikileaks.org has an onion address that is a 'secure site where you can anonymously upload your documents to WikiLeaks editors'.[128] In 2014 the New Yorker magazine dubbed Tor Hidden Services a 'whistleblower drop box' (Bartlett 2014a:110) because Tor has been increasingly used in recent years to protect the anonymity of the user in areas of information leaks. Along with Edward Snowden's usage of Tor with his 'revelations' mentioned in the Prologue, Chelsea Manning's release in 2010 of 250,000 diplomatic documents (Cablegate) and the 'Collateral Murder' video has also showed the power of anonymity through encryption (although she later told Adrian Lamo, who in turn reported her to Army investigators).

> This ties in with a secondary thread in popular coverage of the dark web: its affordances for journalists, activists, and whistleblowers who want to speak freely, despite state monitoring of the Internet…anonymizing software such as Tor can benefit anyone who wants to dissociate speech from identity, including political dissidents' (e.g. Pagliery 2014 cited by Gehl 2014:1223).

[128] https://wikileaks.org/#submit

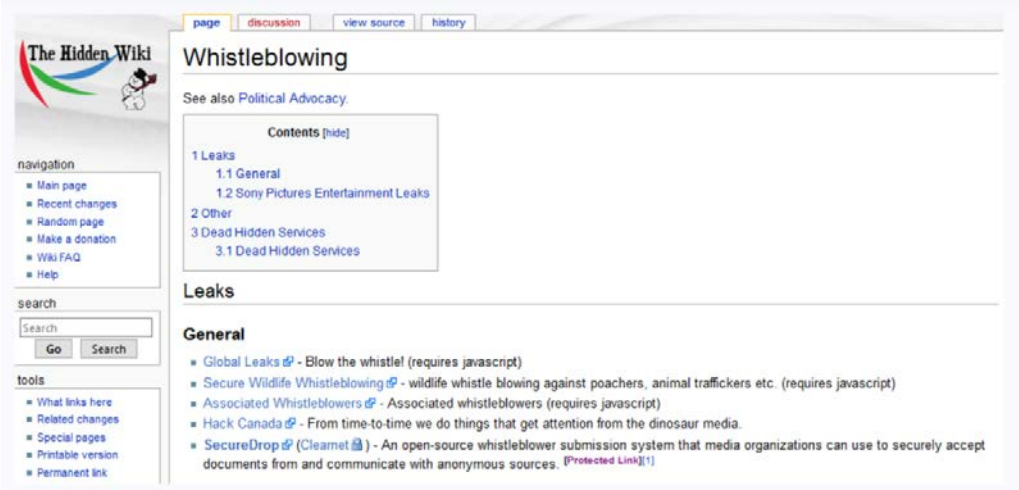


Figure 66: Whistleblowing section of Hidden Wiki

Along with the political opinions of those dissidents in oppressive regimes or even in so-called democracies, some people choose to remain anonymous. Why people wish to be anonymous is varied and, with reference to Marx's 11th rationale, users perceive certain risks if identifiability is the default setting. According to one ethnographic study by Forte et al., which explored p2p users of Tor and Wikipedia in regard to online participation, besides play, trickery or deceit, this 'namelessness' or 'anonymia' can be used for freedom of speech and privacy or to address concerns for safety in life-threatening situations (2017). Although quite comprehensive, Marx's rationale 11, 'To avoid persecution', includes reasons for anonymity such as witness protection, hotlines of communication, protection from human rights abuses, heath care issues such as AIDS testing and journalists reporting on news and media sources (1999). According to the U.S. Supreme Court, 'anonymity is a shield from the tyranny of the majority,' protecting those who disagree from retribution.[129] There are always people who wish to express 'forbidden opinions', such as the racial agitation occurring in most democratic countries with the rise of the alt-right or those fighting for civil and human rights in dictatorships.

> Anonymity networks [such as Tor] serve an important purpose on the Internet. They often provide the only means for citizens to access or distribute censored or restricted content without a threat to their privacy or even safety' (Khattak et al:1).

1.5 Postcypherpunks & Hactivists

Besides its criminal reputation (drug markets, child porn, piracy, etc.) in the media, there is a range of content on the Dark Net/Web that is rich in dissident literature, political debate and what some might consider anarchy. Reminiscent of the 'cypherpunks', who built and shared the early cryptography tools described in Chapter 6, these modern day 'postcypherpunks' have engaged once again in 'crypto-wars' with governments to obtain privacy and anonymity online. With the search engine 'Torch' and the keyword 'anarchy' I found articles on anarchist activities including squatter movements in the Netherlands. (Figure 67)

[129] 'A frequently cited 1995 Supreme Court ruling in McIntyre v. Ohio Elections Commission reads: Anonymity is a shield from the tyranny of the majority. . . . It thus exemplifies the purpose behind the Bill of Rights and of the First Amendment in particular: to protect unpopular individuals from retaliation . . . at the hand of an intolerant society'. Available here: https://www.eff.org/issues/anonymity

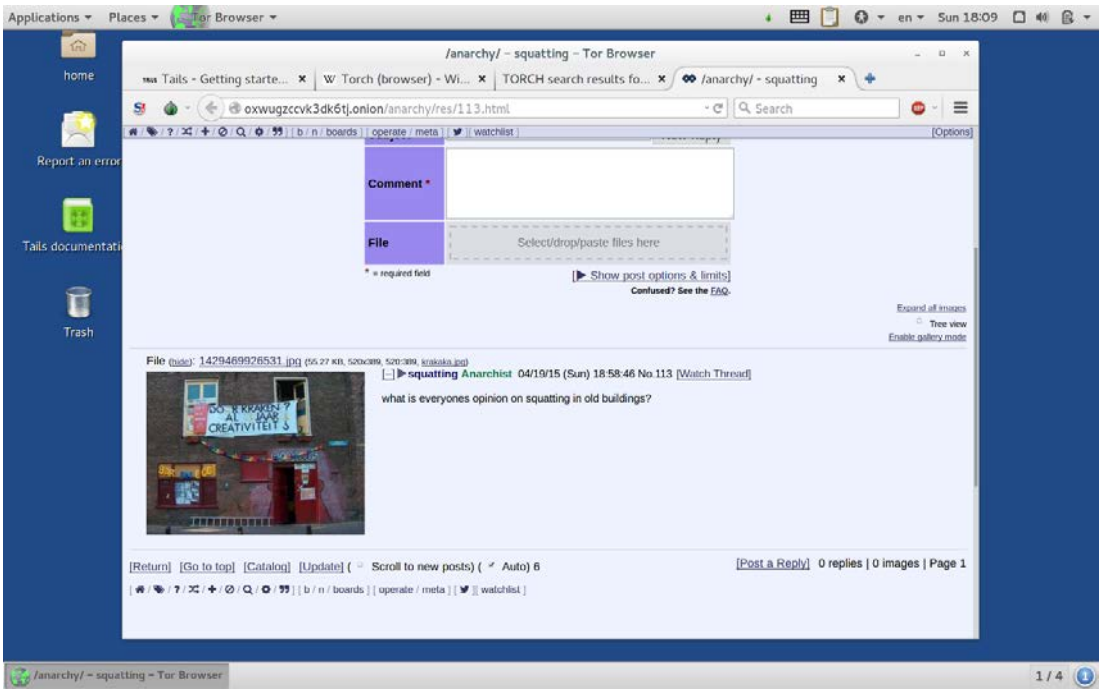


Figure 67: keyword ‘anarchy’ with Torch

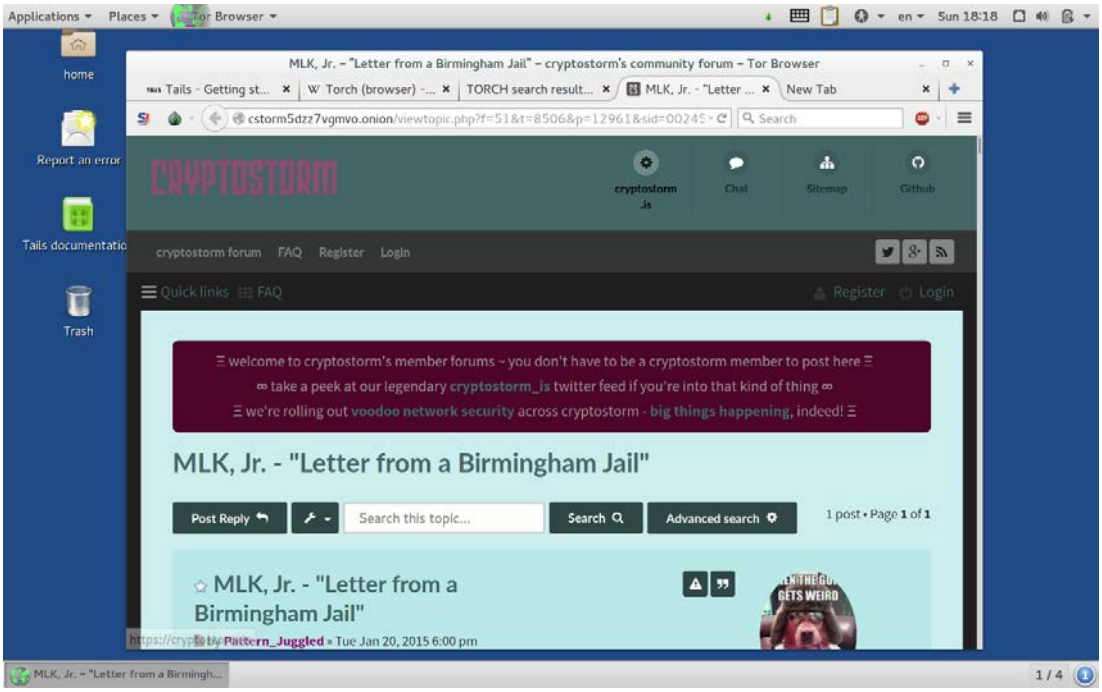


Figure 68: keyword ‘anarchy’ with Torch

Martin Luther King’s ‘Letter from a Birmingham Jail’, which I found on the ‘Cryptostorm’ forum. (Figure 68) Furthermore Torch delivered links for hacking manuals in Spanish on a Russian site (it had both Russian and English listings) and a book entitled ‘My New Gender Workbook: A Step by Step Guide to Achieving World Peace Through Gender Anarchy and Sex’. (Figure 69) Marx’s 13th ‘rationale for anonymity’ pertains to reasons of ‘experimentation and risk-taking’. Translated to the modern-day web and online participation, far-reaching definitions include whether a person is not identifiable as themselves and if they are untrackable or even ‘unreachable’ (Nissenbaum 1999).

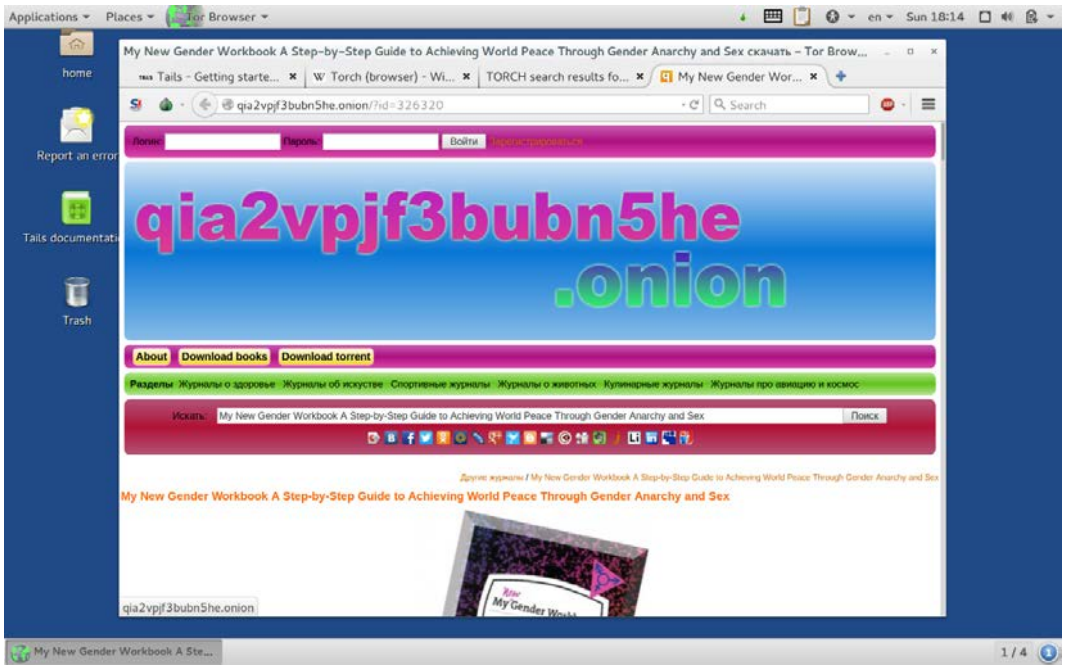


Figure 69: keyword 'anarchy' with Torch

In other words, one's personal identity, or information associated with that identity, are not known publicly. This also pertains to those who operate under a group moniker, such as 'Anonymous'. Gabrielle Coleman's in-depth six-year ethnographic study entitled *Hacker, Hoaxer, Whistleblower, Spy: The Many Faces of Anonymous* provides a comprehensive overview of the history of Anonymous and its communication strategies including but not limited to IRC (Internet Relay Chats) shown below, which are also accessible via Tor (2014). (Figure 70)

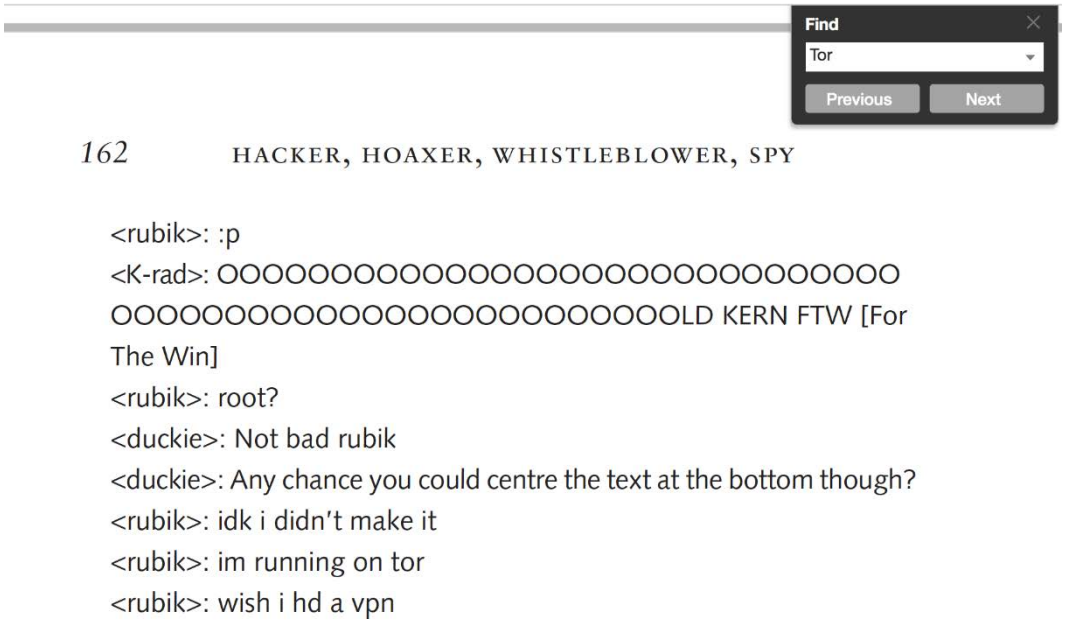


Figure 70: Gabriele Coleman's book with IRC comments about Tor and VPN (2014:162)

I received results for the 'FreeFor' chat forum with the search engine Grams that employs various means of communication tools, including IRC. (Figure 71).

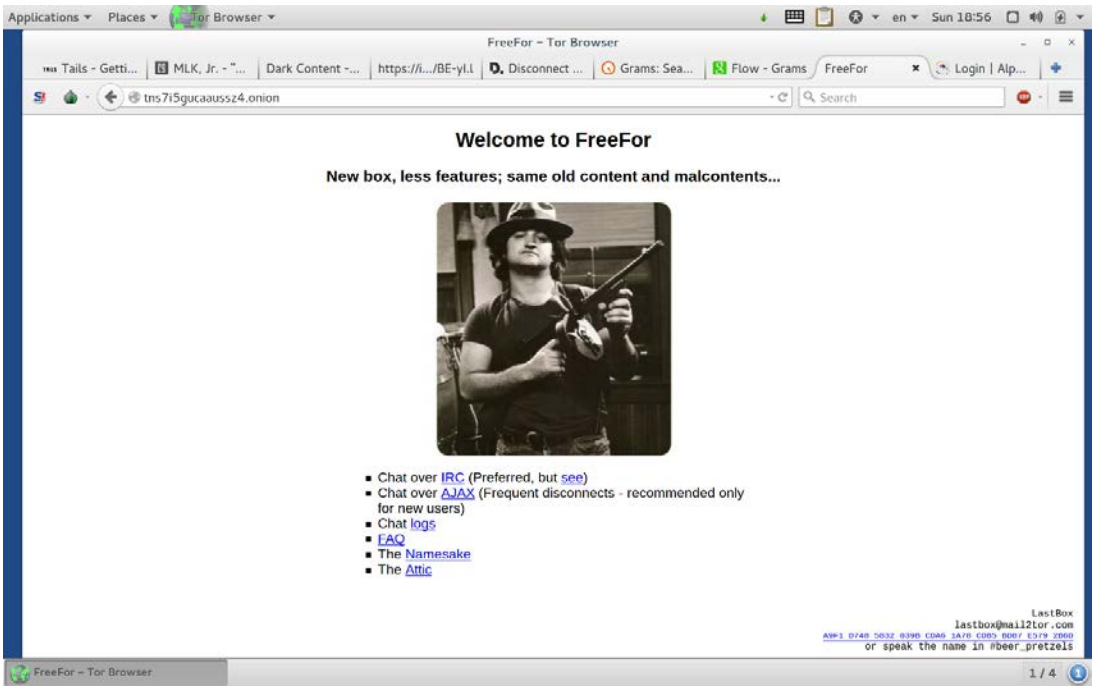


Figure 71: FreeFor forum that uses IRC

Due in part to Coleman, the Anonymous whisperer, along with subsequent interventions, the reputation of Anonymous has shifted from criminal to almost heroic in regard to some of their actions involving ISIS. 'Anonymous might be no friend to the US government: but it recently "declared war" on ISIS. In this space, governments would be wise to work with anyone willing to help' (Bartlett 2014b:12). Furthermore Bartlett's ethnographic research of the Dark Net shows that 'it was Anonymous hacktivists (and then more recently two vigilante hackers)—who helped to identify and remove Tor Hidden Service sites which were hosting child pornography' (ibid). Since 2014 there have been a range of 'operations' identified by hashtags including #Cyberprivacy, as well as the targeting of terrorists and corrupt and racist organisations, such as the 2020 hack and the edited release of over one million police files by DDoSecrets, 'BlueLeaks', in the wake of the murder of George Floyd in 2020. However, the shifting identities of Anonymous members, as well as their activities on the Dark Net, remain hidden.

1.6 Grams' Admin

Described as 'Google searching for secrets', I then applied the navigation tool whose name alludes to its function—Grams, which searches for products (not only drugs) across Dark Net marketplaces and derives its popularity with users from media reports about the Dark Net. Both the empirical study (Spitters et al. 2014) and Grams developed taxonomies of keywords for Tor Hidden Services. In 2015, when searching with Tor, I discovered the Deepdotweb, a now defunct news site that was seized by law enforcement in 2019 because it allegedly received Bitcoin for sending and posting links to the Dark Net.[130] (Figure 72) The anonymous 'admin', who is designing and programming Grams, divulged in an interview on Deepdotweb from May 3, 2014 that they have a career building websites on the clearnet to 'pay the bills', but with the discovery of the 'Dark Net' they became inspired: 'There's a whole sub-culture and hidden web based on anonymity, I decided I wanted to be on the cutting edge of making the Dark Web just

[130] When clicking on this link: https://www.deepdotweb.com/2015/11/15/university-helped-fbi-in-silk-road-2-and-child-porn-busts/ I received this image.

as useful and easy to use as the clearnet' (Anonymous 2014). Already in 2014, the admin shares how specific keywords relating to 'trending' topics were added to Grams, mimicking Google's business model as explained in Chapter 5:

> I'm conflicted by the need to monetize the site, while also creating features that will better the user experience. The next major feature, I think, will do both. I'm calling it 'Gramwords', and basing it the Google Adwords system. Now that the weighted keyword search algorithm is in place, I can really sink my teeth into this project. Vendors and Tor site owners alike will be able to purchase sponsored keyword listings for their sites and/or products (Anonymous 2014).

Figure 72: A 'domain seized' image of deepdotweb. (August 2020)

At that time, Grams was designed to have the look and feel of Google, with a Google inspired colour palette and other Google-like features including the infamous 'I Feel Lucky', which acts as a type of filter to let the user sort the results in different ways: relevance, price and date. Connecting back to Chapter 5, Grams anonymous admin offered some insight into the intentions and the developments of the search engine:

> The big feature I've been working on for a while and finally finished this week was to overhaul the search algorithm. Once I started working on an ad-words feature, I realized that the search algorithm needed to be more keyword based since the previous algorithm only searched the database using full-text method. At the same time, I realized that a lot of searches were for keywords not necessarily related to the marketplaces, which meant I needed a full Tor search engine. I first created a full Tor search engine using weighted

keywords, which analyses webpages based on the number and placement of keywords on a page. After that was done, I created a similar system to find weighted keywords in the Dark Net market listings. Once both systems were completed I merged the database systems so that users will now see Tor sites and Dark Net market listings side by side in the search results. Search results will now be more relevant, accurate and diverse. The system is not perfect yet- I will be tweaking the different weights given to keywords over the next couple weeks, to give users the best results possible (ibid).

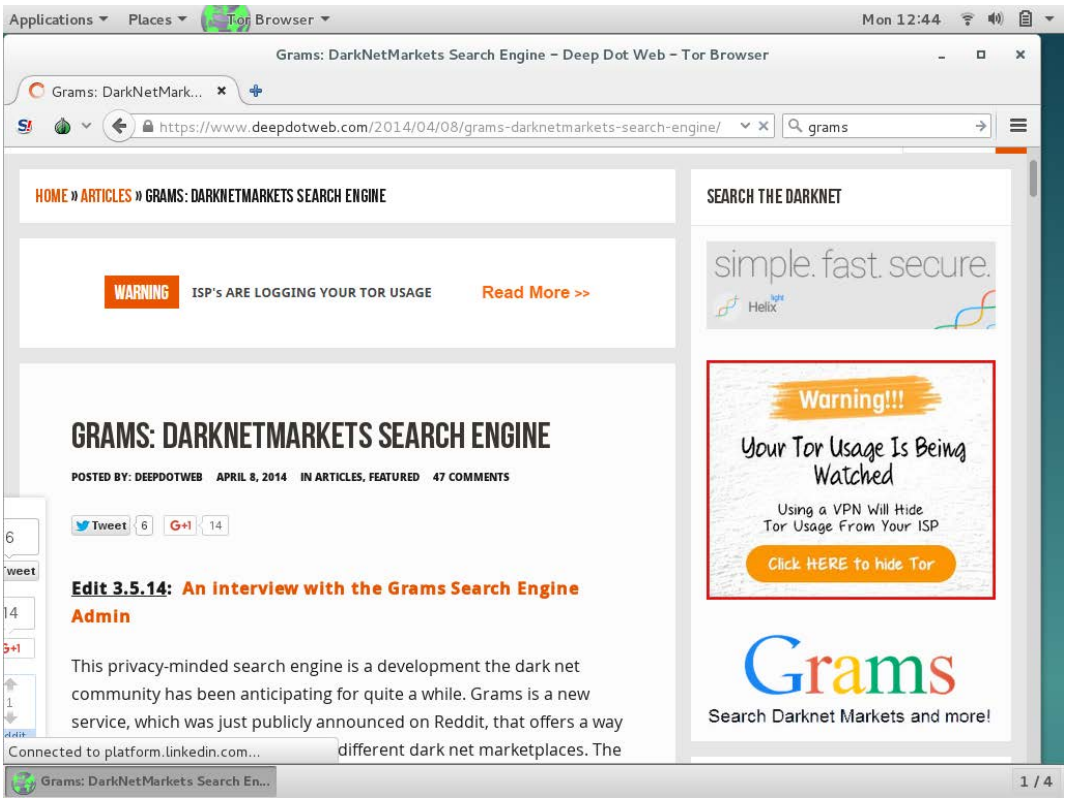


Figure 73: Interview Grams admin on the now defunct 'deepdotweb.com'. Warning: Your Tor Usage is Being Watched

When I first found the site I received a warning and the banner ad to the right, as I was not using TAILS but only Tor at that moment. (Figure 73) I then used TAILS and discovered a special feature of Grams, 'Flow', which gives an overview of specific keywords in alphabetical order that when clicked upon, access Hidden Service sites as an alternative to remembering onion addresses. (Figure 74)

When I clicked on the first keyword of Gram's Flow, 'alpha', this led me to the Alphabay site, where BitTorrent sites such as the Pirate Bay and information repositories on different ways to perpetrate credit card fraud are also prevalent, besides the drugs, guns and 'rent-a-hacker' services.[131] (Figure 75) With the keyword 'Rent-a-Hacker' I received this login. However, I was redirected to the AlphaBay validation again before I could proceed and I decided not to continue. (Figure 76)

[131] As much as Bittorrents were predominant in the earlier days of Tor, there has been much backlash inside the community as the traffic slows down Tor too much and inhibits the TorFlow.

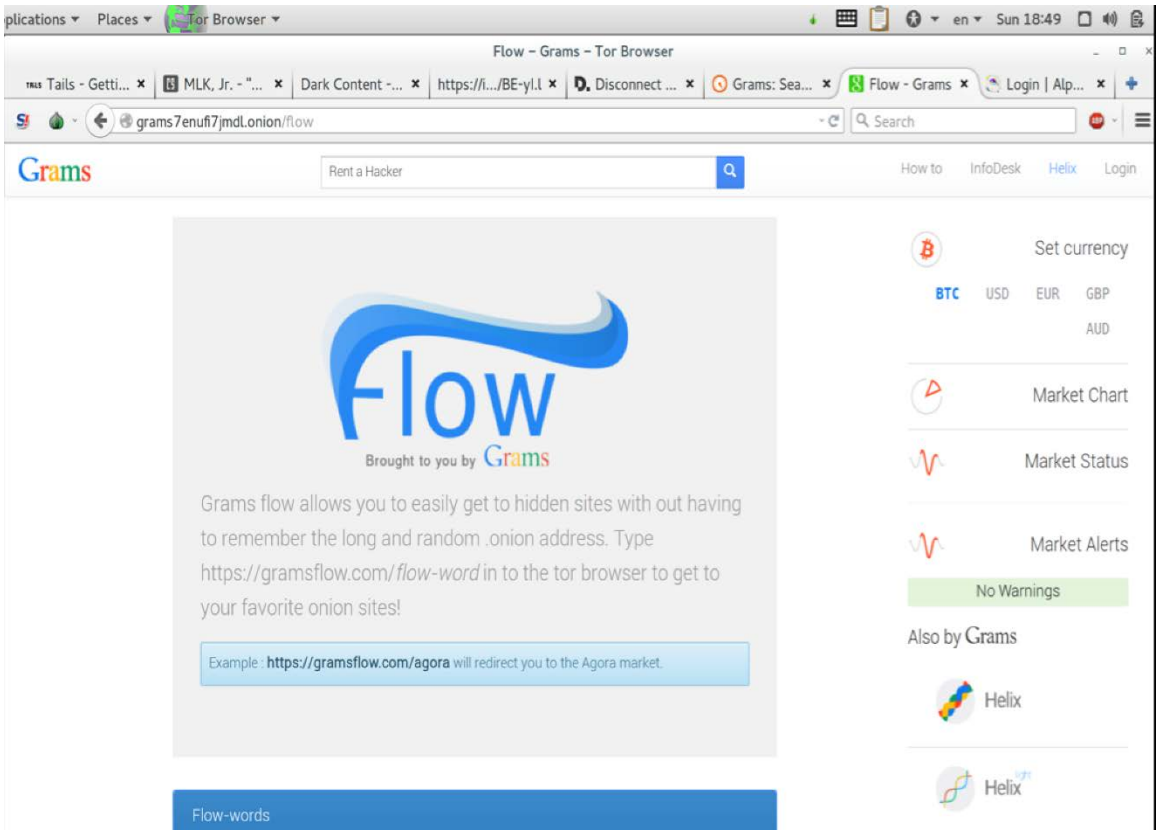


Figure 74: Flow offers keywords for the user to click on

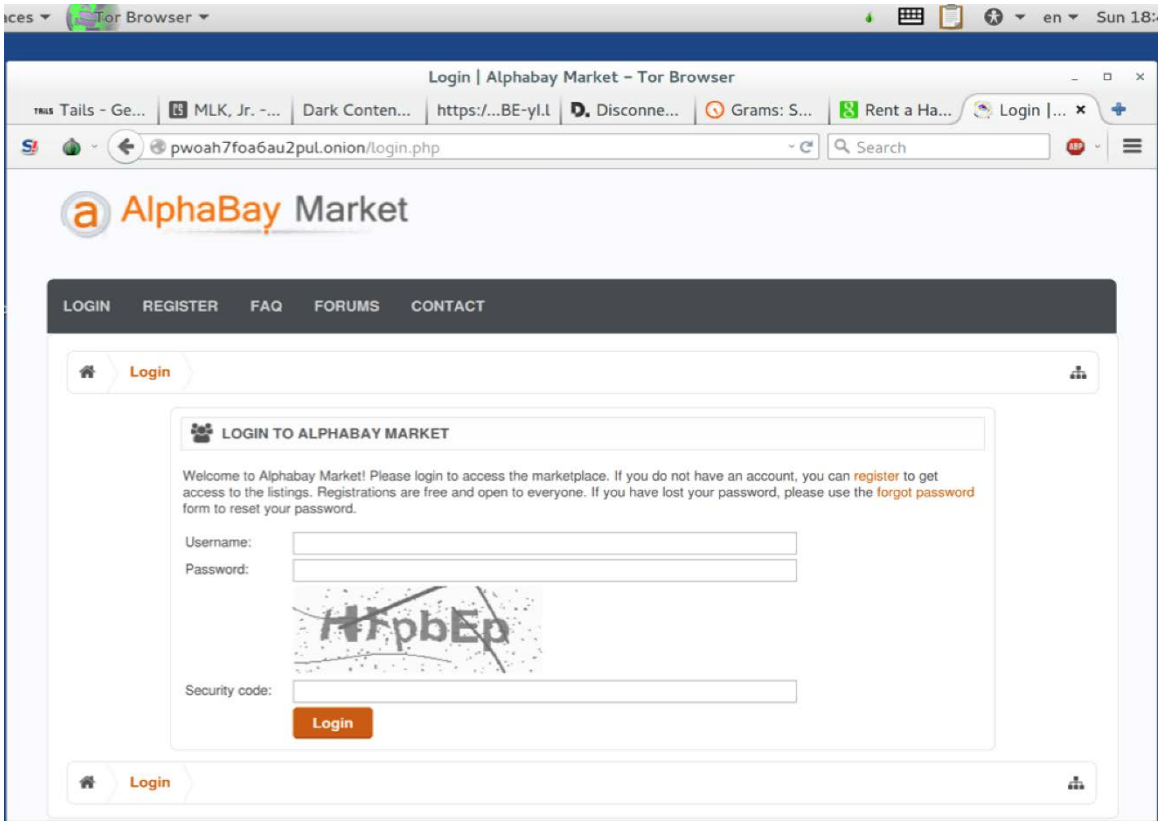


Figure 75: entrance to the AlphaBay marketplace

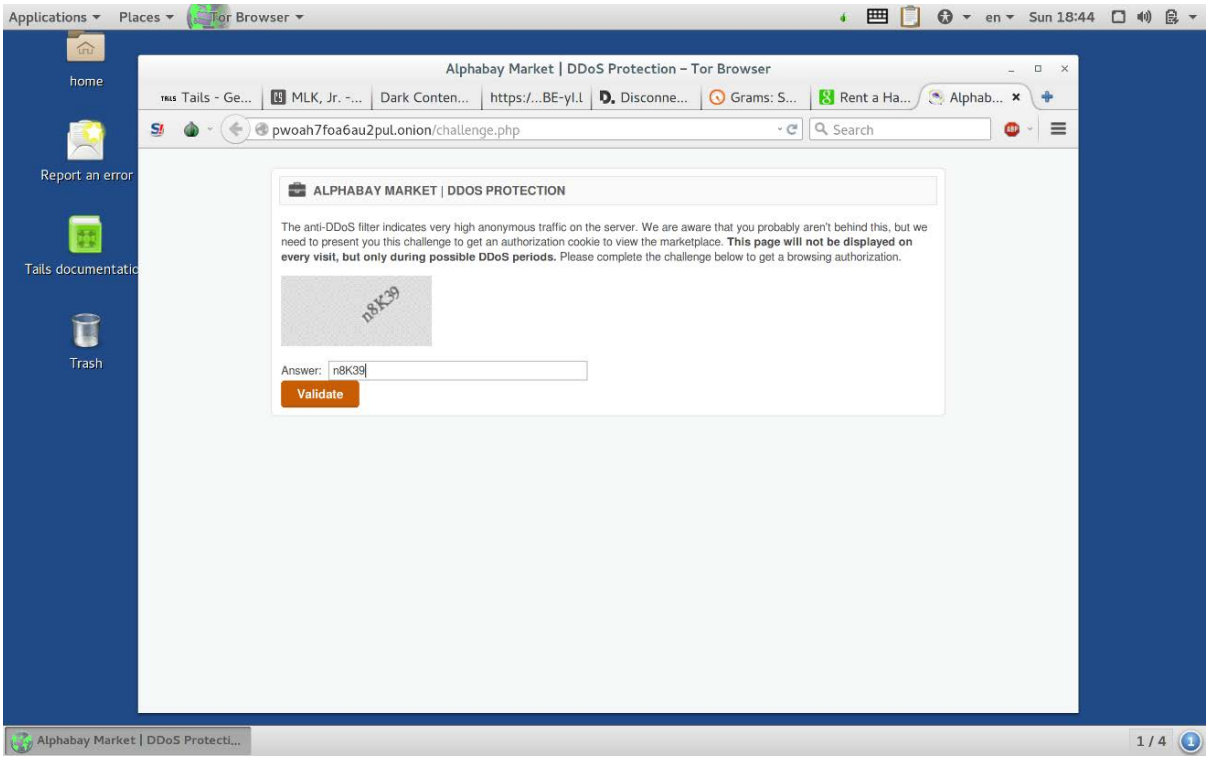


Figure 76: Rent a Hacker redirect

Although 'Grams has recently introduced trending searches and vendors can even buy sponsored keyword listings space for their sites and products' (Bartlett 2014:162), unlike Google's Adwords system, the goal is not to collect all of the users' data and direct targeted advertising. Here the user is not personalised, with little information collected about the users who are anonymous and use pseudonyms and encryption keys, as discussed in Chapter 6. The person wanting to advertise makes an ad and determines which keywords are relevant and how much they are willing to pay a day, much like 'affiliate advertising'. Users are assigned a Bitcoin address, where they need to add Bitcoins on regular basis, so the 'Gramword' system can function. Child pornography is not allowed on any banner ads, in line with Grams' terms of usage.

Also, in contrast to Google, Grams allows any Tor site to be indexed and the user has the ability to 'Add a Site' so that they can add their own site to the index. 'Tor site owners will get the best results on Grams by making sure their metatags such as title and description are accurate, just as they would SEO for a clearnet site' (Anonymous 2014). Ultimately the goal of the Grams developer is to create an up-to-date search engine that offers relevant results for the seeker, only each user does not have a Grams profile. Intentions are clear—the admin hopes to make Grams the most trusted, reliable and popular search engine on the Dark Net. Many sites on Tor's Hidden Service adopted the search engine Grams' specially developed API and there is a subreddit where community members publish feedback that keeps up to date reports on which sites are available, along with providing vendor feedback.[132] Although not having purchased anything on the Dark Net, I explored giving feedback, by reading comments on the Grams subreddit, such as this one:

[132] https://www.reddit.com/r/onions/comments/22z3qe/grams_beta_version/

> Hello, Grams scrape feedbacks from the markets sir, you must sign up to your market, buy some stuff and then place a feedback before you will see it on Grams. If you want to add a Grams review, you must search for your vendor on the InfoDesk, then click on the 'review' button and place it. That's how it works :). (Figure 77)

The numbers in parentheses shown in the image are reviews, not sales and Grams plans to have a vendor directory on the site, where reviews will be collated from different markets.[133] Updating the site every three days, they added new features to the search algorithm to give the user the best results possible, such as a 'scoring system' based on 'keywords, number of transactions, good reviews, and the number of clicks from our site' (Anonymous 2014). The Grams anonymous admin also planned to add other features for discrimination such as price, country and market. When this interview was conducted, sites like Agora, Pandora and Silk Road 2 were still up and running and they were looking to integrate more sites into the search engine.

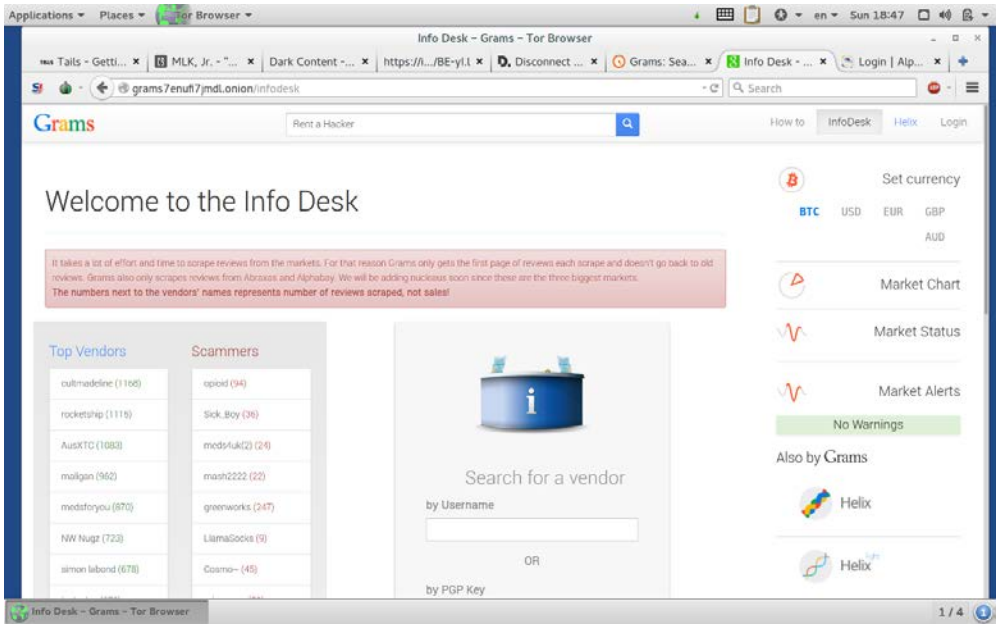


Figure 77: Grams Info Desk

1.7 Platform Criminals

According to 'internet buzz' and the mainstream media, the Dark Net, especially after the takedown of the Silk Road in October 2013 and the subsequent instantiation of Silk Road 2.0 in November 2013, became, by repute, a hive of criminality. In his eponymous book, *The Dark Net*, ethnographer and media theorist Jamie Bartlett states that he went 'in search of them' (Bartlett 2014b:6) in order to demystify the hysteria in the media. He decided to search its secrets, or 'black arts' using the Tor browser and wrote about these invisible and unseen users. In his ethnographic study of a year 'living on the Dark Net' he elucidates his encounters—from drugs peddlers' concerns for buyers being careful on the Silk Road, to anarchist hackers promoting literature but also with women who charge viewers to watch them dance, strip or have sex online. Bartlett surmised that this alternative internet is for those not wishing to be found by mainstream 'clear netters' and these underworlds are not,

[133] https://www.reddit.com/r/DarkNetMarkets

however, buried that deep because ‘cyberspace doesn’t have depth’ (ibid). Yet his caveat is to be careful, cautious and responsible when exploring it as he also relates meeting people who buy guns, illicit drugs or solicit illegal services with their Bitcoin (Bartlett 2014a:240).

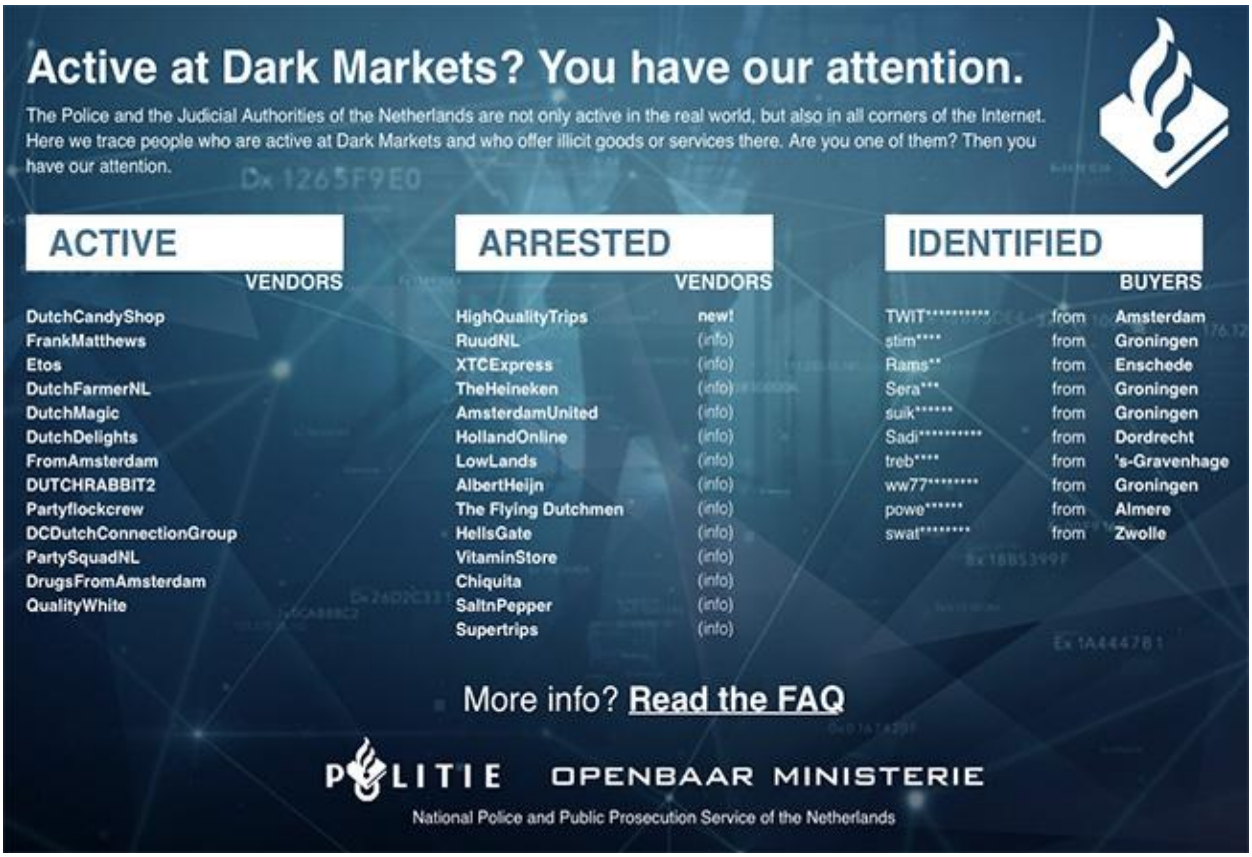


Figure 78: Dutch Police: Active at Dark Markets? You have our attention (2018)

The former Silk Road 2.0 traded in many forms of drugs, marijuana and cocaine ‘were the most popular items but customers could order almost anything, from ecstasy to magic mushrooms, from OxyContin to Valium’ (Glenny 2015:3). When searching with Grams I had access to these Darknet markets, also called cryptomarkets, which rely on anonymising technology and electronic currencies to facilitate licit and illicit transactions among participants in relative anonymity (Dittus et al. 2018:217, 278). As mentioned in Chapter 5, Platform Capitalism is changing the nature of markets through multi-sided platforms where ‘[t]heir business models typically involve rent-seeking in the form of transaction fees, and other forms of profit derived from the hosted transactions and resulting data trails’ (Srnicek 2016:21 cited by ibid). Darknet markets are ‘two-sided markets that operate as Tor Hidden Services, connecting large numbers of vendors to large numbers of buyers, offering a wide range of licit and illicit products that are catalogued in a structured manner comparable to e-commerce platforms’ (ibid:280). In recent years, on Silk Road, police have infiltrated sites and certain markets (Figure 78) as with the site Hansa, which was an eponymous keyword on Grams’ Flow before it was taken down.[134] (Figure 79)

[134] Dutch law enforcement took over the Hansa site in 2017, which led to the arrest and prosecution of many users. ‘While the live Hansa site was protected by Tor, the development server had somehow been exposed online, where the security firm discovered it and recorded its IP address’ (Greenberg 2018).

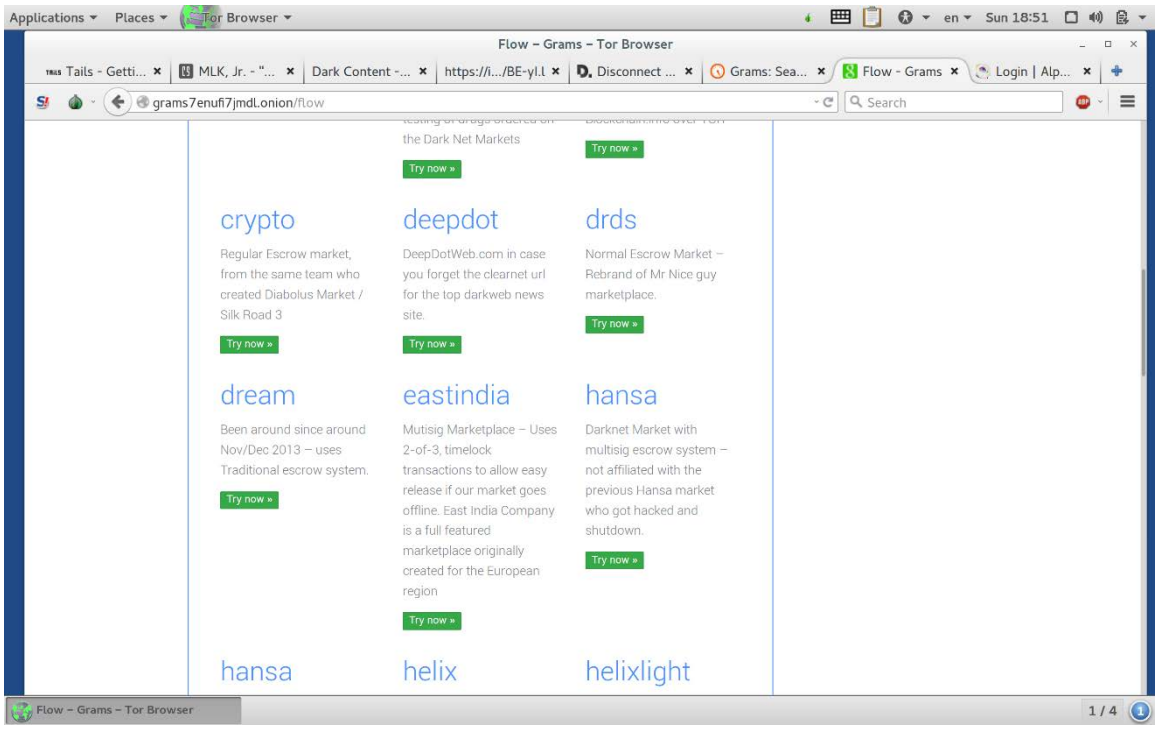


Figure 79: keyword ‘hansa’ with Grams’ Flow

When I clicked on the keyword ‘silkroaddrugs.org’—I received and advert to sign up for the ‘Anonymity’ newsletter, a type of journalism that publishes on the market places. (Figure 80)

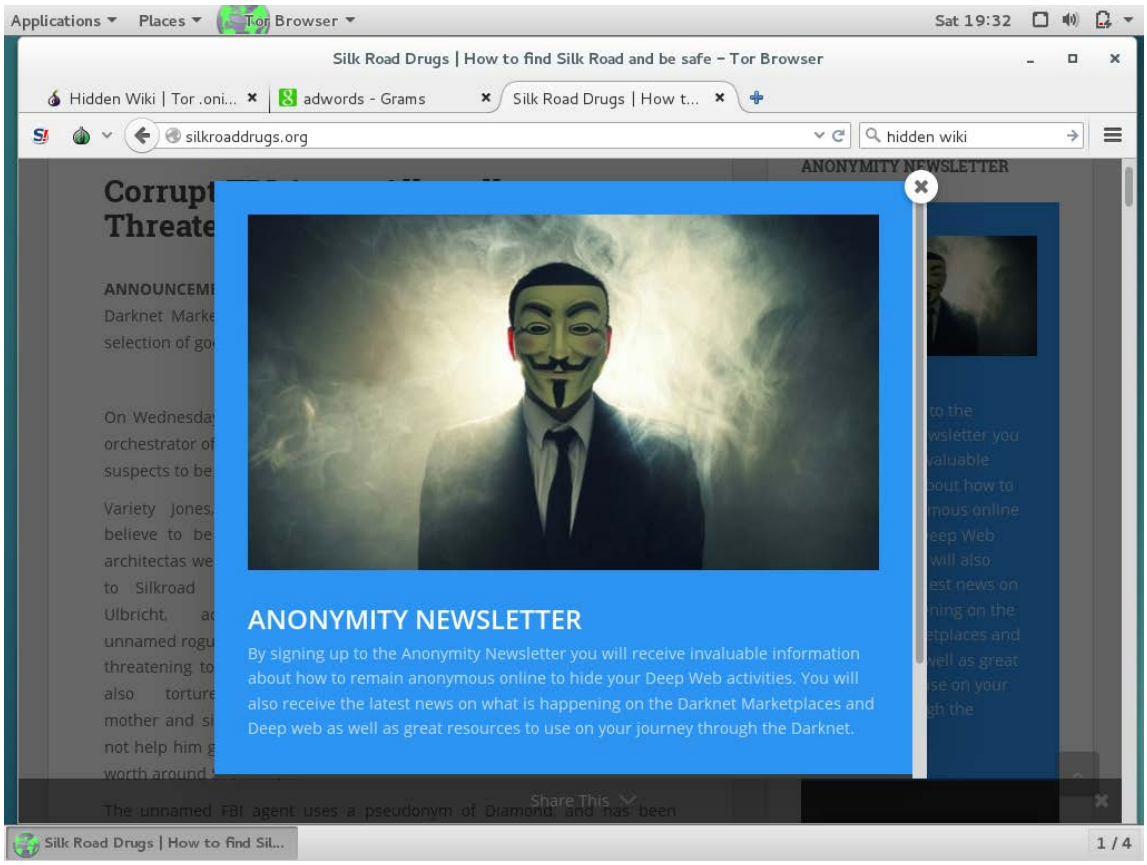


Figure 80: keyword ‘silkroaddrugs’ with Grams’ Flow

The empirical study *Platform Criminalism: The 'Last-Mile' Geography of the Darknet Market Supply Chain* analyses the flow of the online drug market supply chain through the lens of economic geography and the relationships between producers, sellers and buyers worldwide. These Dark Net markets are accessed by users searching with the Tor Browser that hides the user's IP address and blocks locative data, thereby revealing 'the ways in which a layer of the internet specifically designed to conceal geography might in turn alter material economic geographies' (ibid:277). In their conclusion, the authors state that the vendors of cannabis, cocaine and opiates are located in the consumer countries, not the producer countries and that Dark Net purchasing occurs in the countries of consumption, not production, possibly 'leaving the old trafficking routes intact'. Their findings 'suggest that the geography of darknet market trades is primarily driven by existing consumer demand, rather than new demand fostered by these markets' (ibid:285). Furthermore, the study delivers a conceptual contribution to the literature on 'platform capitalism' through their analysis of economic geography by situating 'darknet markets in existing platform discourse', yet showing illicit supply chains.[135]

Their outcomes support the media reputation of the Dark Net and the Silk Road, in particular, as an illicit drug economy, however, the reputation economy concerning customer service is contradictory. Many consider the former Silk Road to be the eBay of the Dark Net (Bartlett 2014b:6; Glenny 2015:3; Dittus et al. 2018:278), where 'the site administrators don't deliver the goods themselves but act as market facilitators between a buyer and a seller, getting a cut from every transaction' (Glenny 2015:3). At eBay and Amazon, customers write recommendations that determine whether other buyers will decide to frequent the services of a dealer. The same could be said of the Silk Road as the Dark Net economic model with my discovery of the Info Desk and Grams subreddit. En route through his own ethnographic excursions, Bartlett discovered the vendor review system and the Dark Net secret:

> The most surprising statistics about the Silk Road 2.0 is not the amount of available drugs (although that is truly staggering); it's the satisfaction scores (2014b:6).

Reflected in the 95% customer satisfaction reports, the reputation economy of the Silk Road supports the theory that 'the real secret of dark net markets is good customer service' (Bartlett 2014:165). Analogous to Google PageRank that determines which websites are more frequently visited, based on ranking through authority (Chapter 3) and advertising (Chapter 5), the reputation economy of the Silk Road enabled its expansion and hegemony, not only its discoverability through Dark Net search engines. 'Once a site like Silk Road acquires a reputation for reliability it tends to grow rapidly' (Glenny 2015:3), exhibiting features of the 'network effect'. This recommendation system on the Dark Net is also analogous to Google's personalisation and sorting people into groups of others 'like them' (Chun 2016) from Chapter 5—yet vendors (and buyers) can switch identities through pseudonymity and onion addresses and there is fierce competition, like any capitalistic market. Trust also plays a salient role in this reputation and recommendation economy.

> These networked relations were found to be shaped by preferential attachment, and a need to develop trusted relationships among anonymous participants. Such structures allow for a diffusion of commerce that is more amendable to new entrants, including

[135] 'We further acknowledge the support of Google for funding the "Economic Geographies of the Darknet" project at the Oxford Internet Institute' (Dittus et al. 2018:285).

bulk traders, potentially supporting subsequent redistribution elsewhere (Dittus et al. 2018:279).

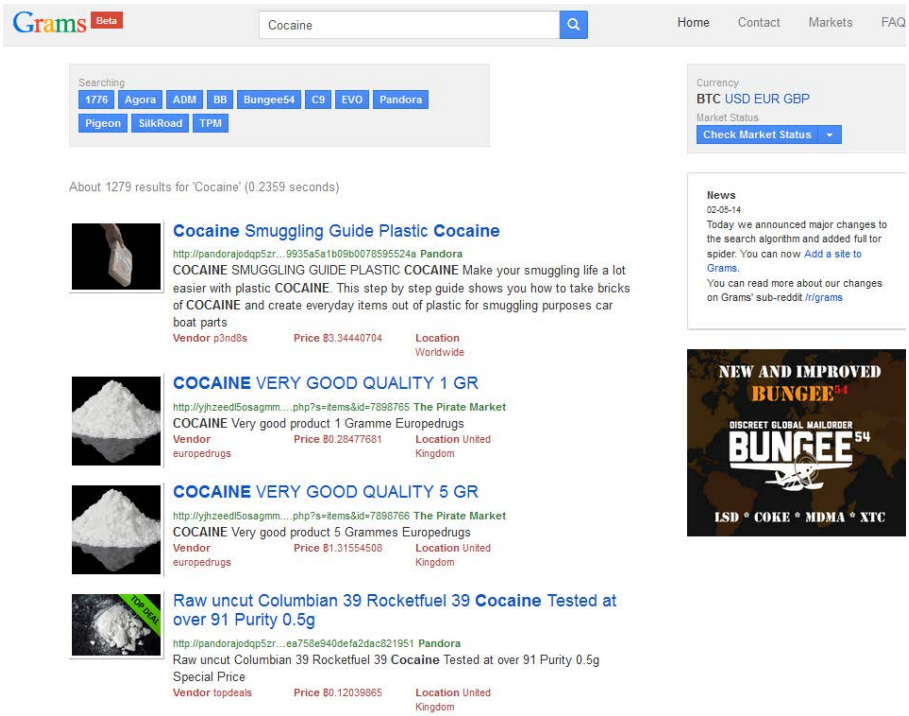


Figure 81: A list of sites with reviews and recommendations (2014).

The reputation economy of the sites increased as well as the number of users who had the tech-savvy to access the site and then became 'happy customers', expressed by their reviews (Bartlett 2014b:6) as shown by the results from a search on Grams with the keyword 'cocaine', along with the announcement of major changes to the search algorithm (May 2, 2014). (Figure 81) 'The user-ranking system provides a safer, systematic and reliable way of determining the quality and purity of the product' (ibid). These platforms put strangers in touch with each other anonymously, who might never have met in the 'physical' world and, if they did, other things might happen.

> The perceived benefits include a reduction of physical danger compared to street trades, increased product quality at cheaper prices, speedy delivery, and the convenience of varied and well-presented offerings from the same source (Dittus et al. 2018:285).

With the help of specialised search engines for Tor Hidden Services, this 'novel form of [drug] retail distribution in consumer countries' offers potentially less risk for both buyers and sellers (ibid).

1.8 The Torist

Returning to Gary Marx once again, his 7th 'rationale of anonymity' describes how the individual needs to protect 'strategic economic interests' with the ability to pay with cash instead of a frequent shopper programme that collects bonus points and behavioural information about consumption habits (1999:103). Nowadays on the Dark Net users must have a cryptocurrency account (mostly Bitcoin). Moreover, it is the 'behaviour as a consumer' in the

realm of marketing, where ‘purchases should be revealed at their [user] discretion and not electronically taken from them’ (ibid:104) that has become contentious. In an era of constant tracking by corporations, digital payments are being promoted, often deeming those who pay anonymously with cash as a criminal with ‘something to hide’. Furthermore, Marx’s rationale ‘to protect personhood’ (14th) could be applied to the present condition of not being unidentifiable to ISPs (if possible) or to the government, or to Tor exit node operators. This user group includes myself, yet I also had an expectation of what anonymity (and to a certain degree pseudonymity) could offer me as a (re)searcher on the Dark Net. I attempted to ‘reimagine search’ through anonymous surfing and to experience the serendipity of finding alternative results, which I did, on Tor Hidden Services.

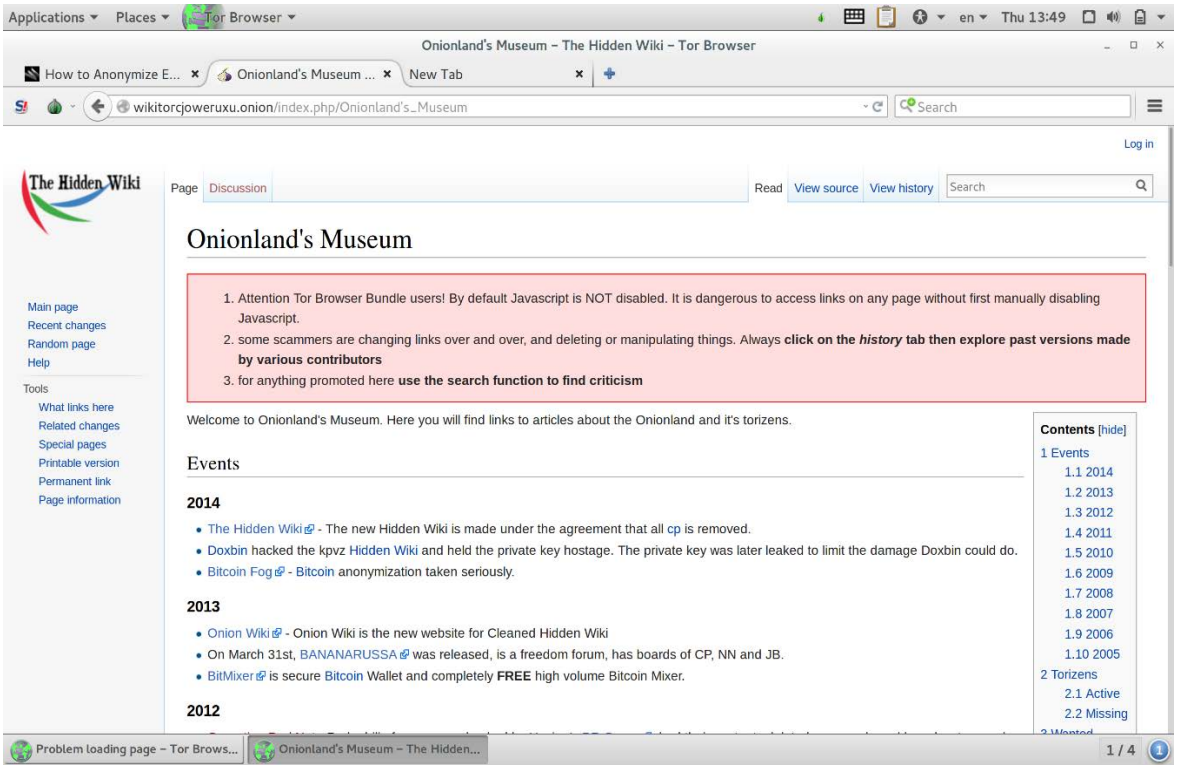


Figure 82: Onionland’s (Tor) Museum

According to Hugill et al., anonymity is what forms the basis for ‘the creative process’ that ‘normally involves a move from the known to the unknown, and sometimes from the named to the unnamed’ (2013:238). This echoes Bartlett’s findings on the Dark Net as a ‘huge, unchartered territory of cyberspace’ that ‘fosters breathtaking creativity’ (2014b:5). In addition to finding writings on anonymity, or guidebooks on how to be anonymous it is also possible to obtain material on how to attain anonymity by not being tracked, not being served targeted advertising and how to exist outside of the purview of Google. Bartlett’s research confirms what I was in search of…and to a certain extent found—there are sections of the Silk Road where not just drugs but ‘alcohol, art, counterfeit, and books!!!’ (2014a:147) are openly traded.[136] However, I was never a ‘happy customer’ as it was not my intention to purchase anything on Silk Road, only to use the search engines on Tor Hidden Services.

[136] He obtained this data by the following method: ‘We have some data because a clever computer programme harvested the details of 120,000 sales that had passed through the site over a 90-day period from January-April 2014’ (Bartlett 2014:47).

Most of my exploits in this region could be described as 'tourism', I located certain landmarks in order to navigate the Dark Net but I didn't linger at any one destination. Moreover, the fact that because 'onion services are private by default', users such as myself had to 'discover these sites organically, rather than with a search engine' (Winter et al. 2018). Although I failed to find the DWSN, I did find the *Onionland Museum* on a Hidden Wiki, reaching out to 'Torizens' like myself. (Figure 82) While conducting research I also received an onion address from the nettime mailing list, to submit texts to the Torist. This merger of Tor and tourist is a

> new online journal, planned for a biannual release, which is dedicated to deepening the relationship between technology and the humanities. In particular, we are interested in how electronic communication intersects with communities (artistic, political, and beyond) (The Torist 2015).

The *Torist* is not explicitly focused on receiving texts about the Dark Net per se, or the critique of massive data gathering by the NSA or corporations but rather solicits submissions of artworks, net art, poetry and fiction. The Admin also explicitly states that you do not have to use your real name, or a full name and that any identity is acceptable—but if you do not give your name you will be labelled as 'Anonymous'. (Figure 83)

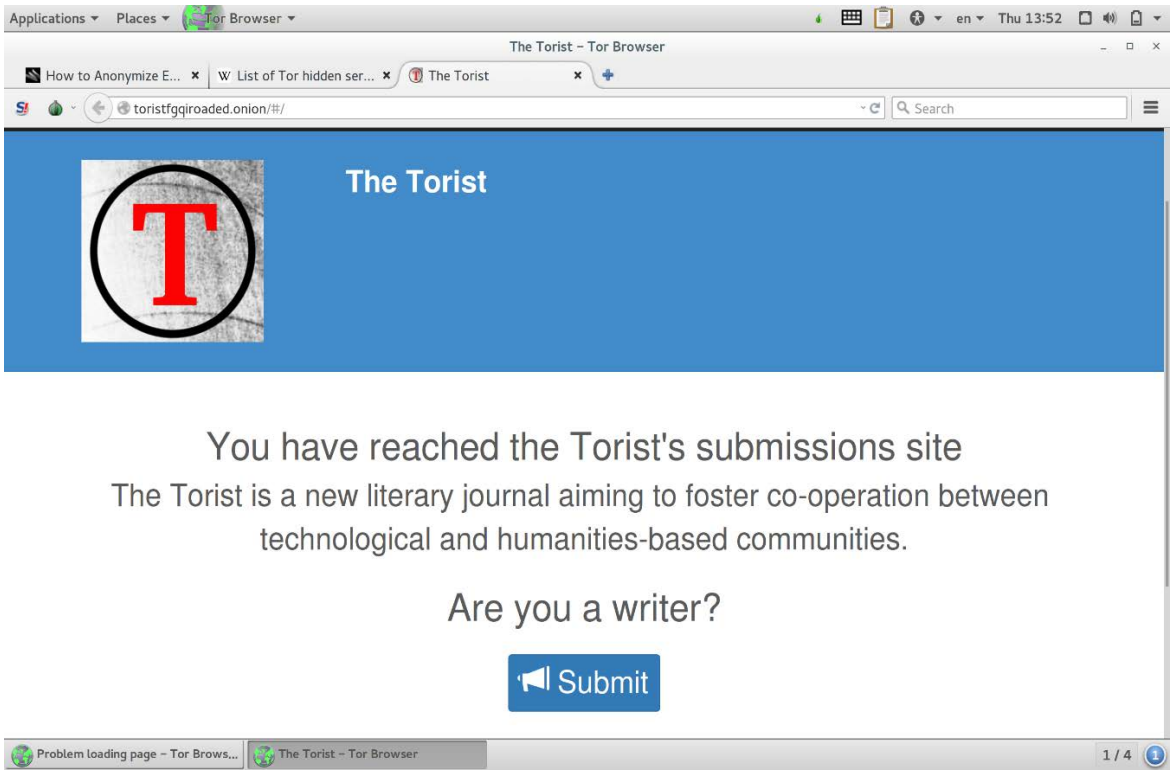


Figure 83: Torist's submission site

2.0 Deanonymised Users

In this section I first discuss 'hacks' carried out on Tor by various actors (governmental, corporate, academic). I then focus on the questionable collection of a Dark Net user's IP address without a warrant in regard to the US 4th Amendment's 'search and seizure', along with users' 'expectation of privacy' when using Tor.

2.1 Crack for sale, no, not that kind

Sometimes fiction imagines the future, other times it attempts to explain the complicated technical consequences of real-world scenarios to its audience. In the opening episode of *Mr. Robot*, "1.0hellofriend", the protagonist, Elliot, reveals to Ron, the owner of the coffeeshop in which they are sitting, that he has hacked him.

"I started intercepting on the traffic on the network, that's when I started to notice something strange. That's when I decided to hack you. I know you run a website called 'playdoo's boys'. Ron first responds with "Pardon me?" Elliot continues. "You are using Tor in the networking to keep the servers anonymous, you made it really hard for anyone to see it. But I saw it. The onion routing protocol, it's not as anonymous as you think it is. Who ever is in control of the exit nodes is also in control of the traffic which makes me, the one in control." Ron interjects, "I must ask you to kindly leave." Elliot ignores his request. "I have everything, all your emails, all your files, all your 'pics'. Ron begins to loose his temper, "Get out of here right now, or I will call…." Eliot finishes his sentence: "The police? You want them to find out about the 100 terabytes of child pornography you serve to your 400,000 users?"

As Eliot walks out the door, is it clear to the viewer that the police will apprehend Ron. Although this is an inaccurate, fictionalised account of how hackers could hack Tor, it serves to illustrate how the reputation of Tor had reached the mainstream, airing on major channels worldwide. Compared to this portrayal of the fictional figure of Elliot, I consulted my online, personalised YouTube feed self-declared hacker, 'Eli the computer guy'. Eli used to work for the US government as a security specialist and has gone AWOL, so to speak, with his own channel supported by numerous followers that enhance his reputational economy. In a nutshell, Eli expresses his position in this way: 'All hackers like to hack hackers. Real hackers hack the hell out of other hackers, it's just kinda what they do' (Eli the Computer Guy 2015).

In one video Eli the Computer Guy describes how various actors, government officials, secret service, or even spammers attempt to compromise Tor exit relays. Specifically, it should not be discounted that Tor exit node operators have access to the traffic going through their exit nodes, whoever they are (Çalışkan et al. 2015:19). Eli explains that if the NSA controls enough of the exit nodes, which they ostensibly have the resources to do, they only need to pay 1,000,000 dollars to rent 10,000 exit nodes for one month, thereby capturing and logging all traffic exiting the nodes, even if they were encrypted. Hypothetically, if the government had control of enough of these relays, they could then match the IP addresses of those exiting Tor into the 'clear web' with help from ISPs or could potentially insert 'malware' into the user's computers as spyware. Or they could implement a man-in-the-middle-attack. There have been reported hacking attempts where a group of Tor relays were 'actively trying to break the anonymity of users by making changes to the Tor protocol headers associated with their traffic over the network' (ibid).

Furthermore, Snowden documents revealed that the NSA, as well as GCHQ, planned to influence the development of Tor, along with a document showing a NSA 'proof-of-concept' attack on Tor. The NSA was 'secretly operating computer nodes in the Tor network, but the success of that effort was "negligible" because the agency has access to few nodes' (Gross 2013). The director of the Tor project, Roger Dingledine, retorted in an email in reaction to the attacks, that it was 'good news' that the NSA attempted to attack Tor, which means that 'there's no indication they can break the Tor protocol or do traffic analysis on the Tor network' (ibid).

As shown previously, there is controversy surrounding the Tor network, connecting it to the so-called 'Dark Net' with its 'hidden services' that range from the selling of illegal drugs, weapons and child pornography to sites of anarchism, hacktivism and politics (Spitters et al. 2014:1). Besides governmental agencies, there are company actors attempting to crack Tor that in turn sells exploits to governments worldwide. The previously mentioned dubious company *Hacking Team* has been known to develop malware, which can be placed on the ISP of the suspect that exposes hidden traffic and their business model in this burgeoning market, which includes offering bounties for 'zero-day exploits'.[137]

> The Italian organisation, which even its CEO called a "notorious" provider of government spyware, was looking to expand its line of products…That included targeting the anonymizing Tor network, where civil rights activists, researchers, pedophiles and drug dealers alike try to hide from the global surveillance complex (Brewster 2015).

In 2015 the company *Zerodium* extended a bounty of up to 30,000 dollars a day to researchers to expose Tor's vulnerabilities and the founder, Chaouki Bekrar, states that attacks targeting Tor nodes and de-anonymising Dark Net users 'are the holy grail of exploits for government agencies in charge of criminal investigations' (ibid). These are not the first monetary rewards offered for cracking Tor's anonymity that have raised questions regarding collusion (Çalışkan et al. 2015:18). In 2015 Russia offered a significant monetary award ($ 110000) to challenge the anonymity of the Tor network. As noted by academic authors (AlSabah et al. 2012, A. Biryukov et al. 2013, Spitters et al. 2014, Chakravarty et al. 2014, Winter et al. 2014, Çalışkan et al. 2015), various empirical studies have shown that different actors can compromise Tor anonymity in numerous ways. Besides governmental agencies, actors in the security industries and companies in the zero-day exploit business, there are also academics carrying out research that attempts to 'hack' Tor.

2.2 Looking for the IP address of a Tor user? Not a problem

In early 2014 the researchers Alexander Volynkin and Michael McCord conducted an empirical study at the SEI (Software Engineering Institute), a non-profit organisation at Carnegie Mellon University (CMU) in Pittsburgh, U.S.A. Ostensibly, SEI started hosting part of Tor's relays but then intentionally set up 'malicious actors' so that they could carry out their research. According to one anonymous source reported at *Motherboard*, SEI had the ability to deanonymize a new Tor hidden service in less than two weeks. Existing hidden services required upwards of a month, maybe even two months. The trick is that you have to get your attacking Tor nodes into a privileged position in the Tor network, and this is easier for new hidden services than for existing hidden services (Cox 2015).

The SEI researchers were supposedly able to bypass security and obtain around 1000 IP addresses of users with their hack. In July 2014 the researchers from SEI at CMU were invited to present their methods and findings on how to 'de-anonymize hundreds of thousands of Tor clients and thousands of hidden services' at the *Black Hat* security conference. This is part of the DefCon conference in Las Vegas in August every year but they never showed up, the reason for their cancellation is still posted on the website (Figure 84).

[137] On 2 April 2019 Hacking Team was acquired by InTheCyber to create Memento Labs, also a change of name to create a new reputation. https://en.wikipedia.org/wiki/Hacking_Team.

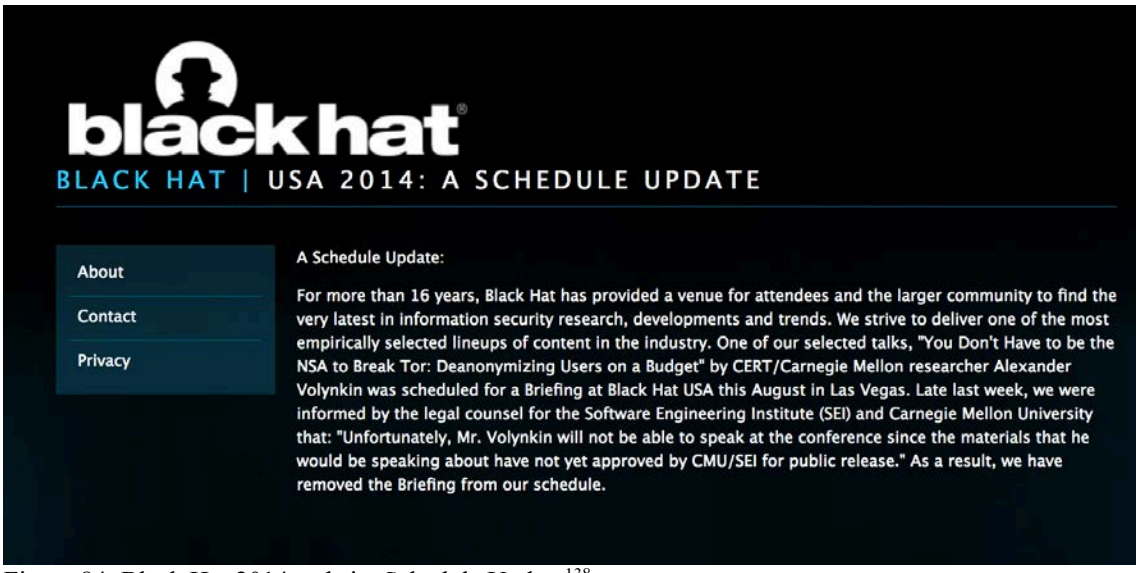


Figure 84: Black Hat 2014 website Schedule Update[138]

However, the Internet Archive's *Way Back Machine* captured the researcher's abstract that elucidates their braggadocio of a low budget exploit of Tor for around $3000, as well as a call out to others: (Figure 85)

> Looking for the IP address of a Tor user? Not a problem. Trying to uncover the location of a Hidden Service? Done. We know because we tested it, in the wild…. (Volynkin 2014).

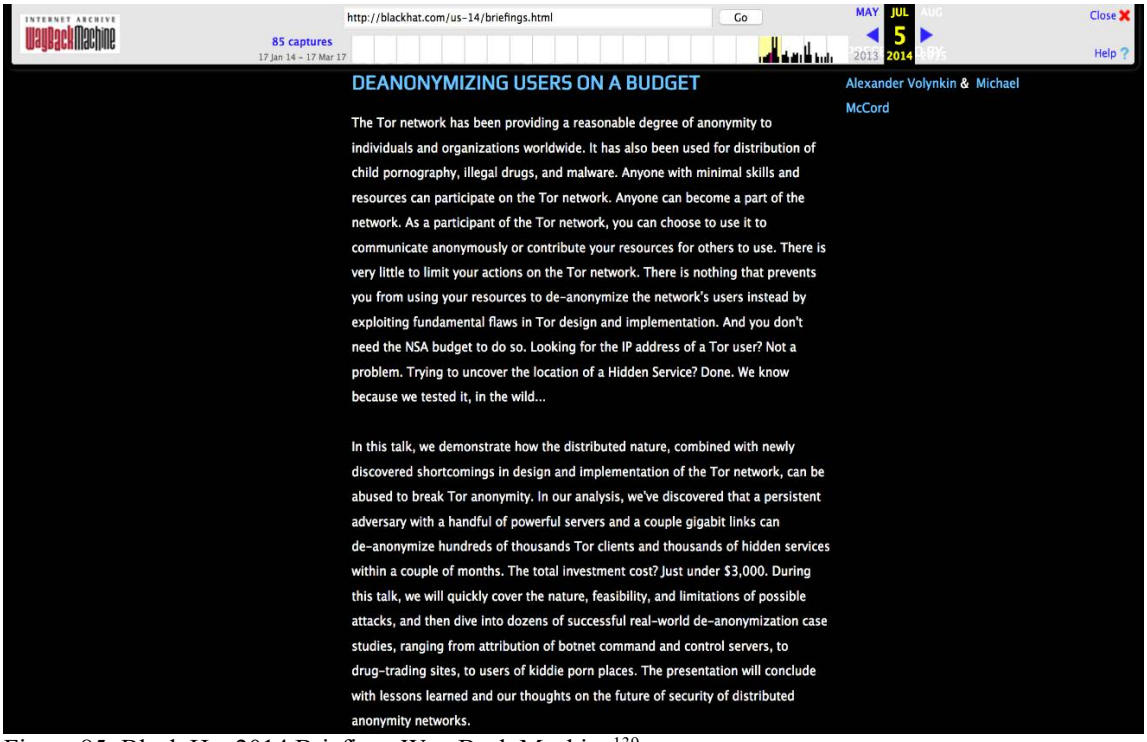


Figure 85: Black Hat 2014 Briefings.Way Back Machine[139]

[138] https://www.blackhat.com/latestintel/07212014-a-schedule-update.html
[139] https://web.archive.org/web/20140625125021/https://www.blackhat.com/us-14/briefings.html#you-dont-have-to-be-the-nsa-to-break-tor-deanonymizing-users-on-a-budget

With regard to ethical research considerations, the researchers' 'anonymous users' didn't realize or know they were participating in a study-cum-hack. Many in the security research community regard this as an infringement of ethical standards included in the *IEEE Code of Ethics* that prohibits 'injuring others, their property, reputation, or employment by false or malicious action' (IEEE n.D.:section 2.4.2). Even when following such an officially recognized IEEE ethical code, 'failure, discovery, and unintended or collateral consequences of success' (Greenwald et. al. 2008:78) could potentially harm 'objects of study'—in this case visitors to Tor Hidden Services. Besides the 'usual suspects' there are also academics carrying out research—contacting users, building their trust and protecting their sources—as demonstrated previously by Gehl (2014) and Bartlett (2014).

Moreover, unlike other empirical studies, the SEI researchers did not inform the Tor Project of their exploits. Instead, Tor discovered the exploits and contacted the researchers, who declined to provide information. Only after the abstract for Black Hat was published online (late June 2014) did the researchers 'give the Tor Project a few hints about the attack but did not reveal details' (Felten 2014). The Tor Project ejected the attacking relays and worked on a fix for all of July 2014 and at the end of the month, released a software update, along with an explanation of the attack (Dingledine 2014a). As this case shows, not only 'malicious actors' but also certain researchers collected data on Tor users, unbeknownst to them. According to the Tor Project director Roger Dingledine the SEI researchers acted inappropriately:

> Such action is a violation of our trust and basic guidelines for ethical research. We strongly support independent research on our software and network, but this attack crosses the crucial line between research and endangering innocent users (ibid).

However, according to Nicolas Christin, an assistant researcher at CMU, SEI is a non-profit, not an academic institution and in 2015 it renewed a 5-year governmental contract for 1,73 billion dollars (Lynch 2015)—therefore the researchers at SEI are not academics but instead are 'focusing specifically on software-related security and engineering issues'. To be precise, SEI is one of the Federally Funded Research and Development Centers (FFRDC), which are unique non-profit entities sponsored and funded by the U.S. government that address long-term problems of considerable complexity, analyse technical questions with a high degree of objectivity, and provide creative and cost-effective solutions to government problems (ibid).

2.3 Search and Seizure: Who's Hacking Whom?[140]

In January 2015, Brian Farrell aka 'Doctor Clu,' was arrested and charged with one count of conspiracy to distribute illegal hard drugs such as cocaine, methamphetamine and heroin at a Tor Hidden Service marketplace (Silk Road 2.0) on the Dark Net (Geuss 2015).[141] His IP address (along with other users) was purportedly captured in early 2014 by researchers Volynkin and McCord when they were carrying out their empirical study at SEI.[142] In

[140] My eponymous 2017 article is a detailed account of this case: https://limn.it/articles/whos-hacking-whom/
[141] (U.S. v. Farrell, U.S. District Court, W.D. Wash., No. 15-mj-00016) Complaint for Violation. Available here: https://cdn.arstechnica.net/wp-content/uploads/2015/01/5498263-0-14302.pdf
[142] Nick Mathewson, the co-founder of the Tor Project made this statement regarding the case: "If you're doing an experiment without the knowledge or consent of the people you're experimenting on, you might be doing something questionable—and if you're doing it without their informed consent because you know they wouldn't give it to you, then you're almost certainly doing something wrong. Whatever you're doing, it isn't science." Available here: http://se.azinstall.net/2016/08/how-tor-users-got-caught-part-2-cliff.html

November 2015, the integrity of these two SEI researchers was again brought into question when the rumour circulated that they had been subpoenaed by the FBI to hand over their collated IP addresses (around 1000). 'Since the beginning of 2014, an unnamed source had been giving authorities the IP address of those who accessed this specific part of the site' (Vinton 2015). Legally, in the U.S., the FBI, SEC and the DEA can all subpoena researchers to share their research. Matt Blaze, a computer scientist at the University of Pennsylvania made the following statement about conducting research:

> When you do experiments on a live network and keep the data, that data is a record that can be subpoenaed. As academics, we're not used to thinking about that. But it can happen, and it did happen (Blaze 2016 cited by Vitáris 2016).

However, the obtained information was not for public consumption but for an agency within the U.S. Department of Justice, the FBI. In an official media statement, CMU's SEI responded by explaining that their mission encompassed searching and identifying 'vulnerabilities in software and computing networks so that they may be corrected' (CMU 2015).

Besides the ethical questions regarding the researchers handing over their findings to governments that have supported them (ostensibly with tax-payer money), the politics of security research and vulnerability disclosure continues to be a heated debate within academia and the general public. The issuing of subpoenas by law enforcement might provide a means to gather data on citizens and to obtain knowledge of academic research—which then remains hidden from the public. Computer security defence lawyer Tor Ekeland commented on the FBI's involvement:

> It seems like they're trying to subpoena surveillance techniques. They're trying to acquire intel[ligence] gathering methods under the pretext of an individual criminal investigation (Vitáris 2016).

In the same month there was another rumour that the FBI actually offered the SEI researchers 1,000,000 dollars for their research on how to crack Tor. It is still not clear whether the FBI was using a subpoena to acquire exploits, or if the SEI researchers were originally hired by the FBI and only later disclosed what happened, stating that they had been subpoenaed. Either way, it would raise the issue of whether the FBI required a search warrant in order to obtain the evidence—the IP addresses.

In January 2016, Farrell's defence filed a motion to 'compel discovery', in an attempt to understand exactly how his IP address was obtained, as well as the two-year history of the relationship between the FBI and SEI through working contracts. In February 2016, the Farrell case came to court in Seattle where it was finally revealed to the public that the 'university-based research institute' was indeed confirmed to be SEI at CMU, subpoenaed by the FBI (Farivar 2016). This statement from the order—*Section II, Analysis*—written by US District Judge Richard A. Jones answered the question of whether a search warrant was needed to obtain IP addresses: 'SEI's identification of the defendant's IP address because of his use of the Tor network did not constitute a search subject to Fourth Amendment scrutiny' (Cox 2016). Search then, is not only an online activity to acquire information with search engines but here is defined in relation to a physical search (and eventual seizure) of one's possessions (IP address) according to the US Constitution's Fourth Amendment.

Historically, this legislation doesn't allow unauthorised searching, unless a warrant had been issued and served and that warrant needs to specify those places and things to be searched, along with a suspicion of wrongdoing. There are jurisdictions that determine where the search can occur, such as an 'open field' or 'curtilage' the area directly surrounding the home. Then there is the question of 'probable cause' — that the search will uncover criminal activity. Regarding the confiscation of Farrell's IP address and how it was obtained, ostensibly, according to the FBI, it didn't warrant a warrant (although there was a warrant for his physical address).[143] It seems that if one does not object, or that the information was given voluntary (in this case an IP address) then the Fourth Amendment can be waived. In order to claim protection under the Fourth Amendment, there needs to be a demonstration of an 'expectation of privacy,' which is not subjective but recognized as reasonable by other members of society. Judge Jones claimed that the IP address, 'even those of Tor users, are public, and that Tor users lack a reasonable 'expectation of privacy' (Cox 2016).

Again, according to the party's submissions, such a submission is made despite the understanding communicated by the Tor Project that the Tor network has vulnerabilities and that users might not remain anonymous. Under these circumstances Tor users clearly lack a reasonable expectation of privacy in their IP addresses while using the Tor network. In other words, they take a significant gamble on any real expectation of privacy under these circumstances (Jones 2016:3).

2.4 Expectation of Privacy

In March 2016 Farrell eventually pleaded guilty to one count of conspiracy regarding the distribution of heroin, cocaine and amphetamines in connection with the Hidden Service Silk Road 2.0 and received an eight-year prison sentence. In the Farrell case, the protection of an anonymous IP address was thwarted in various ways (a hack, a subpoena, a ruling) with regard to governmental participation. Privacy technologists, such as Christopher Soghoian, have provided testimony in similar cases, explaining that the government states that obtaining IP addresses 'isn't such a big deal,' yet the government can't seem to elucidate how they could actually obtain them (Kopstein 2016).[144]

> If the NSA and the Russian government can't crack Tor – Russia's Ministry of Internal Affairs has offered a prize of 3.9 million roubles to anyone who succeeds – then it's hard to imagine that the FBI did it on its own (Glenny 2015:4).[145]

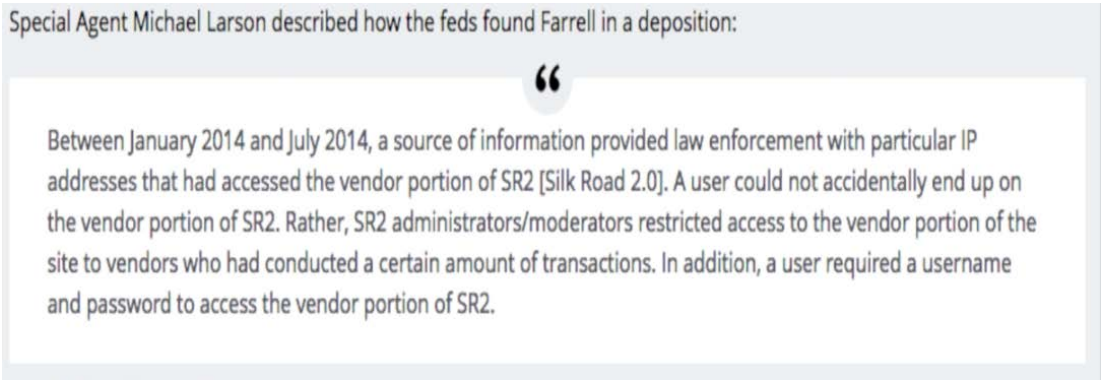
---

[143] An FBI Source of Information (SOI) provided "reliable IP addresses for TOR and hidden services such as Silk Road 2." […] the warrant goes by this: "The SOI also identified approximately 78 IP addresses that accessed a vendor.onion address," […] When Farrell's case was held in the court, the defense made this statement: "On October 12, 2015, the government provided defense counsel a letter indicating that Mr. Farrell's involvement with Silk Road 2.0 was identified based on information obtained by a 'university-based research institute' that operated its own computers on the anonymous network used by Silk Road 2.0"'. Available here: http://se.azinstall.net/2016/08/how-tor-users-got-caught-part-2-cliff.html

[144] 'Civil liberties are under attack if law enforcement believes it can circumvent the rules of evidence by outsourcing police work to universities. If academia uses "research" as a stalking horse for privacy invasion, the entire enterprise of security research will fall into disrepute. Legitimate privacy researchers study many online systems, including social networks — If this kind of FBI attack by university proxy is accepted, no one will have meaningful 4th Amendment protections online and everyone is at risk' (Dingledine 2015).

[145] 'We teach law enforcement agents that they can use Tor to do their investigations ethically, and we support such use of Tor — but the mere veneer of a law enforcement investigation cannot justify wholesale invasion of people's privacy, and certainly cannot give it the color of 'legitimate research' (Dingledine 2015).

Whoever wanted to know the IP address would have to be in control of many nodes around the world in the Tor network and one would have to intercept this traffic and collect IP addresses at the exit nodes. Besides the difficulty factor, correlation techniques cost time and money and certain exploits, including the one from the SEI researchers, were possible in 2014. Even if IP addresses are considered public when using Tor, they are anonymous unless they are correlated to a specific individual's device. As shown below in the deposition, a 'source of information' provided the IP addresses to law enforcement; in turn The FBI obtained the list of IP addresses from Farrell's ISP provider, Comcast and his housemate was identified who in turn provided testimony against Farrell. (Figure 86)


Special Agent Michael Larson described how the feds found Farrell in a deposition:

"

Between January 2014 and July 2014, a source of information provided law enforcement with particular IP addresses that had accessed the vendor portion of SR2 [Silk Road 2.0]. A user could not accidentally end up on the vendor portion of SR2. Rather, SR2 administrators/moderators restricted access to the vendor portion of the site to vendors who had conducted a certain amount of transactions. In addition, a user required a username and password to access the vendor portion of SR2.

Figure 86: Garrett Fogerlie's blogpost (August 2016), which cites an agent's deposition

Additionally, the court denied the defence's motion to compel discovery about the 'Network Investigative Technique' (NIT). The judge's cited reason was that IP addresses are in and of themselves not private, as people willingly provide them to third parties such as with using the telephone (Buckner 2008:504), which I mentioned in Chapter 6. Judge Jones also argued that Farrell didn't have an expectation of privacy because he handed over his IP address to strangers running the Tor network:

> [I]t is the Court's understanding that in order for a prospective user to use the Tor network they must disclose information, including their IP addresses, to unknown individuals running Tor nodes, so that their communications can be directed towards their destinations. Under such a system, an individual would necessarily be disclosing his identifying information to complete strangers (2016:3).

Judge Jones stated that IP addresses are public and reasoned that Farrell didn't have a reasonable expectation of privacy because he willingly gave his IP address to an Internet Service Provider (ISP), in order to have internet access. Moreover, the citation (precedent) that Judge Jones drew upon to uphold his order, *United States v. Forrester,* ruled that individuals have no reasonable 'expectation of privacy' with internet IP addresses and email addresses:

> The Court reaches this conclusion primarily upon reliance on *United States v. Forrester*, 512 F.2d 500 (9th Cir. 2007). In *Forrester*, the court clearly enunciated that: Internet users have no expectation of privacy in …the IP address of the websites they visit because they should know that this information is provided to and used by Internet service providers [ISP] for the specific purpose of directing the routing of information (ibid:2-3).

However, as Deborah Buckner argues in her 2008 article, *Internet Search and Seizure in United States v. Forrester,* in regard to 'electronic surveillance', unlike telephone numbers, 'pen registers' such as IP addresses and emails are not just 'routing information' but have more content, not only identifying the user but visited webpages and entered search terms. Buckner therefore states that a user's expectation of privacy is much greater and that the court erred in the *United States v. Forrester* ruling, which I describe in more detail (Appendix H).

2.5 Control and Freedom

In recent years the reputation economy of the Tor network, as seen by the media, law enforcement and the general public, is that of hiding Dark Net markets, child porn, criminals and terrorists. Yet '[d]espite this dominant idea of the dark web as only useful to pedophiles, assassins, and junkies, recently more Internet users have started to use Tor and even 'hidden.onion' sites' (Gehl 2014:1222-3). In an era where multinational companies and governments organise the personalised subject's online experiences around advertising, data tracking and surveillance, users also apply Tor in order to anonymously search the 'vanilla internet'. Statistics indicate that in 2015 the Tor project estimated 'that about 3.4% of client traffic is hidden-service traffic and 6.1% of traffic seen at a relay is hidden-service traffic' (Kadianakis 2015).[146] In other words, around 96,6% of Tor traffic was NOT Hidden Services.

Returning to Chapter 1 and the changing names of the address offices representative of their functions, Tor also renamed its 'hidden services' to 'onion services' in 2015, 'to reflect the fact that they provide more than just the "hiding" of a service—more importantly, they provide end-to-end security and self-certifying domain names' (Winter et al. 2018). This change in nomenclature follows the development of Tor and its historical 'onion routing', along with its present dynamic 'reputation economy' among its various user groups. As shown by numerous studies (Forte et al. 2017; Khattak et al. 2016; Gehl 2014) there is discrimination of Tor users. The 2016 study *Do You See What I See? Differential Treatment of Anonymous Users* explores how Tor users receive 'differential treatment' that 'involves websites providing Tor users with degraded service, resulting in them effectively being relegated to the role of second-class citizens on the Internet' (Khattak et al. 2016:1). Khattak et al.'s research specifically focuses on censorship; it is not directed at content or traffic *entering* Tor but rather in relation to traffic *exiting* Tor (2016). By examining the characterization of websites and IP addresses, they were able to establish how Tor users are discriminated against at the network layer, 'a straightforward technique for services to block Tor is to filter traffic from publicly listed exit nodes' (ibid:3). Concomitantly their research proposes that 'at least 3,67% of the top 1000 Alexa web sites block Tor users at the application layer' (ibid:1).

Another study *Privacy, Anonymity, and Perceived Risk in Open Collaboration: A Study of Tor Users and Wikipedians* (Forte et al. 2017), supports this by demonstrating how other websites 'block Tor users from posting outright by blacklisting IP addresses that are known to be Tor exit nodes. One of these sites is Wikipedia' (Forte et al. 2017:9). Forte et al. also describe how Tor users repeatedly received CAPTCHAs (Completely Automated Public Turing Test To Tell Computers and Humans Apart), stating that 'the web hosting service Cloudflare (a large content delivery network, or CDN) presents Tor users with CAPTCHAs to such an extent that a

[146] Tor's executive director Roger Dingledine also stressed that hidden services websites only account for 2% of total traffic using Tor's anonymising technology: a warning not to confuse websites hiding themselves with individual internet users using Tor to surf the web anonymously (Dredge 2014).

defeated interviewee returned to using a normal browser (ibid:8).[147] This is echoed by my own study. Whilst collecting my small data set on Tor I would sometimes get a CAPTCHA, which interrupted my data gathering process. I postulate that this happened because I was perceived as a Tor user using google.com as the Tor search engine. (Figure 87)

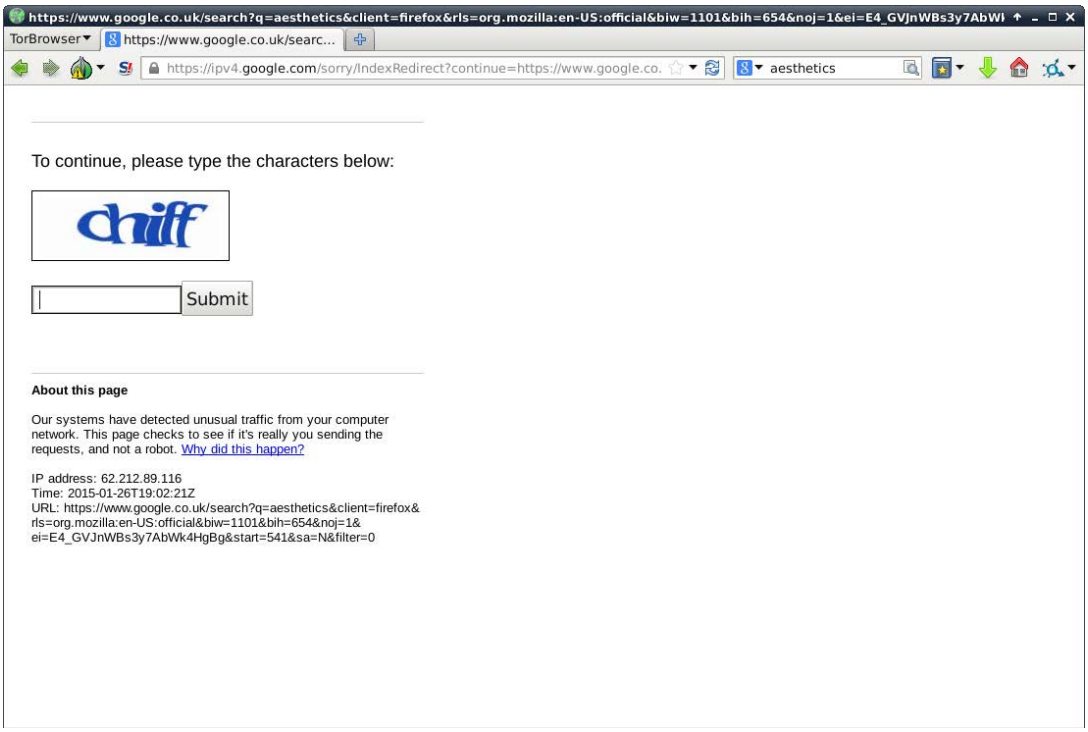


Figure 87: CAPTCHA in Tor Browser, January 26, 2015

Websites and servers 'by default assesses the "reputation" of each client IP address in terms of how much malicious traffic it has been observed to send, and blocks attempted access by clients with sufficiently poor reputations' (Khattak et al. 2016:8). However, this differential treatment of Tor users is often unfounded, resulting in second-class treatment of anonymous users, often deemed criminal by association when using Tor. As discussed previously, '[t]raditional threats to Tor involve deanonymization attacks that reduce user privacy, or governments blocking access to the Tor network' (ibid), which still occurs in certain countries. Tor fights censorship but Tor is itself often censored**.**

Over the past few years Tor has increasingly been applied to protect the anonymity of the user in areas of protest and freedom of speech issues. Actually it is anyone desiring to be anonymous, 'a list that includes journalists and their sources, human rights activists, political dissidents living under oppressive countries and many others who have various reasons for needing to shield their identity and their online activity' (Zetter 2015).[148] As shown above, Tor is also used

[147] In another study, for one interviewee, 'the kind of identity knowledge he wanted to circumvent by using Tor was not his name, but his IP address' (ibid). 'He explained that he does not care if blog owners know who he is when he comments, he cares that they are not able to hack or locate him' (Forte et al. 2017:9).

[148] 'Ostensibly the NSA would like to label anyone who is investigating privacy and anonymity online as an 'extremist' and would collect their IP address for inclusion in its database. Not only those using Tor or investigating privacy services such as TAILS are being monitored, and 'fingerprinted' (tracking IP addresses) but people who visit websites, read articles from journals such as the Linux Journal (according to the NSA an 'extremist forum') and even WIRED, which is where this source comes from, are being collated in the huge troller of data' (Zetter 2015).

by whistleblowers and the controversial content they report, but also lawyers who need to discuss their cases with undocumented immigrants, or by average citizens to disseminate information about reproductive rights in countries where abortion is illegal. De-anonymisation is a security threat to their lives. As explained in Chapter 6, all around the world some of these activists and users are also operating relays for Tor's network, or creating Tor 'bridges' when necessary.

> However, XKeyscore records any connections to the bridges.torproject.org server and uses a microplugin to then read the contents of the email that the Tor Project sends to requesters in order to obtain the address of the bridge (ibid).

Edward Snowden has described XKeyscore as a 'front-end search engine' that collates all users' data, metadata and especially searching activities, what it calls 'raw traffic'. It catalogues the searched URLs of Tor Onion Services in the Dark Net, thereby verifying that the NSA is tracking those interested in privatising services and tools.[149] With this XKeyscore system they are basically 'saying that anyone who is looking for those various [services] are suspicious persons' (ibid). Moreover, TAILS software also did not go overlooked: 'The NSA clearly regards Tails as a sinister tool, however, referring to it in one comment in the source code as "a comsec mechanism advocated by extremists on extremist forums"'(ibid).

As explained above Tor 'does not guarantee perfect anonymity; if you don't use a Virtual Private Network in addition to Tor, people can still see you're using Tor even if they can't necessarily see what sites you're visiting' (Emerson 2016). In order to reach and 'tour' Tor Hidden Services (a.k.a. Onion Services), I had to pass the 'admissions test', what Gehl defines as 'techno-elitism' (2014), in this case, the Tor Browser in combination with TAILS. My research also indicates that Tor had certain design flaws, which enabled malicious attacks, penetration and even infiltration of the Tor network, as confirmed by the empirical studies (Biryukov et al. 2013; Chakravarty et al. 2014). These have been patched. Tor continues to anonymise the origin of the traffic and ensures encryption inside the Tor network yet it 'does not magically encrypt all traffic throughout the Internet' (Emin et al. 2015:30). In other words, Tor does not offer 100% anonymity since the exit node operators are in a position to capture any traffic passing through it (ibid:29).

Although it has been shown that the anonymity network Tor is not without risks Wendy Chun reminds all readers and users, worldwide,

> from our position of vulnerability, we must seize a freedom that always moves beyond our control, that carries with it no guarantees but rather constantly engenders decisions to be made and actions to perform (Chun 2005:30).

According to the 2013 Snowden revelations, the NSA 'describes the [Tor] network as the king of high-secure, low-latency Internet anonymity' (Gross 2013) and admits that they can only de-anonymise a few users. However, the 'NSA has had "no success de-anonymizing a user in response" to a specific request, the document said' (ibid). Snowden's documents also revealed the frustration of the NSA, reflected by their Powerpoint presentation *Tor Stinks*:

[149] https://web.archive.org/web/20140128224439/http://www.ndr.de/ratgeber/netzwelt/snowden277_page-3.html

'We will never be able to de-anonymize all Tor users all the time' (ibid).[150] Whether this frustration will continue remains to be seen. In an era of so-called Big Data, as more user data is collated (by Google, governments *and* researchers) correlation becomes easier and deanonymisation occurs more frequently, yet Tor and TAILS still provide certain degrees of anonymity. In the next chapter I compare the effects of Google Search to the effects of 'reimagining search' when using Tor.

[150] The Guardian. 2013 "Tor Stinks, NSA Presentation Document." Oct. 4, 2013. Link is no longer available. (http://www.theguardian.com/world/interactive/2013/oct/04/tor-stinks-nsa-presentation-document).

Chapter 8:

Black Box vs. Black Bloc

1.0 Black Box, Black Bloc

In Chapter 1 I elucidated 'the pre-history' of search engines before the age of Google through the 'address offices', institutions spanning the 16th-19th centuries where people came to exchange information at urban centres (Tantner 2015). In order to contact the parties involved, these offices collated the addresses of newcomers who arrived in cities looking for work and wrote them down in registers or 'Protokollerbücher'. These 'Protokollerbücher' also organised information about employment, housing, health care and social services with 'users' paying to advertise, as well as to view the public register. However, there were also private 'Protokollerbücher', or 'anonymity machines', where only certain parties had access to the information (Tantner 2015). Although these 'address offices' enabled access to information for most people, they operated as a bureaucratic institution that applied protocols (and 'Protokollerbücher') as a means of standardising captured data and as a process of agreement (contract). In Chapter 2 I described various early internet protocols, such as hyperlinks using HTTP with the World Wide Web, which enables the authority of PageRank (Chapter 3)**.** As shown in Chapter 4, 'commands, address and data' (Kittler 1999) constitute these 'protocols', not only for the registers of the former address offices but also for how contemporary search works, as I explained in Chapter 5.

There I employed the OSI (Open Systems Interconnect), an abstract model comprised of 7 layers, to describe how communication is facilitated between computers and networks, with the internet primarily operating within the 4th and 3rd layers—TCP/IP (Transport Control Protocol/Internet Protocol). Alexander R. Galloway's book *Protocol: How Control Exists after Decentralization*, articulates that one of the ways the protocological 'determines control after decentralisation' is through the configuration of the relationship between TCP/IP and DNS (Domain Name Server), as they are 'political technologies' (2004). TCP/IP is now the standard internet protocol suite and DNS, which runs parallel to HTPP in the application layer (7th), is responsible for translating the domain names into numerical IP addresses in order to identify devices and locate them within the network protocols. In *The Personalised Subject*, I explained how the IP address facilitates not only communication between parties, but the identification of 'subjects' searching with Google. Conversely, for *The Anonymous User* (Chapter 7) the IP address is part of the protocol that facilitates the transport of data but the IP address is not revealed because of layers of encryption within the Tor network.

In this chapter I use the IP address as an organisational architecture to answer my subresearch question, *what are the effects of search engines on (us)ers?* I combine this with Alexander R. Galloway's *Black Box, Black Bloc* lecture (2010), which provides a conceptual (and technological) framework to synthesise results from my two empirical chapters (5 and 7) and is, coincidentally, analogous to the arrangement of my thesis.[151]

> The black box: an opaque technological device for which only the inputs and outputs are known. The black bloc: a tactic of anonymization and massification often associated

[151] Thanks to an 'opponent' at my WIP2, Nanna Thylstrup, who suggested I read this text.

> with the direct action wing of the left. Somehow these two things come together near the end of the twentieth century. Is there a reason for this? (ibid:3).

At the end of *The Personalised Subject*, I discussed the black box (Appendix G) and the Intellectual Property (IP) of Google's proprietary search algorithm through '[in]visibility management' (Flyverbom et al. 2016) and the 'media arcane' (Beyes and Pias 2019). This 'blackness' of the black box is also found in the Black Bloc that is analogous to tactics of 'obfuscation' (Brunto and Nissenbaum 2015), such as the Tor Browser, which hides the user's IP address, as elucidated in Chapters 6 and *The Anonymous User* (7). Drawing on the results from my methods and by applying the IP address as a conceptual hinge, I demonstrate the following effects through a progression of human/algorithmic interaction. With Google Search the IP address swings open to collect data on the subject to produce various *Subjectivities of Search*. With the Tor Browser it also closes back, preventing the user from being identified and instead enables *Agencies of Anonymity*.

Figure 88: Data subject

The structure of both these 'effects' begins with search algorithms interacting with myself as a researcher collecting data online and as a 'data subject', who is protected by law. (Figure 88) Since May 25, 2018 the GDPR (General Data Privacy Regulation) has been implemented in Europe, which regulates the collating, processing, storage and transmission of personal data of EU citizens, or 'data subjects'.[152] According to Article 4, the 'data subject' is an end user whose personal data can be collected through 'direct identification' with an IP address. It is the recognition of citizens as 'data subjects' by their IP address that facilitates certain effects and the organisation of those searching online—through degrees of personalisation (black box), or not, with degrees of anonymity (black bloc).

[152] https://gdpr-info.eu/art-4-gdpr/

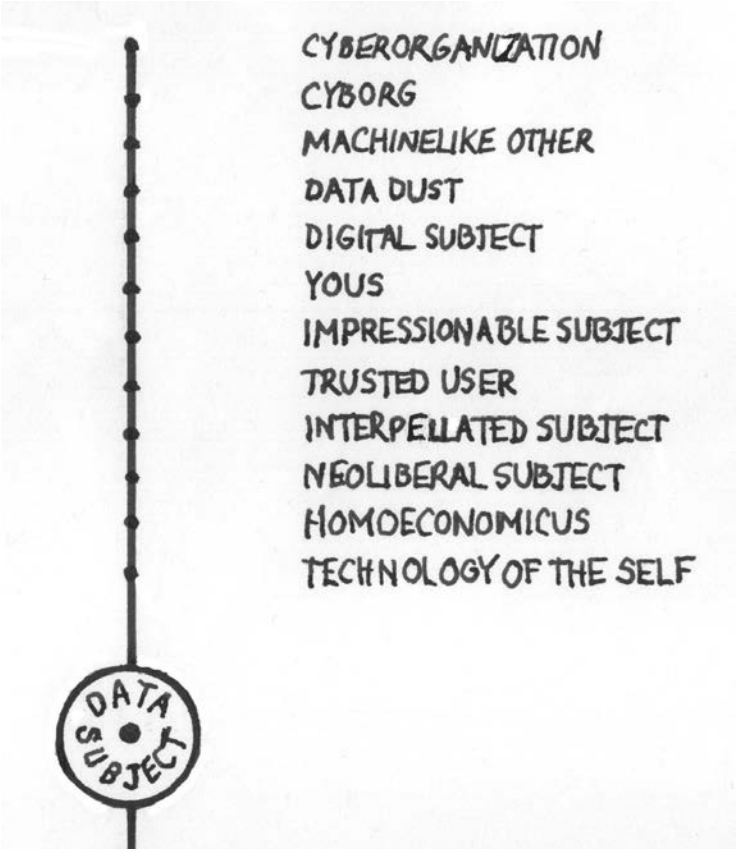


Figure 89: Subjectivities of Search

2.0 Subjectivities of Search (Black Box)

In the section, I incorporate my results from Google Search (Black Box) within a structured discourse analysis to postulate degrees of human-algorithmic interaction into a range of effects: *Subjectivities of Search*. (Figure 89)

2.1 Technology of the self

As described in Chapter 4, the philosopher Michel Foucault acknowledges Montaigne and his oikos (Chapter 1)—the capturing of one's private data in accounting books or registers—as a 'technology of the self'. During my 'experiment in living', I interacted with Google Search algorithms and used tools, such as my computer, as memory extensions or *hupomnemata,* to 'note down' and collect data on myself—keywords and search results. This 'technology of the self' can be used as an instrument to analyse the relationship between subject and truth, where the 'personalised subject' explores power constructs—how the subject constituted itself in one form or another, where 'power is games of strategy'. With my 'critical ethnography of the self' and following Foucault, I also showed that it 'is precisely the historical constitution of these various forms of the subject in relation to the games of truth' (ibid:290-291). As my search histories are constantly collected by Google Search, the 'personalised subject' is not a substance but a form, which is 'not primarily or always identical to itself' as it is changes in different contexts and situations. Through diverse practices such as online interaction with search algorithms as a 'truth game', this modern day 'technology of the self' transforms me as a (data) subject.

As mentioned in the Introduction, Google 'shapes the web'; concomitantly it shapes the subject––one needs to be indexed (Introna and Nissenbaum 2000) and found through search engines in order to accrue value in the reputation economy. Chapters 3 and 5 explained that the pages of results and the ranking of links determines the public visibility of the personalised subject to others using the web. These personalised subjects and content producers make themselves 'algorithmically recognizable in all sorts of ways,'(Gillespie 2014:88), which reflect the accumulation of online social value and the building of a reputation economy.

> But algorithms can also function as a particularly compelling "technology of the self" (Foucault 1988) when they seem to independently ratify one's public visibility. It is now common practice to Google oneself: seeing me appear as the top result in a search for my name offers a kind of assurance of my tenuous public existence (Gillespie 2014:186).

Personalised subjects google themselves at some point, either to measure their attention economy—where everything is based on visibility—or to see what has been written or published about them by search results and that they are indexed. It has become the meter to measure success: Appearing higher in Google's ranking, adding to one's visibility, is a particular kind of attention seeking that embodies Foucault's figure of the 'Homoeconomicus', or 'economical man'(1988:16).

2.2 Homoeconomicus

Foucault was interested in the subject and more specifically, 'the way a human being turns himself into a subject' and part of his scholarship looks at the history of discipline and what he came to term 'biopolitics'. Instead of a negative understanding of sovereign power, Foucault mapped out a new form of power, discipline, with 'its starting point being the physical separation of individuals' (Borch 2015:10).

> Making the individual the object of attention facilitates intervention —partly because isolating the individual enhances the potential for collating and generating knowledge about them…Consequently, discipline, this "political technology of the body", is deployed at a level of detail (ibid).

As Foucault pointed out, the technology of disciplinary power was also carried out spatially, with the transition from discipline as a public form to power to an enclosed and hidden prison, as expressed through Jeremy and Samuel Bentham's *Panopticon* (1791).[153] The mere power of observation, where the subject (or prisoner) is aware of being watched, yet at the same time does not know if they are being watched, enables 'self-discipline predicated on the mere possibility of being watched' (Foucault 1977:201 cited by ibid:10).[154] As shown by my results in Chapter 5, the contemporary condition of the Panopticon is that people are now aware of being watched as well as tracked when carrying out online Google searches in the Firefox browser, as exemplified by the *Panopticlick*, (Appendix B).

[153] This type of architecture enables a 'concerted distribution of bodies, surfaces, lights, gazes; in an arrangement whose internal mechanisms produce the relation in which individuals are caught up' (Foucault 1977:202 cited by ibid:11).

[154] Simone Browne's book Dark Matters (2015) investigates the surveillance of blackness through various lenses, from slave ships, to the Bentham brothers' Panopticon to the contemporary condition as a social and political norm.

Foucault (1975) described this 'regularizing control' of discipline, what has come to be 'focused on the targeted control of individual conduct' (Flyverbom et al. 2016:389) in regard to institutions and bodies. Going beyond sovereign power and discipline, biopolitics 'aimed to foster life' and, similar to discipline, has a 'productive form of power' yet differs in the way it intervenes: whereas discipline was individual, biopolitics 'strives to regulate individuals as an aggregated totality, as a population' (Borch 2015:12). Later Foucault used the term governmentality (1978), a portmanteau of government and rationality, to elucidate how populations are controlled through biopolitics. Additionally, governmentality is not part of a linear development after sovereignty and discipline; rather during the course of his research Foucault changed tack and came to present them as a triangle: 'sovereignty, discipline and government management, which has population as its main target and apparatuses of security as its essential mechanism' (Foucault 2007:107-8 cited by Borch 2015:13).

> Foucault uses the term 'biopolitics' in order to elucidate how political power is carried out on every aspect of human life, making individuals and the Homoeconomicus someone who is eminently governable (Foucault 2008:226 cited by Mirowski 2013).

Calculated practices (such as Google Search), permit individuals to govern themselves, which epitomises the biopolitical and lies at the core of neoliberalism.

2.3 Neoliberal Subjects

> Neoliberal subjects—small sovereigns—are always searching, rarely finding. Shifting from the zoom to the overview, from search term to search term, they defer and extend decisions: the end, like that mythic pot of gold, is never reached. At the same time, though, users' searches produce data that make users findable, even as they wander (Chun 2016:77).

Although the Homoeconomicus is often considered a rational agent in pursuit of self-interest, these subjectivity-defined ends are part of an economic civil society that operates through production and exchange, which is part of the technology of liberal governmentality.
As *Panopticlick* shows, Google search facilitates online tracking and (self) surveillance, simultaneously optimising searching subjects. This biopolitics is organised by

> the image, idea, or theme-program of a society in which there is an *optimization* of systems of difference, in which the field is left open to fluctuating practices, in which minority individuals and practices are tolerated, in which action is brought to bear on the rules of the game rather than the players, and finally in which there is an environmental type of intervention instead of the internal subjugation of individuals (Foucault 2008:259-60 cited by Borch 2015:14 emphasis mine).

This prescient 'optimization' could nowadays be applied to personalised search results, which although declared individualised, organise the subject into categories based on others with similar interests, characteristics and behaviour patterns, as elucidated in Chapter 5. This is supported by Foucault's analysis of power that concerned itself with subjectification, or the way in which 'subjects are constructed as specific subjects' (1982:212) where 'on the one hand, one can be subject/subjugated to the control of others. On the other hand, one can subordinate oneself' (Borch 2015:14). What is crucial to note is that Foucault focused on the

‘interrelationships between the two’ yet opposed the idea of a static, ‘ahistorical, autonomous core identity’ and instead contributed to the field of knowledge with how ‘new forms of subjectivity have been produced throughout history’ and with the proposal that power is ‘productive/creative by nature’ (ibid:14-15).

> This form of power applies itself to immediate everyday life which categorizes the individual, marks him [sic] by his own individuality, attaches him to his own identity, imposes a law of truth on him which he must recognize and which others have to recognize in him (Foucault 1982:212 cited by Borch:15).

Through this interaction, the subject is both recognised, and subject to, the law.

2.4 Interpellated subject

In Chapter 1 I presented the two primary definitions of address for this thesis, namely the physical address, which allowed people to be identified and assigned a location and the act of being addressed, in which the individual acknowledges that they are being recognised as a subject. In the early days of neoliberal capitalist ideology, before it was defined as such, Louis Althusser articulated forms of address through the framework of ideology, which he deemed ‘interpellation’ (1971). In his essay *Ideology and Ideological State Apparatuses (Notes towards an Investigation)*, Althusser put forth the relationship of power structures and the individual to the state through a Marxist lens. With the constitutive process of interpellation, this ideology is recognised by the individual’s acknowledgement of becoming a subject that complicates their domination and subjugation. According to Althusser, ideology is also a type of mediation, as is protocol and this middle role is ‘the imaginary relationship of individuals to their real conditions of existence’ (1971:162). Terry Eagleton described Althusser’s concept of ideology

> as a particular organization of signifying practices which goes to constitute human beings as social subjects, and which produces the lived relations by which such subjects are connected to the dominant relations of production in a society (1991:18 cited by Mager 2014:31).

As members of society, systems of hegemonic power reproduce themselves by hiding governing factors as they simultaneously incorporate the subject into the structure of power. The classic example is that of Althusser’s policeman who shouts at a passer-by “hey, you there!” in public where the individual then responds by turning around. ‘[B]y this mere one-hundred-and-eighty-degree physical conversion, he [sic] becomes a *subject*’ (1971:174).

Subjects are thereby complicit in their own domination and this form of self-governing is incorporated into daily activities, such as responding to officials of the state when addressed and also when searching with Google. ‘The ideological superstructure and the economic base meet with and feed each other in every single Google query’ (Mager 2014:32). Previously it was the police who asked the question: ‘Hey you there?’ Nowadays ‘personalised subjects’ enhance the power structures of Google by recognising themselves as subjects when searching online, who are interpellated as ‘subjects’ by automatically acknowledging the ideology of Google Search by deciding to use it. It then becomes crucial for individual users to recognise their own interpellation when using ‘transnational informational capitalism’ (Fuchs 2011a) and how they relate to the capitalist ideology of Google search algorithms (Mager 2014).

2.5 Trusted User

As argued in Chapter 3, the capitalistic ideology of Google is embodied by PageRank's 'authority', where high rank reflects what is considered valuable or important, and '[m]ost likely, these are the pages to which someone performing a search would like to direct his or her attention'(Google 2014). In the Feedback section of *The Anatomy of a Large-Scale Hypertextual Web Search Engine* (1998), Brin and Page explain that parameters such as 'type-weights' and 'type-prox weights' need to be determined for ranking search results. Brin and Page further divulge that '[f]iguring 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' (1998:110). It became Brin and Page's 'trusted user' who interacts with the search engine and reinforces this 'preferential attachment' by clicking on links mostly found on the first page of Google. The 'trusted user' thereby actuates the 'relevance' and 'quality' of the search results by supplying feedback and data that Google saves and incorporates it into subsequent search results.

However, as discussed in Chapter 5, authority has transitioned to authenticity, with the language of Trusted Users' queries determining not only ranking but personalisation. Brin and Page recount their experiments with PageRank of 'increasing the weight of a user's home page or bookmarks,' thereby alluding to the notion of personalisation as another key innovation to Google's search engine (ibid:112). As they mention in their *Future Work* section they planned to

> support user context [locative data], and result summarization. Then when we modify the ranking function, we can see the impact of this change on all previous searches which were ranked (ibid:111).

These adjustments determined the ranking function, as did the previous searches [search history] and the impact of these changes resulted in personalisation. Brin and Page also made it known they had not done an 'extensive user study' to test the quality; instead they invited the reader to test out Google, to click on links, some of which were advertisements.

2.6 Impressionable Subject

User feedback has always played a salient role in the history of advertising. As elucidated in Chapter 3 and 5, the way in which audiences were created for advertising in early 'push' media such as television was one directional and there was no possibility for immediate feedback. As Lucas D. Introna points out, the business model of advertising regarding the present-day internet is 'historically contingent', that is, it is 'neither inevitable, nor the only possibility' (Introna 2016:26). Introna draws on a Foucauldian genealogy of the technical practices that make up the history of online display advertising and breaks the narrative down into four 'enactments': the gazing subject, the animated subject, the individuated subject and the branded subject, in order to map out how the impressionable subject 'becomes'. The gazing subject is captured and controlled by the GUI (which was invented in 1979, before the hyperlink) and unlike the push media curated by publishers of yesteryear, the pull or agency of the subject is enacted with their clicking on banner ads (the first currency of the internet) (ibid:29-32). With advances in interface development, such as the GUI and personalisation, viewers could see advertisements that were targeted at them. At exactly the right moment the 'animated subject' shows interest by clicking and interacting, with the amount of revenue generated determined by the 'conversion rate' of impressions (ibid:32-35).

As shown in Chapter 5, 'platforms intervene' (Gillespie 2014), with the impressionable subject primed to receive advertising that is relevant to their wishes, desires and needs based on previous interactions with the application interface, Google Search. By interacting with content on various websites, including Google Ads, visits are recorded by tracking cookies that are instilled by another actor, the 'ad server' and the 'individuated subject' produces a shadow (ibid:35). In this orchestration of individuals and algorithmic actors on the stage of the internet, the efficacy of 'online display advertising' comprises not only the agency of human actors but algorithmic ones.[155] This is what Introna deems the 'individuated, branded subject', who is cultivated and constructed by an elaborate 'choreography' and where 'algorithmic agency' plays an increasingly sophisticated role (ibid:47). By tracing the 'flow of agency' in humans and non-human actors, Introna states that *they not only act they also simultaneously enact*—to be more precise they are *performative* (Barad 2007; Pickering 1995; Butler 1990; cited and emphasis ibid).

In addition to facilitating traceability, processes of becoming reflect how the Impressionable Subject is enacted (Foucault 1980 cited by ibid:26), not in a merely technical manner but rather by incorporating the mechanisms of 'power/knowledge' as 'performative, material, discursive practices' (Orlikowski and Scott 2015 cited by ibid). These

> regimes of knowledge—based on historically contingent mechanisms of inscription, recording and calculation—produce the very conditions under which our subjectivity becomes negotiated, and freely taken up by us, as being this or that type of person (Foucault 1980 cited by Introna 2016:26).

The Introduction and Chapter 3 described the Kittlerian 'inscriptive' technologies, 'commands, data, address', which are the mechanisms that capture the actions of the subject and in turn positions them.

> Indeed, what makes online advertising different to other media is the diversity of methods, techniques, technologies (mostly algorithmic) for the production of a particular domain of knowledge—that in turn function to choreograph certain subjects positions, meticulously (ibid:27).

These interactions with advertisements reciprocally affect how the subject comes to understand themselves, echoing my advertisement results with Google promoting 'their books' (Chapter 5) where I recognise my personalisation through my 'critical ethnography of the self. In the flow and 'circulation of agency' (Ingold 2011 cited by ibid), 'through its hyperlink journeys and its search queries…*The impressionable subject is produced as what is being visited and what is being searched'* (ibid:37).

2.7 YOUs

In Chapters 3 and 5 I explained *how Google search works* with the construction of the

[155] As described by Bruno Latour with ANT (Actor Network Theory), there is no producer [director] as such in a 'heteroneous array of actors in the sociomaterial assemblage' (Introna 2016:28). ANT (Actor Network Theory) was introduced by Michael Callon, John Law and Bruno Latour (1977), which explores the ever-changing relationship between human and non-human actors, leading to the formation of new relations, or networks.

personalised subject, further elucidating how subjects are not actually personalised but assigned to categories of others like them, what Chun deems YOUs (2016). With the targeting and categorising of the 'impressionable subject's search habits, 'the production of YOUs, resonates strongly with Louis Althusser's theorization of ideology', which 'represents the imaginary relationship of individuals to their real conditions of existence' (Althusser 1971 cited by Chun 2016:121). Search, as conducted by these neoliberal subjects, provides data to Google Search facilitating the tracking of these YOUs. However, there is no boundary to define the single YOU because these 'associative linkages' cultivated by the interactivity of liking, friending and following, produce massive amounts of data, interconnectivity and distances. Chun applies this distance to YOUs value concerning 'networks that track users, their captured actions involuntarily send back representations of receptions' (ibid:122). Whether there is caring of the self (Foucault 1994) or caring of the network (the other YOUs), with the Interpellated Subject's interaction with platforms such as Google Search, it becomes an effect of algorithmic ideology.

> The "you" is imagined as both singular and plural. It directly addresses the viewer, but it also offers her an image of herself in a plural form'[…]This "you," which you can only become through an act of misrecognition (you aren't really these "yous"), is also a demand. By accepting the "you," a "we" emerges, but this "we" is deliberately temporary: it is brought together for a specific task now. The "we" indicates a specific moment, a now, in which "you" becomes a "we" at the moment of contact; in contrast, the "you" is curiously permanent. It precedes and remains after the "we" fades' (ibid:36).

Sorted together with others 'like them' (Chun 2016), versions of multiple 'data selves' are fed back through never-ending and recursive algorithmic loops. This intertwined relationship of 'selves'—the 'we','you','YOUs' now becomes '(u)s'—comprised of a complexity of user subjectivities where the YOU addressed by Google search is, crucially, both singular and plural. Furthermore, Chun notes that '[i]n its plural mode, though, it still addresses individuals as individuals' (ibid:64), reflecting what has come to be called the digital subject.

2.8 Digital Subject

As previously iterated, the 'data subject' used in present-day EU legislation is a juridical term for the corporal, or human subject, in this case myself. In contrast, the digital subject is something that is 'mapped onto living persons' according to media theorist Olga Goriunova (2018). This concept includes the 'subject of a data profile' embodied as 'a history of browsing or search engine queries' and is considered an 'extension of the self or person: the digital self and digital personhood' (ibid:127). Expressed otherwise, the digital subject is an abstracted persona created from various data, records and archives, aggregated together to form what comes 'after the subject, requiring new ways to understand how it connects to the subjectivities of living persons' (ibid:126). Goriunova elucidates how the little traces of identities become minimal subjectivities produced and distributed through software:

> Software takes the digital subject apart, as it is leaking in all directions but also brings it back together, linking it to one unit, or, in other words, the beginning point is "me" at the generation of data and the end point is an aggregation (2015).

Echoing Chun's interpretations of Althusser (1971), Goriunova argues that a 'misrepresentation' occurs in the distance between the subject and the 'hey you'. Furthermore, these relations are

unclear and there is always a transformation in the middle because it is computational, with Goriunova using the analogy of a Moebius strip to visualise how 'we're on one side and the digital subject is on the other' (ibid). (Figure 90) She draws on the work of Gottfried Wilhelm Leibniz to describe the relation between these two entities as 'distance', which doesn't completely disaggregate but remains thick and viscous. With distance, dots of data comprise the identities of the digital subject that becomes different things at divergent moments in time, where probability evaluates and constructs profiles, computationally (Goriunova 2018:133).

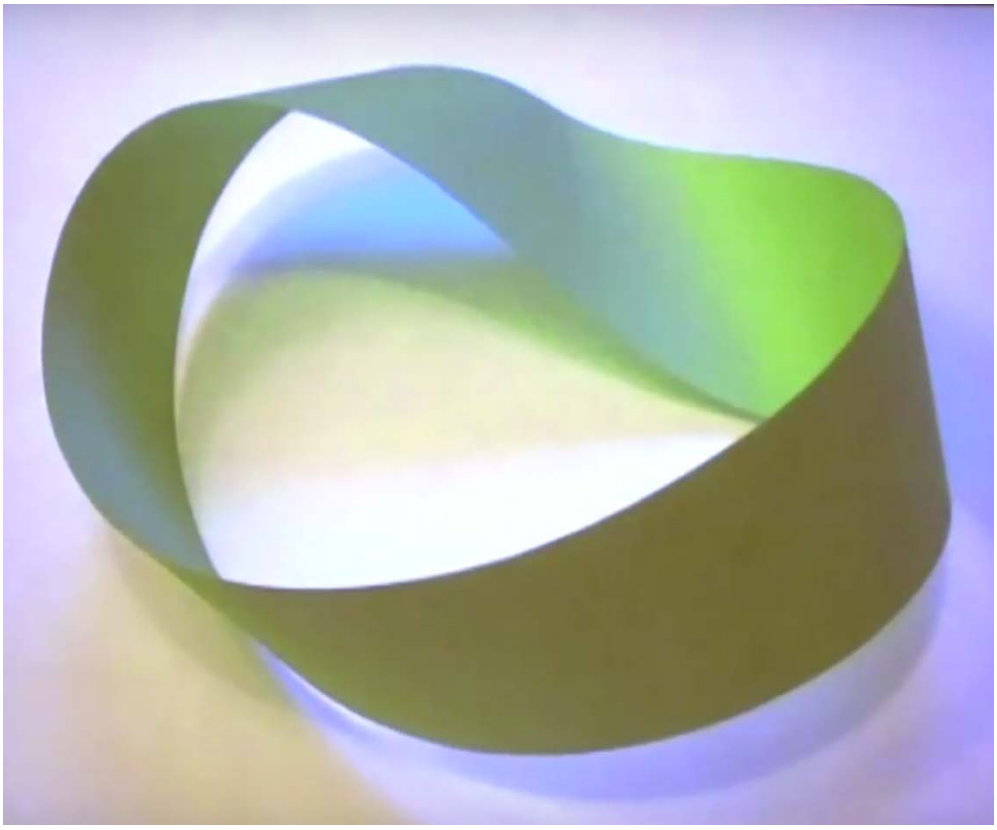


Figure 90: Moebius strip, from Goriunova lecture (2015).

> Digital subjects are values […]Digital subjects are future oriented […]they span different spatio-temporal scales: they can differ in length of alphanumerical strings, in complexity, in forms of composition and proposition, in proximity of evaluated parameters, in number of units, in frequency of occurrence, and in the types of future they propose (ibid:133).

With profiles in constant flux, temporalities and degrees of correlation supposedly enable better personalisation and '[i]t is also through the distance that digital subjects become more or less personalised or multiple—put together and disaggregated' (ibid:131). As shown by my results and the collaborative filtering practices described in Chapter 5, the data subject is not actually personalised, instead what they are presumed to be, or even will become is based on search results—what s/he previously viewed or clicked on as well as real time information about location, IP address and other browser fingerprinting. Despite the granularity of data points,

> [t]he digital subject here is not personalized; it corresponds to an aspect of a population. A digital subject, in fact, rarely corresponds to a classically constituted individual: *it is always more and less than a human* (ibid:133).

In other words, the person's 'representation' is not a representation—'data are not traces' (Rouvroy 2013; Gitelman 2013 cited by ibid:129). Whilst promised extreme individualisation, the data subject is assigned to groups of others like them in databases. 'What is important is that these digital subjects don't represent the world as there is no direct mirroring [...] they are constructed, they are not natural, they are products' (Goriunova 2015).

2.9 Data Dust

Deleuze's profound description of capitalism in the *Society of Control* is one in which services are sold and stocks bought and 'individuals have become *dividuals* and masses, samples, data, markets, or *banks*' (1992:5). This coded 'dividual' is a 'torn chain of variables' that is composed of data-points which can be 'scrutinized, calculated, circulated, assembled and reassembled for specific purposes' (Weiskopf 2018:14). Nanna Thylstrup argues that the modern human subject is produced within the interstices of a continual negotiation between the virtual and real, comprised of the bits and pieces of 'dividual' selves, the splintered figure of the subject (2014:33). In a similar way, rather than a complete composite profile of an individual, digital archival users are effectively split into two entities, their 'real' self and the user on the backend as a 'data shadow' (ibid:30). Drawing on the work of Andrew Pickering, Goriunova writes that the shadow is part of the 'representational paradigm in science' and that this 'type of idiom is an attempt to reflect nature by mirroring, yet this representation appears as *shadows* of themselves" (Pickering 1995:5-6 cited in Goriunova 2018:135). Articulated differently, what algorithms can

> collect and process will always be just a tiny fraction of information that exists in the world. It (e.g. the produced image, the representation) can only be a simulacrum of reality, like the shadows on the wall of Plato's cave (Weiskopf 2018:8-9).

Others have articulated the data shadow as a data double: 'discreet and constantly aggregated data bodies, Deleuzian dividuals, rather than the "figures and products of modernity"' (Raley, 2013:127). Antoinette Rouvroy elucidates how the data double is affected by 'indexicality' that arrives 'from elsewhere, and its production is different from the laboratory practice of science and its regimes of objectivity' (Rouvroy 2013). Moreover, Louise Amoore's research shows how the abstraction of data traces are reconfigured into 'data derivatives' (2011:27)

> The data derivative comes into being from an amalgam of disaggregated data – reaggregated via mobile algorithm-based association rules and visualized in "real time" as risk map, score or colour coded flag (ibid cited by Weiskopf 2018:17).

As 'we become our data, granulating our subjectivities in the data worlds we inhabit', it is crucial to point out however, that with certain bespoke programmes data is actualised in real time—'data is not a priori testimonial' (Goriunova 2018:135,137).

This 'real-time' collation of data on the 'data subject' a.k.a. a Trusted User of Google Search, creates algorithmically produced entities of 'Data Dust' as an object of knowledge, which is comprised of dividuals, data shadows, data doubles and data derivatives. These simulacra are not representative of real 'individuals'. Instead they are encompassing elements, bits, points—'the fragments of registered behaviour, which are extracted from the flow of data for specific purposes' (Stalder 2016). These commodities are what users pay with in the 'back-end systems of archives' or databases while the 'front-end provides empowerment' (Thylstrup 2014:33).

Additionally, these entities of Data Dust find others like themselves not in the 'meat space' apartments of urban cities, or in towns scattered across flatlands, mountains and valleys, but in worldwide data centres.

> These server farms store our Facebook friends, the articles we read, *our Google search history*, our pictures, our videos, our emails, our blogs, our social and economic activity. Every part of our online lives, that is, our lives, full stop. The dust of human beings. Fragments. Powder. This is where our vital information is calculated, where we make purchases, where we come to meet other humans (Degoutin and Wagon 2015 emphasis mine). (Figure 91)

Figure 91: 'Human dust, fragments, powder' Degotin and Wagon, *World Brain* (2015), Episode 3.

2.10 Machinelike Other

As shown previously, with computation, the holistic figure of the self has been replaced by splintered fragments of Data Dust and is endlessly divisible. These numerous entities of the subject—subjectivities—are constructed by the very data that 'personalised subjects' give away when habitually searching online.

> It could be argued that subjectivity itself, partly an effect of cultural and disciplinary individuating techniques (Foucault 2005) and partly a product of work on the self through technology, such as the diary (Kittler 1990), is formed in structural coupling with its computational environment, making all subjects digital today (Goriunova 2018:126).

Nowadays the construction of the 'digital subject' stems from the continuous production of inscribed personal data (Kittler 1999) resulting in 'new forms of subject construction that arise out of computational procedures and are employed by various forms of power to distinguish, map, and capture not only subjectivities, but also non-humans and physical things that inhabit the world' (ibid:127).

As with black box (in)visibility management (Flyverbom et al. 2016) of the 'media arcane' (Beyes and Pias 2019) explicated in Chapter 5, non-human agents observe humans and as Peter Galison notes, cybernetic philosophy includes the hidden, the non-transparent and is 'premised on the opacity of the Other'—in this case the algorithm (1994). (Appendix G) Furthermore, it was Friedrich Kittler who remarked that '[a] simple feedback loop—and information machines bypass humans, their so-called inventors. Computers themselves become subjects' (1986:258). In the 1970s, before cybernetic theory became popularised in literary science fiction novels (e.g. Philip K. Dick), the psychotherapist and philosopher Felix Guattari became predominantly interested in the emerging forms of a subjectivity that is detached, not just from the person or subject, but also from the human. According to the media theorist Eric Hörl this is specifically what undermines the notion of subjectivity that has been serialised, standardised and normalised by the mass media (2015). The philosopher Slavoj Zizek articulated the transition of subjectivity and a lack of human agency through the 'constitutive ambiguity' of the term 'mediatization' by questioning today's 'progressive computerization of our everyday lives':

> [T]he subject is also more and more 'mediatised', imperceptibly stripped of his [sic] power, under the false guise of its increase? When our body is mediatized (caught in the network of electronic media), it is simultaneously exposed to the threat of a radical 'proletarization': the subject is potentially reduced to the pure $ (the divided subject), since even my own personal experience can be stolen, manipulated, regulated by the machinelike Other (1997:65).

2.11 Cyborg

The term 'cyborg' is a portmanteau of 'cybernetic organism' originally neologised by Manfred Clynes in 1960. Clynes was a neurophysiologist and space scientist who researched the development of 'artefact-organism systems'—cyborgs—in order to free astronauts from routine flight-maintenance, so 'leaving man free to explore, to create, to think, and to feel' (Clynes and Kline 1995:31). Donna Haraway's *Cyborg Manifesto* (1991) proclaimed that humans have already been cyborgs for decades, with pharmaceutical drugs for ailments, depression and birth control or limbs replaced by prostheses and later on, in science fiction and television series, augmented with 'bionic implants'. She describes humans as cybernetic systems, such as those with heart implants and these

> [t]ransformations are effected by virus vectors carrying a new developmental code, by transplant surgery, by implants of microelectronic devices, by analogue doubles, by other means (ibid:179).

In an era of informatics, Haraway's modern coinage of the term cyborg emphasises its more hybrid forms and thereby its potential as 'a cybernetic organism, a hybrid of machine and organism, a creature of social reality as well as a creature of fiction' (1991:150). Whereas Clynes' cyborg was 'anthropomorphic', a prosthesis that 'augmented human powers' and was designed for humanistic goals, Haraway's hybrid is 'multiple, without clear boundary, frayed, insubstantial' (1991:177 cited by Parker and Cooper 2016:236-237). What is of interest to Haraway is not only the machine, nor only the human. 'The machine is not an it to be animated, worshipped and dominated. The machine is us, our processes, an aspect of our embodiment' (1991:164 cited by ibid:248).

This type of embodiment is manifested as a hybrid form that also reacts, as shown in my interaction with Google Search in Chapter 5, engaging with the machine as another actor and 'understanding how the digital is made and makes [which] thus requires recognition of its specificity and its operational hybridity' (Goriunova 2015). As explained in Chapter 4, nowadays users search with smart phones and computers as *hypomnemata*—(memory) extensions of the human body, along with devices such as RFID that are implanted into bodies, or even tiny tracking monitors reflecting the fact that 'digital hybridity is the de facto mode of contemporary existence' (Goriunova 2018:126).

Instead of an object-subject dichotomy, Haraway's cyborg is then a site of contestation, continuously questioning the 'unitary categories and hierarchical divisions of conventional form' (ibid: 242, 249). The cyborg challenges any traditional demarcation line between the human and the machine and instead defers to a space of patterning through the very technologies that comprise the database. Recombinant formation of new patterns and forms comprise this 'hybrid mediation,' with the cyborg moving at 'the edge of chaos' where it is a 'living form in the making, surviving, living-on, becoming' (Goodwin 1994 cited by Parker and Cooper 2016:243). On contemporary server farms the cyborg exists—the Data Dust of human beings, fragments and powder—as database constructions that are permeable to information, lending themselves to reconfiguration through commodification. With the disappearance of the organism, Haraway describes the commonality between 'communication science and modern biologies', as that of 'translating the world' into a 'problem of coding':

> a *search* for a common language in which all resistance to instrumental control disappears and all heterogeneity can be submitted to disassembly, reassembly, investment and exchange[…]In a sense, organisms have ceased to exist as objects of knowledge, giving way to biotic components, that is, special kinds of information processing devices (Haraway 1991:164 emphasis mine).

2.12 Cyberorganization

Articulated as the 'continuous disassembly and reassembly of new forms and patterns', without clear and delineated borders, in 1995 Haraway's cyborg had been taken one step further, recognising itself as an ongoing 'plasticity of informational patterns (e.g. databases, electronic money)' (Cooper and Law 1995:268 cited by Parker and Cooper 2016:237). Nowadays it is the concept of 'posthumanism', where fluid identities and emergent 'becomings' enable heterogeneous perspectives that centre around the human body within cultural and technological constructions. Drawing on Haraway's seminal work (1991), in *How we Became Posthuman* (1999) Katherine Hayles asserts that the digital age has altered the notion of [dis]embodiment—how information is processed and separated from that which transmitted it.

> The first idea was what Hayles called the uncontroversial or well-established notion of technogenesis: that the human species is defined by its co-evolution with various tools and technologies; or, that the inside—subjectivity—is always contaminated by the outside—technics. The second basic idea was that of distributed cognition: that cognition involves more than the neocortex, but also the body and its extended material and technological environment (Rodgers 2010).

Expressed otherwise, with the reconfiguration of subjectivities through technics and distributed cognition, 'human bodies as cyborgs—as human machine systems—are in turn systematically combined into modes of "cyberorganization"' (Parker 2000:73). With Cyberorganization 'as [a] continually shifting set of relationships' (ibid:81), the subject has become the site of data collection as well as being constantly evaluated by algorithms and 'technology, or standards, precede meaning, and enable it —similarly to how they enable the being of the subject' (Siegert 1999). As with the cyborg, Cyborganization is not an extension of the human agent but perhaps rather 'the agent is an extension of the machine' (Parker and Cooper 2016:243). The mediation of the human agent is constantly evolving in a world of information processing, where Cyborganization is the 'complicity/complexity between human agents and machines that makes continuous pattern-making possible' (ibid:242). The organisation of these patterns out of chaos is a 'never-ending process of ordering' with the becoming process part of the 'development of human cognitive systems' that reflect 'cyborganizational complicity/complexity' (ibid:238, 242). Cooper and Law emphasised the re-production of such 'unfinished heterogeneity' and in this way Cyborganization 'remains forever unfinished' because without a destination, it 'keeps on moving all the time maintaining its differences, its spaces of mimetic division', which simultaneously controls the flow of information back to users (ibid:267, 243).

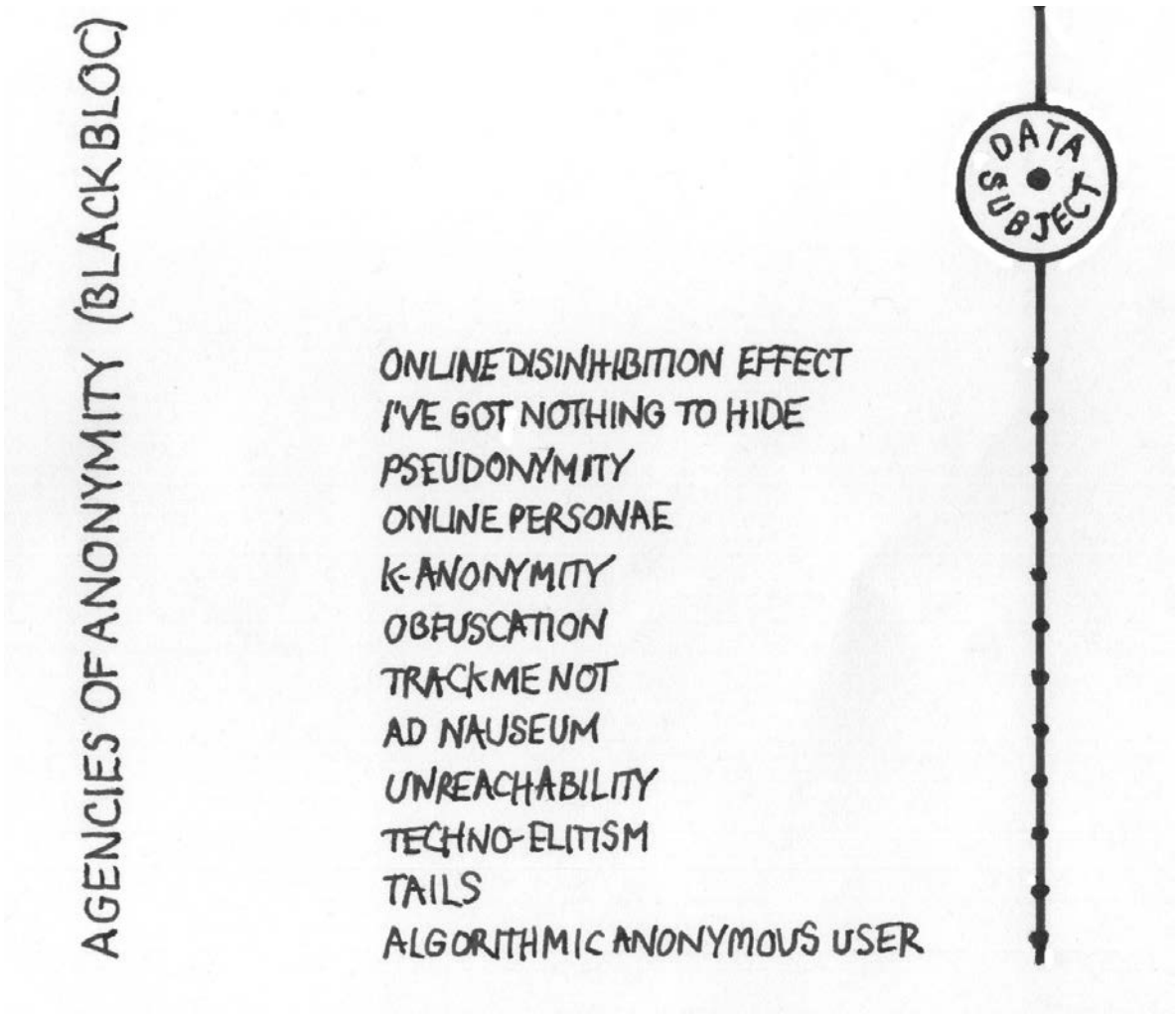


Figure 92: Agencies of Anonymity

3.0 Agencies of Anonymity (Black Bloc)

In the preceding section, I explained the effects (*Subjectivities of Search*) according to the Black Box (Google Search). I return now to the other half of Blackness, the Black Bloc. Based on my method of searching with the Tor browser and other search engines on the Dark Net, I elucidate some of the effects as 'agencies of anonymity', through a structured discourse analysis. (Figure 92)

3.1 Online Disinhibition Effect

In 2005 John Suler described what is now known as the 'Online Disinhibition Effect', arguing that online behaviour differs from 'real' world behaviour. Sometimes these online actions are reflected by gestures of human kindness, or by showing lavish gratitude or generosity.

> We may define benign disinhibition as a process of working through—an attempt to better understand and develop oneself, to resolve interpersonal and intrapsychic problems or explore new dimensions to one's identity (Suler 2005:184).

The reverse is toxic, resulting in aggression, anger, perversion and hate speech, where people act out things they would never do in the 'real' world (ibid). In other words, 'toxic disinhibition is simply a blind catharsis, a fruitless repetition compulsion or acting out of pathological needs without any beneficial psychology change' (ibid). However, instead of a binary opposition between benign and toxic disinhibition, Suler applied a system of categories to differentiate the manifold types of disinhibition. These range from adopting user names to provide 'anonymity' and the renunciation of responsibility for one's own behaviour or, with 'invisibility', concealment or the inability to physically see the reactions of others (ibid). With Suler's 'asynchronicity' there is a delay in communication response time or, with his 'solipsistic interjection', the conjuration of imaginary scenarios (ibid). Suler deems certain actions online as a game, 'dissociative imagination' in which emboldened by 'attentuated status and authority', rules are ignored and users can speak their minds (ibid).

Though speech acts could nowadays also occur using VOIP (Voice over Internet Protocol), Suler emphasises the notion of invisibility because when people communicate with each other online they mostly do so in written form, they do not see each other and thus are more apt to express themselves without inhibition. Furthermore, Suler argues that these types of interlocutions serve to 'unlock the unconscious':

> The online self becomes a compartmentalized self, a dissociated self. In the case of expressed hostilities or other deviant acts, the person can evade responsibility, almost as if superego restrictions have been temporarily suspended from the online psyche. In fact, people might even convince themselves that those online behaviours 'aren't me at all (ibid).

With this in mind I embarked on my journey with search engines on the Dark Net, deciding to conduct research on privacy and anonymity in reaction to the ubiquitous media statement: *I've got nothing to hide*.

3.2 I've got nothing to hide

A common critique from mainstream media is that having too much online privacy, in the form of anonymity, enables crimes to be committed by 'miscreants' on the Dark Net. As shown by my exploration and the results I derived employing Dark Net search engines (Chapter 7), there are anonymous uploads of texts and speech yet '[t]here is a common narrative about anonymous online speech that it is hateful. Only satisfying taboo or illegal desires' (Gehl 2014:9). In his text, *I've got nothing to hide and other misunderstandings of privacy,* Daniel Solove questions the inherent value of privacy and whether there needs to be legislation to protect it (2008:747). Solove articulated his stance in a well-known New York Times Op-Ed from 2011, *The Virtues of Anonymity*, in which he addressed surveillance,

> [t]he loss of anonymity might make many people more civil in their speech and more circumspect in their actions. That's a good thing. But it might also chill a lot of valuable expression. A world where everything people said and did was monitored, recorded and scrutinized would be an oppressive place to live.

Yet this force of revealing identities is seen as justice against vigilantes. In his 2002 article *Anonymity on the Internet: Why the Price May Be Too High,* David Davenport supports this supposition and argues against online anonymity on the basis of accountability. 'Accountability requires those responsible for any misconduct be identified and brought to justice. However, if people remain anonymous, by definition, they cannot be identified, making it impossible to hold them accountable' (Davenport 2002).

Former Google CEO Eric Schmidt made similar statements on CNBC in 2009, arguing that 'if you have something that you don't want anyone to know, maybe you shouldn't be doing it in the first place' (Esguerra 2009 cited by Van der Nagel and Firth 2015:3). Schmidt continued further, stating

> but if you really need that kind of privacy, the reality is that search engines including Google do retain this information for some time, and it's important, for example that we are all subject in the United States to the Patriot Act. It is possible that that information could be made available to the authorities (Quartiroli 2011).

The utterance of 'I've got nothing to hide' by people around the world has been one of the most common responses to the Snowden revelations mentioned in the Prologue. When asked whether spying on citizens is justified and whether citizens should be willing to exchange privacy and anonymity for increased security, the predominant rationalisation is that state surveillance is positive as it protects us from terrorism and 'I've got nothing to hide'. However, privacy and anonymity are crucial for a democratic society and the public sphere. As much as governments and certain multi-national corporations would like to agree with former Google representatives Davenport and Schmidt, there are counternarratives. Researcher danah boyd points out that people often feel immune because they haven't committed a crime, yet

> [a]ccountability is commonly raised as one of the reasons behind which people should provide identifiable information in online settings. When people prefer not to share their names, they're assumed to have something to hide (boyd 2012:30).

3.3 Pseudonymity

As demonstrated by my results in Chapter 7, pseudonyms have been used by a slew of writers, poets and novelists both in centuries past and nowadays. With online activity, they are even more commonly employed in order to protect identity and prevent the inhibition of free expression. Lisa Nakamura points out that in the early days of the web, '[t]he ability to participate anonymously or, as was and remains far more common, pseudonymously was an integral part of why Barlow and other net utopians saw the Internet as valuable' (Nakamura 2014:1). In the 1990s users had 'monikers' when they signed into chatrooms, later on they created email addresses for each different service they signed up for. As explained in Chapter 6, besides 'traffic analysis' and 'mixed networks', the basis for Tor encryption, David Chaum's other contribution to knowledge included 'digital pseudonyms', which is 'a public key used to verify signatures made by the anonymous holder of the corresponding private key' (1991:1). Chaums' idea was to apply this to 'form[ing] rosters of untraceable digital pseudonyms' from selected applications such as a 'record-keeping organization' with a unique pseudonym, or with verifying digital signatures in elections (ibid:8).

The 'Godfather of ubiquitous computing', Mark Weisner, described 'digital pseudonyms' as a means to insure privacy within networked societies. 'For example, schemes based on "digital pseudonyms" could eliminate the need to give out items of personal information that are routinely entrusted to the wires today, such as credit card number, social security number and address' (1991:8). These real-life situations or offline worlds often require a different 'hat' depending on the situation. In the digital realm, pseudonyms are handy and are put in use to manage context-specific impression management and persistent content online (Hogan 2012:304). People are able to present one part of their self to one group of people while presenting another part of their self to other groups.

> Through pseudonyms people can express their competitive urges in gaming environments, their health concerns on specialist sites, their sexual urges on pornographic sites, and their political appetites on blogs without these getting in the way of each other or personal and professional obligations (ibid).

In the contemporary online world certain moments call for complete transparency whilst other situations demand (partial) obscurity. As the sociologist Bernie Hogan notes, '[p]seudonyms are both an antecedent to this situation and also a partial solution. We may live in a global village but our huts still need curtains' (ibid). When individuals do not wish to share personal details but would like to contribute information, pseudonymity can be a possible solution.

> Pseudonyms are a practice, which is often meant to facilitate nonidentifiable content. Individuals can employ pseudonyms for one-time use or employ them as persistent alternate markers of identity (ibid:294).

3.4 Online Personae

Instances of pseudonymity do not mean that one is completely anonymous, rather various pseudonyms can be linked together to form either an 'online identity' or 'persona'. The Online

Persona was perhaps first articulated in 1984 by Sherrie Turkle in her book *The Second Self* with her leitmotif of the Rorschach metaphor, which 'had come alive in a new social practice: the virtual world as context for explorations of identity' (Turkle 2005:288). In the updated MIT Press version from 2005 she describes how the virtual world is a new environment for experimentation.

> In online life, computer users cycle through personae, cutting across "real life" distinctions of gender, race, class, and culture. The online personae we write into existence function as evocative objects: the virtual can be used to reflect constructively on the real. Even a gesture as simple as choosing an online name can be fraught with implications (ibid).

Geert Lovink's notion of the *Synthetic Self* builds upon Turkle. He articulates the second self as a parallel universe identifying a 'cult of the self', along with how it is constructed online (2017).

When one experiences the impulse to create a fictitious identity it is likely a humorous gesture or a discreet reference to a real-life existence. One reason to build these constructed online personas is to encode digital identities, which serve a particular 'use-value' depending on the situation as they 'also offer sites of reinvention, liberation, and play. Fake accounts and performed identities testify to that' (Goriunova 2018:127). Persona building goes hand in hand with crawling the Dark Net in that it is replete with pseudonyms and false identities. Returning to my analysis of DWSN in Chapter 7, Robert Gehl described how to circumvent capture in regard to his ethnographic study, where 'one would have to create a new account every time one used the site, taking care never to link statements one makes to a coherent pseudonym' (2014:1226). In attempts to reach 'total anonymous freedom' DWSN members 'customise their aliases, avatars, pseudonyms through widgets and in such a way that [o]ne does not use a fake account every time; one builds a *persona*' (Gatson 2011:232; Marwick and boyd 2010 cited by ibid). In this way users can also be empowered to construct their own identifies, instead of user data being structured for marketing purposes, as discussed in Chapter 5 with personalisation and in Chapter 6 in relation to the 'nymwars'. When all of these pseudonyms are collated to construct personas, together they can form an online profile. However, there are tactics to prevent this from happening.

3.5 k-anonymity

Latanya Sweeney begins her article *k-Anonymity: A model for protecting privacy* (2002) with the observation that '[s]ociety is experiencing exponential growth in the number and variety of data collections containing person-specific information as computer technology, network connectivity and disk storage space become increasingly affordable' (Sweeney 2002:557). She further goes on to introduce the situation with how 'data holders' (companies, institutions, agencies) 'operate anonymously' and have to release (anonymous) data in order for the continuation of the database, yet 'failing to provide proper protection within a release may create circumstances that harm the public or others' (ibid). When exposing data publicly, all specific identifiers need to be removed (name, address, telephone number), making the leftover data appear anonymous. For example, this data conundrum applies to US Census summary data where demographic values could be combined with geographic locations, structured as fields of information or attributes such as zip code, gender and data of birth that, if released, would not be considered anonymised. Therefore, Sweeney designed a protection model, *k-anonymity*, which

would alter the released information based on scaling, where the greater number of candidates, the 'more ambiguous the linking, and therefore the more anonymous the data' (ibid:559).

Following Sweeney, multilevel relational databases (MDB) allowed data storage at diverse security levels, thereby enabling various security clearances (2002). Formerly 'aggregation interference problems' could be 'solved by database design' and dataholders operated anonymously without considering other sources of data (ibid:560). Another anonymising technique is to add 'noise' but sometimes this destroys the integrity of the records. The main method used to control the flow of data is called 'suppression', where a conscious decision has been made not to release sensitive data. However, '[s]uppression can drastically reduce the quality of the data, and in the case of statistical use, overall statistics can be altered, rendering the data practically useless' (ibid:560-561). Additionally, 'querying restriction' is a common solution that prohibits queries that can reveal sensitive information. For example in health databases, querying 'itemizing medications' and then 'associating patients with their prescribed medications may be sensitive and even deanonymise data because medications typically correlate with diseases' (ibid:561). In other words, preventing what is called 'disclosure— e.g 'explicit or inferable information about a person was released that was not intended' through a method of 'disclosure control' attempts to limit the release of identifying data or to ensure that what is released is 'sufficiently anonymous' (ibid:563).

3.6 Obfuscation

Analogous to the increased strength of anonymity with a larger amount of users with the Tor browser, the efficacy of 'scaling' tactics with *k-anonymity* are nowadays limited due to fact that the storage of data continues to become easier and cheaper as computational power increases and companies can examine this information in 'real time'. As indicated in Chapter 5, with regard to recommendation systems and customisation, the preferences of users are correlated with other peoples' data and retained indefinitely. Moreover, it is uncertain what will happen in the future with search data with machine-learning algorithms, how that data could travel and whether, with new techniques of analysis, other patterns could be revealed that were never anticipated (Brunton and Nissenbaum 2015). In regard to the problems above, *Obfuscation: A User's Guide for Privacy and Protest*, by Finn Brunton and Helen Nissenbaum is a response as to why obfuscation is necessary in an era of online tracking and argues that its use-value is in 'mitigating and defeating present-day digital surveillance' (2015:1).

According to the authors, obfuscation, a cryptic term that needs some deciphering, is 'the deliberate addition of ambiguous, confusing, or misleading information to interfere with surveillance and data collection' (ibid). Connecting back to Sweeney's *k-anonymity* techniques, '[o]bfuscation, at its most abstract, is the production of noise modelled on an existing signal in order to make a collection of data more ambiguous, confusing, harder to exploit, more difficult to act on, and therefore less valuable' (ibid:46). Moreover, these tactics prevent various shields of recognisability—how signals or information could be interpreted, transmitted or shared by enemy parties. (Figure 93) The authors chose obfuscation 'because it connotes obscurity, unintelligibility, and bewilderment and because it helps to distinguish this approach from methods that rely on disappearance or erasure' (ibid). Applying tactics of secrecy and deception to combat asymmetrical relationships of power, they show various methods of resistance that, while they might be considered 'weapons of the weak', afford some amount of autonomy for the user (ibid:55).

Figure 93: *Dattoo* (2015) by KABK students Christina Yarashevich and Janne van Hooff is a digital mask that provides anonymity through obfuscation. Image credit: http://www.onlineopen.org/index.php

Their publication is not just a scholarly book on present-day obfuscation with the Tor Browser, it also documents historical obfuscation practices (30 cases). They also address the contemporary quandary of how information is already distorted by the powers that be, the ethical and moral complexities of censorship and the manifold reasons why obfuscation tactics are necessary in an age of privacy erosion. Obfuscation can bolster an 'existing strong privacy system, in covering up some specific action, in making things marginally harder for an adversary, or even in the mere gesture of registering our discontent and refusal' (ibid:58). Harkening back to the scholarship of Gary Marx in Chapter 7, along with foreshadowing the concept of 'contextual integrity' (Nissenbaum 2010) that I will discuss in Chapter 10, Brunton and Nissenbaum state that there needs to be a balance between user consent and refusal that determines where and in what situation obfuscation is appropriate to enable privacy or even, in certain cases, anonymity (2015).

3.7 TrackMeNot

In 2005 it was reported by journalists that in 'anonymized search query logs', 'the identities of certain searchers had been extracted from personal information embedded in search terms' (Nissenbaum and Howe 2009:431). As a reaction to the public realisation that search companies (notably Google) were logging, storing and analysing the search query logs of individuals Helen Nissenbaum and Daniel Howe produced a Firefox browser extension, 'TrackMeNot' (TMN), 'designed to achieve privacy in web search by obfuscating users' queries within a stream of programmatically generated decoys' (ibid: 417;420).[156] Unlike Tor and proxies for anonymisation, TMN provides 'protection against aggregation and profiling of individual search queries' (ibid:429). The authors state that there are at least four mechanisms that enable a search query to be tracked by to a user:

[156] http://trackmenot.io/

> (1) identifying information included in search queries (name, zip code, phone number, Social Security number, etc.), (2) *IP addresses linking searches across sessions*, (3) explicit login to search engines (often for mail or other services), and (4) persistent cookies linking any of the preceding items to users' search activities (ibid:426 emphasis mine).

As indicated by my search results in Chapter 5, with locative data and personalisation the IP address plays a crucial role in the identifiability of the user (myself) along with the ever-increasing amount of data collected and shared between search engines and third parties.

Since 2006 TMN 'protects Web users against data profiling by simulating HTTP search requests to search engines with queries extracted from the Web', along with human-like behaviour simulations (ibid:421). Various 'technical mechanisms' are applied to mimic the search behaviour of users, such as 'dynamic query lists (with RSS-based initialization), real-time search awareness, live header maps, burst-mode queries, and selective click-through' (ibid:418). In other words, TMN enables automatic dummy requests simultaneously to 'a user's real search query' (Hermann et al. 2014:1), thereby obfuscating the user's real searches to instead produce camouflage. TMN is a direct response to the tracking and surveillance of search engines such as Google along with placing agency in the hands of the user by (re)configuring the 'malleability of IT and the openness of network protocols' and these 'political commitments' can actually attempt to 'mediate transactions in politically charged ways' (Nissenbaum and Howe 2009:431). By changing the relationship of the user to their search queries and providing affordances of agency, TMN embodies what Gary Marx describes as a 'practical means of resistance' (1999), 'where individuals take advantage of the blind spots inherent in large-scale systems of surveillance' (Nissenbaum and Howe 2009: 431).

3.8 AdNauseum

Although historical accounts are endless, modern day technological applications also exist that intentionally divert the attention of not only the user but of the search browser. Another 'tactical media' project from the designers of *TrackMeNot* is *Ad Nauseum*, a browser extension that addresses the lack of standards for tracking, privacy issues, user profiling and 'excessive universal surveillance' (Nissenbaum, Howe, Zer-Aviv 2014). By flooding Google search with false queries, *AdNauseum* clicks and likes *all* ads, concomitantly visualizing the ads over time.[157] In this way it leverages obfuscation with various search engines, including Google, by clicking on ads in the background to automatically pollute the database and to decrease profiling and profit. Firstly, it collects the ads in a 'vault' where they can be studied and analysed by the user, just as they are studied and analysed by algorithmic profiling. Secondly, by simulating clicking on the ads, it interferes with and 'confuses trackers' thereby decreasing the value [and profits] of the 'aggregated tracking data' (Howe and Nissenbaum 2017:57).

In January 2017 Google banned it from their store, stating that it violated their Terms of Service because 'an extension should have a single purpose that is clear to users' and even disallowed manual updates, preventing users from accessing their own data (ibid:60). Howe and Nissenbaum responded that '[t]he single purpose of *AdNauseam*, we would argue, is quite clear—namely to resist the non-consensual surveillance conducted by advertising networks, of which Google is a prime example (ibid). Public response followed, on sites like Ycombinator,

[157] https://adnauseam.io

with comments pointing out the power of Google, which runs both the browser (search) and the ad network and that the ban was not surprising.[158] (Figure 94)

Figure 94: Ycombinator thread 'Hacker news'

Howe and Nissenbaum admit that the aim of *AdNauseum* is 'to disrupt business models that support surreptitious surveillance' and responded to their critics by questioning the 'real name' policy demands of Facebook and Chrome (Howe and Nissenbaum 2017:63-64), once again harkening back to the 'nymwars' mentioned in Chapter 6. In this way advertisers [and Google] are able to 'exploit our most human endeavors (sharing, learning, searching, socializing) in the pursuit of profit' (ibid:64). Moreover, Howe and Nissenbaum believe that 'effective privacy protection must be infused throughout a system' (2017:57) and although obscuration offers many shades of privacy, a 'true inner self' can only emerge in anonymity (Brunton and Nissenbaum 2015:45).

3.9 Unreachability

One could argue, for a variety of reasons, why users should value (offline) anonymity as I demonstrated in Chapter 6, and whether a person is identifiable (or not) as themselves, if they are untrackable or even 'unreachable' (Nissenbaum 1999). In past centuries people moved to urban centres to find work but also to reinvent themselves, as I described in Chapter 1 with the address offices. 'People strolling through a foreign city are anonymous because no-one knows who they are' (Nissenbaum 1999:141). Helen Nissenbaum's text, *The Meaning of Anonymity in an Information Age* explores what underpins anonymity in a networked, computerised society and answers three questions: 'Should anonymity be protected in electronic interactions and communications? Would this be a good thing for community, responsibility, free expression, political participation, and personal fulfillment? If so, when and why?' (ibid). Continuing further, Nissenbaum explains the necessities of why users need anonymity in the information, which concurs with my findings from 'Anonymous Users' on the Dark Net and Bartlett and Gehl's ethnographic studies (2014).

> For situations that we judge anonymity acceptable, or even necessary, we do so because anonymity offers a safe way for people to act, transact, and participate without accountability, without others "getting at" them, tracking them down, or even punishing them. This includes a range of possibilities. Anonymity may encourage freedom of thought and expression by promising a possibility to express opinions, and develop arguments, about positions that for fear of reprisal or ridicule they would not or dare not

[158] https://news.ycombinator.com/item?id=13327228

> do otherwise… Anonymity may also provide respite to adults from commercial and other solicitations (Nissenbaum 1999:141).

The crux of Nissenbaum's article concerns the shift from the natural definition of anonymity, the ability to carry out actions 'namelessly' in offline space, to how users can conduct themselves anonymously online, through the concept of unreachability. In 1999 Nissenbaum articulated the complexity of the computer-oriented society, where being nameless 'is not as the end in itself of anonymity' but the 'traditional means' to obtain unreachability (ibid:143) and where its absence can be used for positive expression, not just criminality (as shown in Chapter 7). However, in the digital sphere the issue at hand does not necessarily concern what is named and what is not 'because although it preserves a traditional understanding of anonymity, it fails to preserve what is at stake in protecting anonymity' (ibid). The larger argument for real-name 'policing' is that anonymity and pseudonymity are used for questionable or even illegal activities and behaviours online. Echoing the previous sections, mainstream media purports that hacking, doxing, flaming and trolling are made possible by being 'anonymous' without accountability (Davenport 2002).

Although a simplification, Hogan retorts that '[a]nonymity is a state implying the absence of personally identifying qualities' (2012: 294). Moreover, there is a degree, or sliding scale, moving from real names through pseudonyms to anonymity as they 'work in tandem' and 'anonymity and pseudonymity are not neutral states' (Van der Nagel and Frith 2015). Additionally, Nissenbaum presciently predicted the era of 'surveillance capitalism' (Zuboff 2015), where every atom of data is collected, kept and analysed, which I will address in Chapter 9. 'Information technology has made it possible to track people in historically unprecedented ways. We are targets of surveillance at just about every turn of our lives' (Nissenbaum 1999:141). This technological condition extends its tentacles, grasping at and identifying users who should paradoxically aspire to a state of Unreachability.

> [T]he value of anonymity lies not in the capacity to be unnamed, but in the possibility of acting or participating while remaining out of reach, remaining unreachable. In other words, this unreachability is precisely what it at stake in anonymity (ibid:143).

3.10 Techno-elitism

According to Robert Gehl, the Dark Web Social Network (DWSN) is an experiment with power and freedom through anonymity and infrastructure, going beyond imposed limits (2014:1231). As discussed in Chapter 7, in my searching of the 'Dark Net' I sought to find the DWSN site without success, which these 'technical elites' consciously protect through tactics of obscuration. 'Clearwebbers' cannot 'reach' it and this 'unreachability' reflects an acquired knowledge—'tech savvy'— that enables DWSN users to transcend the boundary of the clear web into the Dark Net of Tor's Hidden Services (a.k.a. Onion Services). Spurred by agencies of 'control' over one's data, it is a type of self-determination—knowing how to obtain anonymity and remain hidden from search engines and governmental agencies. Phrased differently, it is because of the technical savvy of its members that they have been able to arrive at the Dark Net/DWSN in the first place. Yet Gehl also points out the hypocrisy that motivates DWSN: it wants to produce an alternative social network that is open to others who want to escape from the clearweb's practices of surveillance and data collation yet it remains selective—only those with a certain skillset can reach it (ibid:1228).

> Through techno-elitism—that is, through the desire to keep the site hidden to those with the technological "eyes to see"—DWSN members and administrators establish a power relationship between themselves and those on the clear web: We are technical elites, and you are hapless clearwebbers caught in the nets of the NSA (ibid).

Although Gehl points out the productive aspects of techno-elitism (coding, hacking, DIY attitude), DWSN mimics the clearweb with its inherent power constructs, which inhibit its usage by the general public and cultivates a crowd of users who possess the know-how to use it. Like many initiation rituals, from college frats hazing to rights of passage, these members undergo a series of trials in order to reach the destination—DWSN. Gehl shares one quote in particular by a member cum administrator who confirms this, noting that

> there is also the theory that 'The teacher will appear when the student is ready'. [New members] might make it to [DWSN] by google search but accessing the URL via TOR [requires] effort and in my opinion, [is]the admission test (ibid:1229).

In other words, to join and to help build the DWSN community one first needs to have the tech savvy to gain access to it (which I didn't). Controversy exists on DWSN because some 'techno-elitists' think that only certain types of searchers should be able to reach it whilst others are more open to 'clear webbers', should they be able to find the site. [159] However, having Google make any content from the DWSN appear on the 'clear web' would make it too easy for novice users to discover the site. As explained in Chapter 7, there are those 'techno-elitists' who do not wish to be indexed by Google and desire to remain 'unreachable' to search engines.

3.11 TAILS (The Amnesiac Incognito Live System)

My (RR) interview with an Anonymous Hacker (AH)

RR: What makes Tor unsafe?

AH: When you use Tor you are just a client. But the exit nodes are a real problem. We do not know who is running the servers of these exit nodes. They could be anyone in the world, also governmental officials, FBI, CIA, SIS, M16, etc.

RR: Can I be anonymous on the internet?

AH: There is no way to be anonymous on the internet actually. Or, if you would be anonymous, it would be temporary and it would cost much effort and money. If you wish to be anonymous you would need to hack a wireless network somewhere, anonymously, by sitting in a car in the street for example. The computer or device you are using needs to not be registered to you, or purchased by you because its MAC (media access control) address is traceable. Every device has a MAC address, but there are ways to remove it. After using the internet for whatever you

[159] One admin noted that the DWSN 'is a living thing now, let's see what happens:)', implying that some 'clear web' coverage might be tolerable. Another admin, however, bluntly stated 'I'd prefer it just be deleted and disappear.' Another member expressed both views: 'i'm only here [since] a few days, need to find my way a bit and I think a bit more users to make [the DWSN] more "alive" would be cool … but with more coverage on the clearnet, who knows what kind of people start visiting' (ibid).

want to do you would then need to destroy the computer or get rid of it in some way, pass it on, knowing full well that you have been able to be tracked. Nowadays the way you type, how long it takes, rhythm, keystrokes, (e-biometrics) are also personally identifiable.

RR: What is the best way you know of to be anonymous on the internet at this moment if I cannot carry out what you describe above?

AH: TAILS (The Amnesiac Incognito Live System) is an operating system that is installed on a USB stick that you boot with your computer.[160] Using TAILS that has Tor already configured complicates things a bit so you are harder to track but the good news is that everything is deleted afterwards. TAILS is designed to leave no traces on your computer. If you do want to save something you should either back it up on another device, like a USB stick, or a DVD or send it through the internet (always tricky, depending on whether you wish to have the information compromised). Saving webpages, taking screenshots, etc. for your research would only work if you set yourself up with admin account and deliberately save them on the computer you are using, but then you compromise the whole purpose of using TAILS for anonymity.

3.12 Algorithmic Anonymous User (Random Darknet Shopper)

During my research on the Dark Net I discovered the tactical media project *Random Darknet Shopper (2014-ongoing)*, where a programmed bot autonomously searches the Dark Net and carries out purchases, completely anonymously, of course, albeit through code.!Mediagruppe Bitnik describes the project on their website:

> *The Random Darknet Shopper* is an automated online shopping bot which we provide with a budget of $100 in Bitcoins per week. Once a week the bot goes shopping in the deep web[sic] where it randomly chooses and purchases one item and has it mailed directly to the exhibition space. As soon as the items arrive they are unpacked and displayed, each new object adding to a landscape of traded goods from the Darknet.

The exhibition *The Darknet - From Memes to Onionland. An Exploration* displayed artefacts ranging from guns, baseball caps with hidden cameras, a book on 'The Art of French Cooking', a list of email addresses, a British gas bill, two USB bitcoin miners, Viagra and even ecstasy pills obtained from the former Dark Net marketplace *Agora,* which was shut down in August 2015. When the ecstasy pills arrived at the Kunst Halle St. Gallen, in January 2015, even the Swiss police took notice of contemporary art.

The Random Darknet Shopper reflects the genre and spirit of Mail Art, an ongoing conceptual art movement that utilises mail systems as a distribution model in the postinternet era. According to a review in the Guardian, the artist Weisskopf from !Mediagruppe Bitnik declared that'[t]he arts should be able to mirror something that is happening in contemporary society in a contemporary way' (Power 2014). However, this raises an important question, can one prosecute and punish an algorithm that commits a crime, in this case purchasing illegal substances and sending them to the Kunst Halle by post? The Swiss artists *!Mediengruppe Bitnik*, Carmen Weisskopf and Domagoj Smoljo state:

[160] See end of Chapter 6 for a more detailed explanation of TAILS.

> We are the legal owner of the drugs – we are responsible for everything the bot does, as we executed the code. But our lawyer and the Swiss constitution says art in the public interest is allowed to be free (ibid).

!Mediengruppe Bitnik and their lawyer are using the same legal argument which has kept some artists alive (or at least out of jail) with the defence 'it's *just* art.' This echoes Jean-Luc Godard's statement regarding representation, 'this is not a just image, this is just an image' (1980).[161] The programmers cum artists concluded with a response to the future searching of the Dark Net and decentralised systems, asserting that 'dark markets' have always existed offline or online and that they are here to stay:

> People have realised [with bitcoin] that money is not an absolute. They realised they could shape it. They could create their own things with maths, P2P networks, decentralisation and cryptography. Whether Tor survives or not, you will soon be able to run darknet nodes on your own computer, which can't be taken down (Power 2014).

4.0 Collaborative Collectives

I return now to my 'experiment in living' and to my results from comparative searching using Google's personalisation (Chapters 4 and 5) and Tor's anonymity, which I explained in detail in Chapters 5 and 7. From my results I learned that Tor delivers 'Google-like' results with its default search engine at that time (Disconnect), only without personalisation and targeted ads because there is no locative data or search history captured by my IP address. However, if I assume that both is the case—on the one hand, I am assigned as a Tor user and on the other that Google assigns me to groups of people like me (an assumption I cannot fully prove with my experiment but is the most likely scenario to explain its outcomes) — the original framing of my experiment has to be specified. Instead of a personalised search as opposed to an anonymised search, I would have, in fact, conducted a Google search that is collective-via-users-like-me on the one hand, versus on the other hand, a search that is collective-via-all-Tor-users. At stake, therefore, are two collectives that take different *forms*.

In the collective-via-users-like-me-search it is Google's algorithms which construct the collective I am part of and, as I elucidated in Chapter 5, assign me to this or that collective (e.g. collaborative filtering). I have no access, no knowledge and no agency in regard to the collectives which I am made part of via Google. The forces (identification markers: IP address, search histories, cookies, supercookies and locative data) that sort me into a collective and the collectives that I am organised into, i.e. the categories that Google sets up and assigns me to, are not transparent to me. Moreover, Google collects my individual search activities and, in future scenarios, will probably state that they 'personalise' search even further based on data collated in the past and present.

Tor's collective, on the other hand, is both an anonymity network *and* a browser—Tor is mostly Firefox code (95%) that incorporates patches to Firefox ESR (Extended Support Release) and facilitates anonymity. The Tor collective is at least partially known to me as university labs worldwide run the major nodes but I do not know who is running the relays (it is an anonymised network). I can, however, look at the 'exit address' list, which is constantly updated and shows

[161] *Ce n'est pas une image juste, c'est juste une image*. Colin MacCabe *Godard: Images, Sounds, Politics* (1980)

the IP address, though I cannot identify them.[162] By joining the Tor collective I decided to trust the exit node operators, also in regard to my 'expectation of privacy', as discussed in Chapter 7. The key difference is that whereas Google organises me into particular collectives through their non-transparent process of collaborative filtering, I decided to be organised into the 'anonymous Tor collective'.

Both search collectives, the one determined by Google's algorithms as well as the one created by the choice to use Tor, add to specific filter bubbles. The filter bubbles are, however, structurally different: in the case of the bubble produced by Google's algorithms, Google collects the data of its users and incorporates user feedback into subsequent search results. When I search for different things, I am merged into different clusters with *other* people 'like me' or 'YOUs' (Chun 2016). I would then add to the feedback loop by continuously adding to my own personalisation by clicking on the links that are delivered to me as results. I do not have access to the Google collective itself—I am constantly switched into a different cluster by an algorithmically organised process that I have no control over and cannot adjust. There are also constant updates and tweaking of signals being carried out on the algorithm, as I showed in Chapter 5. Therefore, the results of my small data study propose that various degrees of Google Search's personalisation organise the corporal 'data subject' into effects: *Subjectivities of Search.* With the *Black Box*, as Google's personalisation increases so does the amount of computational agency, at the same time the degree of anonymity (privacy) decreases.

The filter bubble of the Tor users, on the other hand, is one where I stay in the same group that shares the same filter—no matter how much I change my search behaviours (what I click on or not). As a Tor user, the variable is what Tor uses as their default search engine (Startpage, Disconnect Search or presently DuckDuckGo) and if this default (still) delivers Google Search results without locative data. Therefore, the results of my small data experiment postulate that the user is assigned the category of 'Tor user', which can be seen from outside the Tor network. Unlike Google search, privacy-enhancing technologies and diverse settings enable the user to organise themselves into effects: *Agencies of Anonymity*. With the *Black Bloc*, as the amount of anonymity (privacy) increases with Tor so does human agency, at the same time the degree of personalisation decreases. (Figure 95)

When I use Tor I am part of an p2p anonymity network, which increases in strength the more users use it. Exactly and only because I am anonymous and unknown, I have a small voice in a choir of the manifold decisions that make up the p2p-collective of Tor users, whereas I would lose this voice if I were to join the constant flux of the algorithmic clustering of personalisation. With the *individualisation* of the pseudo-autonomous objects of Google's personalisation, my 'data dust' is atomized and fractured; as a 'digital subject' I have no agency to decide where I am assigned. To partake anonymously in a p2p-collective *individuates* me more than personalisation does. Bernard Stiegler's entire pharmacology of care revolves around this new ecology, which can be understood to be threefold, as a 're-articulation of psychic, collective and technical *individuation*' (Hörl 2017:35). At stake is an *individuation* in the sense of Stiegler's reading of Simondon—an *individuation* that is marked by being collective and psychic alike. According to Simondon the genesis of how an object comes to exist is through operations of *individuation,* or 'ontogenesis' and a living being exists in a state of becoming between individuations, never in isolation but in collective, social as well as psychological constellations.

[162] Available here: https://check.torproject.org/exit-addresses

The Tor browser (p2p network) embodies this individuation of collective singularities, which are dependent on the other.

> After all, an individual exists and is only capable of individuating as a result of the relations it establishes with others and that others establish with it (Stiegler 2007:77).

Phrased otherwise, the possibility of choosing the Tor Browser collective over the *individualisation* of Google Search might be a modest resolution to the problem of 'how one would define a singularity that could be a collective singularity' (Lazzaroto 2012:14).

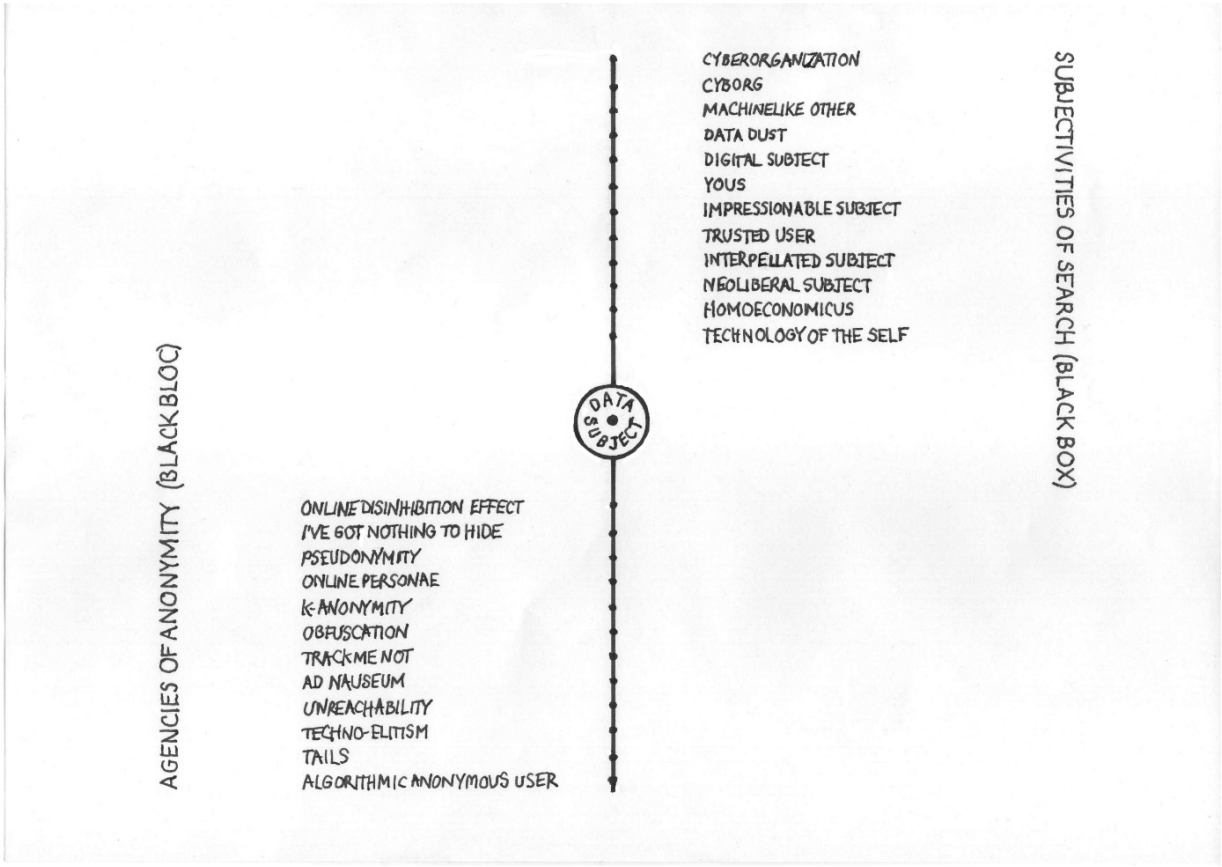


Figure 95: Agencies of Anonymity vs. Subjectivities of Search

The experience of setting up my 'experiment in living' has opened up a *view* on how Google Search works and my exploration of the Dark Net using Tor (and TAILS) reimagined what search could look like. Without having to become a 'personalised subject', Tor offered me 'relevant' search results as an 'anonymous user'. With the Tor Browser I am not commanded (Franklin 1989) as I am with Google Search, instead I chose which collective I wished to be part of, in this case Tor. The effects of the Black Bloc (*Agencies of Anonymity*) enabled me to intervene with my 'tech savvy' and this type of searching enabled degrees of anonymity, embodied by various levels of user agency. Aside from its other merits, Tor is also one, albeit not the only, strategy to challenge the protocols facilitating Google's 'surveillance capitalism' (Zuboff 2015), which I will explain in following chapter.

Then, slowly, we incorporate the whole notion of systems: we'll link registration data to surveillance, to contract compliance, thought the director. Finally, we arrive at tautology: the data prove the need for more data! We think we are creating the system, but the system is also creating us. We build the system, we live in its midst, and we are changed (Ullman 1997:91).

Chapter 9:

Black Utopias: Surveillance Capitalism

1.0 Black Utopias

In Chapter 1, I explained how analogue address offices in European cities organised information and thereby can be considered the 'first' search engines (Tantner 2015). With their registers or 'Protokollbücher' these address offices often served as a 'Borgesian universe', offering a 'long tail' of products and service, and, as a platform of capitalism, constituted a 'virtual marketplace' (Blome 2010) with printed advertisement journals. Yet the analogy between modern search engines and their historical antecedents also hinges on how personal data is collected, managed and organised (Tantner 2015:273). As the address offices collated more and more data in the registers, they also took on an authoritarian function, such as the Fielding Brothers' 'Intelligence Offices', which profiled clients and servants for surveillance and policing that in turn affected their behaviour. Tantner describes how these 'Black Utopias' were sometimes met with scepticism by residents who were concerned about their dual agenda of registering people coming to the city whilst acknowledging that an address makes them not only searchable but findable. By possessing a registered 'address', people were identified and contacted, in a manner comparable to contemporary online querents who are identified by their IP (Internet Protocol) addresses, as shown by my results in Chapter 5 (Authority, Authenticity and Authorship). With particular reference to my effects Subjectivities of Search, as discussed in Chapter 8, in this chapter I address my main research question: *how does Google Search organise (us)sers?*

I begin with Shoshana Zuboff's article *Big Other: surveillance capitalism and the prospects of an information civilization* (2015), in which she elucidates how Google's 'logic of accumulation' regarding data, or 'behavioural surplus', enables 'surveillance capitalism'. During the course of this chapter I interweave her discourse analysis of Google Chief Economist Hal Varian's documents with my own inquiry and the results from previous chapters along with Tantner's notion of 'Black Utopias'. By depicting the technological developments and data structures that make surveillance capitalism possible, I demonstrate how Google simultaneously organises (us)ers in the process. Furthermore, I show how Google transitioned from organising the world's information and making it accessible to how, and to whom, it makes it 'useful', with its 'pervasive digital tracking' (Christl 2017) of (us)ers.

2.0 Logic of Accumulation

'It's not about the information in the world but the world's users' information' (Stalder and Mayer 2009)

I introduced this thesis by referencing Michel de Montaigne, who extolled his father's conception of Europe's first address office and his documentation of household minutiae (oikos) (Tantner 2015). These events were captured in daily journals or ledgers by applying memory tools, or *hupomnemata*, which I discussed in Chapter 4 with my 'critical ethnography of the self'. With the contemporary logging, capturing and documenting of one's life through data collection tools such as the computer, this Technology of the Self (Chapter 8) reflects how the data subject becomes a personalised subject by interacting with search algorithms in 'games of

truth' (Foucault 1994). In her 2015 article, Shoshana Zuboff elucidates how the 'logic of accumulation' of user data in regard to computer-mediated transactions became the basis for Google's business model. Drawing on her previous text from 1981, *The Psychological and Organizational Implications of Computer-Mediated Work*, she explicates how 'computer mediation' differed from the automated and mechanised forms of industrial labour in centuries past.

According to Zuboff it was not just about replacing the human body with machines; with information technology there is an additional transparency to both the process of imposing and producing information (2015:76). This ability incorporates the past reflection of activities 'because it has the unique capacity to simultaneously *automate* and to *informate*' (ibid). Zuboff also describes how computer-mediated work extends to what she deems 'organizational codification', 'resulting in a comprehensive "textualization" of the work environment', where the flow of work is captured by various technologies, as a new division of learning *from* work, alongside traditional divisions of labour (ibid). Besides learning from how workers work as a result of the informating process, computer-mediated work extends the organisational codification of 'real-time' data collection and management systems into what she calls the 'electronic text' (ibid). This, in turn, is organised by the 'logic of accumulation in which it is imbedded and the conflicts inherent to that logic:

> [It] organizes perception and shapes the expression of technological affordances at their roots. It is the taken-for-granted context of any business model. Its assumptions are largely tacit, and its power to shape the field of possibilities is therefore largely invisible. It defines objectives, successes, failures, and problems. It determines what is measured, and what is passed over; how resources and people are allocated and organized; who is valued in what roles; what activities are undertaken—and to what purpose (ibid:77).

As recounted already with the search subjectivity, Machinelike Other (Chapter 8), Zuboff's *Big Other* text (2015) is ostensibly a reference to the reality TV show which features human contestants tracked by cameras 24/7 (itself referencing Orwell's *Big Brother*) yet here the government is replaced with Google. Zuboff conducts a close reading of what she deems two 'extraordinary documents', written by Varian (2014, 2010), which 'disclose the logic and implications of surveillance capitalism as well as "big data's" foundational role in this new regime' (Zuboff 2015:76). Zuboff proposes that 'big data' is not a technological object or an effect or an autonomous process nor a thing in itself but rather it is 'embedded in the social' and is 'both a condition and an expression' of surveillance capitalism (2015:77). Furthermore her argument is structured around Varian's 'new uses' stemming from computer-mediated transactions: 'data extraction and analysis,' 'new contractual forms due to better monitoring, ''personalization and customization,' and 'continuous experiments' that 'provide insight into the logic of accumulation' (Varian 2014 cited by ibid:78).

Zuboff first points out that Varian divulges the 'pervasive and continuous' recordings of computer-mediated economic transactions, which facilitate 'rendering an economy transparent and knowable in new ways' (ibid:78):

> The computer creates a record of the transaction … I argue that these computer-mediated transactions have enabled significant improvements in the way transactions are carried out and will continue to impact the economy for the foreseeable future (Varian 2010:2 cited by ibid).

In Chapter 4 I discussed Foucault's observation of the computer as the modern *hypomnemata*, which also facilitates the capturing and measurement of human habits such as searching. Wendy Chun addresses how 'surveillance and the older notions of privacy, remain through our habits' along with how they 'remain central to discipline' (2016:61), a notion I discussed in Chapter 8 with the Homoeconomicus. Alternatively, Chun points out that '[c]apture systems are all about habitual actions. They seek to create new, more optimal habits; they record habitual actions in order to change them' (2016:61). As early as 1994 Philip Agre's essay *Surveillance and Capture* addressed these two phenomena in the technological era. Agre predominantly associates the cultural phenomenon of surveillance with that of the 'classical political sphere of state action' whereas for him, 'the capture model has deep roots in the practical application of computer systems' (Agre 1994:107). However, the two are not mutually exclusive and Agre emphasises their contingency, with the capture model dealing with the 'deliberate reorganization of industrial work activities' to allow computer tracking in real time (ibid:101). These tracking activity systems are 'thoroughly integrated with distributed computational processes' and formed Agre's motivation to map out the phenomenon of capture.

Agre begins with the historical reference of Weisner's 'ubiquitous computing' where 'computational machinery distributed throughout the physical environment' tracked workers and thereby already enounced the creation of 'smart buildings' (ibid:102). Although the computer maintains a centralized database it is widely distributed, with histories and trajectories continuously updated; 'each tracking system is capable of closing the causal loop between the entity and the computer' (ibid:105). Yet tracking processes are not limited in time and space, as with the 'calling up information in the file', because those agents doing the 'calling up' could be tracked as well (ibid).

> As human activities become intertwined with the mechanisms of computerized tracking, the notion of human interactions with a "computer"—understood as a discrete, physically localized entity—begins to lose its force (ibid).

Agre then goes on to describe how computers are used as models of organisations—what he defines as 'grammars of action' which are 'built upon linguistic metaphors' and become crucial to the capture, storage and transmission of information. Following Agre, a 'grammar of action has been imposed on an activity' in regard to how (micro) exchanges are now facilitated with transaction costs, where the

> discrete units and individual episodes of the activity are more readily identified, verified, counted, measured, compared, represented, rearranged, contracted for, and evaluated in terms of economic efficiency (ibid:119).

Recalling the description of Franklin's 'prescriptive technologies' in Chapter 4, with relation to Chapter 5 and 'device cultures' such as Google Search, users are prescribed into the advertisement interface through these 'grammars of action' (Weltevrede 2016:14). Additionally, these 'grammars of action' facilitate socio-technical arrangements and Agre

> hypothesizes that this relation, which "presupposes that the entire world of productive activities can be conceptualized, a priori, in terms of extremely numerous episodes of exchanges among economic actors," constitutes the political economy of capture (Agre 1994:121 cited by Chun 2018:72).

In this way Varian's 'computer-mediated economic transactions', or Zuboff's 'informal actions', are actually Agre's tracking activity systems, or, in the words of Chun, 'most succinctly: capture systems transform all transactions into market-based ones so that computerization = liberalization' (ibid).

The role then of 'information technology' in regard to markets cannot be underestimated, yet it is 'not synonymous with the capture model' because it can have other consequences than just the 'reduction of transaction costs' (Agre 1994:121). Along these lines, Agre's concern lies in the increasing productive efficiency of markets with social benefits due to capture but 'whose social costs ought to be a matter of concern' (ibid:122). Merging with Zuboff, it is then perhaps not the *Big* (Br)*Other* activities of the state—surveillance—that Agre forewarned but the corporative and predatory nature of Google as it captures 'grammars of actions', such as users' search queries. Moreover, Agre's statement that 'capture is never purely technical but always sociotechnical in nature' (1994:112) is aligned with Zuboff's 'logic of accumulation' that comprises 'surveillance capitalism', which decides what is left out, what is accumulated and how it is organised, producing 'its own social relations and with that its conceptions and uses of authority and power' (2015:77).

Figure 96: Google Data Centre, near Mons, Belgium

3.0 Databases of Intentions

'Data banks are the Encyclopedia of tomorrow. They transcend the capacity of each of their users. They are "nature" for postmodern man' (Lyotard 1979:51).

Building upon the address offices, in Chapter 2 I described human endeavours that attempted to organise the world's information by creating publicly available offline indexes, musings on world encyclopaedias and visions of machines for search. The second half of 20th century ushered in new technologies and media and, in the 1980s, the hyperlinking of the World Wide Web. Eventually academic and commercial players built online databases and search engines such as Google.com became one of 'the major nodes' or 'Citadels of the Web', with link distribution reaching about 90% (Warnke 2013:86-88). In this way 'open spaces' became hyperlinked, enclosed Google territories that comprised its 'datascape' (Pasquinelli 2015), with user data stored and processed in 'highly connected, automatically operated, air-conditioned and high-security data centres' (Warnke 2013:86-88).[163] Although users could take part by sharing their information without much 'technical expertise, these 'citadel rulers' dictate what occurs in database operations (ibid:87). (Figure 96)

Following Lev Manovich's seminal article from 1999, if the database is itself a medium, in what ways does the user navigate computational data? As information became distributed throughout millions of servers worldwide it needed to be collated and prepared as data in order to be processed by algorithmic intervention (Gillespie 2014:169). In the electronic era, data consists of bits (a combination of the term **b**inary and dig**it**), annotations, characters, texts and numbers that are 'machine readable'.[164]

> Since the development of relational and object-oriented database architectures, information can be organized in more flexible ways, where bits of data can have multiple associations with other bits of data, categories can change over time, and data can be explored without having to navigate or even understand the hierarchical structure by which it is archived (ibid:171).

Continuously updating and restructuring data aggregated into computational forms, 'relational databases' have separate tables or 'relational variables' that facilitate the ingestion or deletion of new data without affecting the structure (Gitelman 2013:9). As 'object-oriented databases' have appeared data can be arranged and connected in manifold ways when it is retrieved.

> Data are effectively made independent of their organization, and users who perform logical operations on the data are thus "protected" from having to know how the data have been organized (ibid).

As users continuously shared more data about themselves, Google delivered 'relevant' search results based on its collected data. Social production organised this user participation, the general intellect (Chapter 3), with network effects resulting in a 'positive externality' of platform capitalism (Chapter 5). Moreover, this 'free labour' of hyperlinking and user's data—billions of search queries—facilitated the construction of Google's proprietary database. Analogous to the registers of the address offices (Chapter 1), Battelle's 'database of intentions' (2006) is the instantiation of Zuboff's 'logic of accumulation' and this 'gathering of user information is the backbone of digital media economics' (Jarrett 2014:18-19).

---

[163] I attempted to enter one of Google's citadels after visiting *The Mundaneum* in Mons, Belgium, but I was unable to enter or to receive a tour.

[164] As a unit that measures information, '[t]he bit is the fundamental particle… a binary digit, a flip-flop, a yes-or-no' (Gleick 2011:10).

Zuboff's section titled *Personalization and Communication* emphasises the non-reciprocal relationship between Google and its 'users' searching, what she describes as a '21st century Faustian pact' (2015:83). Baited with 'hooks that lure', users were caught up in the 'logic of accumulation' and became habitually accustomed to search services, evermore reliant on the results (ibid). During the early 2000s, Trusted Users (Chapter 8) entered keywords and interacted with the search interface that delivered results and, by clicking on the top SERPs, supplied feedback to Google (Chapter 3). With Google's capturing of search queries, a.k.a. 'grammar of actions', these Trusted Users supplied what Zuboff calls 'Behavioural Data' that then was recirculated with Interactive Learning and Product-Service Improvements. These refined search results were made possible because of Google's 'extensive retention of search data' (Anderson, 2010; O' Brien and Crampton, 2007 cited by ibid:80) and enhanced user search experience.

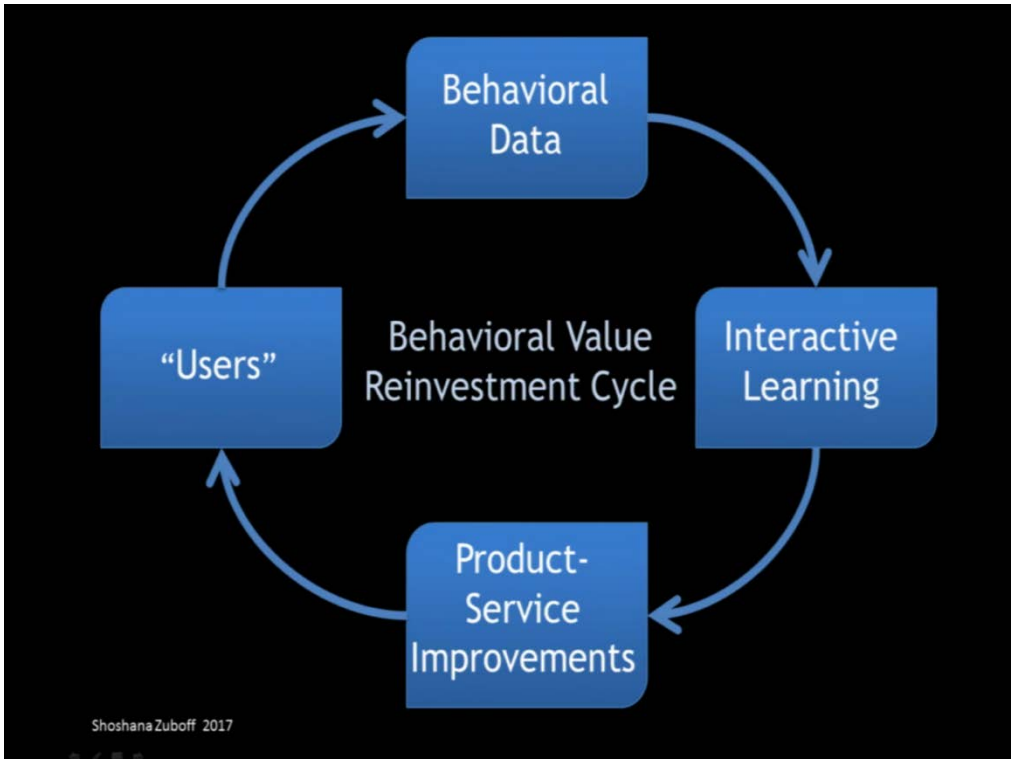


Figure 97: Shoshana Zuboff's diagram 'Behavioral Value Reinvestment Cycle' (2017)

> It was a complete 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). (Figure 97)

In 2010, Battelle 'updated' his statement on the 'database of intentions' to move beyond 'web search' by redefining what is considered a 'field', or specific information that makes sets of fields (called records) and how these records constitute the database (Battelle 2010).
As explained in Chapter 5 with regard to the SEO industry, Google's ecosystem became much larger and contained more signals than just the query, including what Battelle calls 'The Social Graph', 'The Status Update' and 'The Check-in'. Morever, these echo conclusions from the article *Personal Web Searching in the age of semantic capitalism: Diagnosing the mechanisms of personalisation* (Feuz et al. 2011). 'The Social Graph' includes data about not only 'who we

are' but 'who we know' (Battelle 2010) and this directly refers back to the linkages of PageRank that mapped how the webpage is situated in the graph of the web through connections and applied weights to these 'ties' (Granovetter 1973), as described in Chapter 3. 'The Status Update' reflects what is on the user's mind in real time and exudes immediacy', harking back to the offices and their concern about information being current (Chapter 1) as well as echoing Chun's research in her book, *Updating to Remain the Same: Habitual New Media* (2016). In signal terms 'The Check-in' means 'locative data', where the virtual world is connected to 'terra firma', using GPS coordinates and other positioning metrics (cell towers) that are now automatic (if location is turned on), as shown by my results in Chapter 5.

With 'personalisation', results are based on users' signals (IP address, keywords, search history, locative data and browsing habits), as described in Chapter 5 and above.[165] Alaimo and Kallinikos note that this personalisation is also 'an organizational practice' that regulates interaction between various actors and 'operates by re-engineering the space of consumption,' which in turn is 'shaped by iterative feedback loops' (ibid). Google's improvement of search services continued, with personalisation and targeted ads affecting how the Impressionable Subject becomes individuated and produced through its hyperlinked journeys (Introna 2016) (Chapter 8). However, as Zuboff cogently points out, the Behavioural Value Reinvestment Cycle was not enough because it used up all of the value users created (Zuboff 2017). Over time, with the 'logic of accumulation' of user 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).

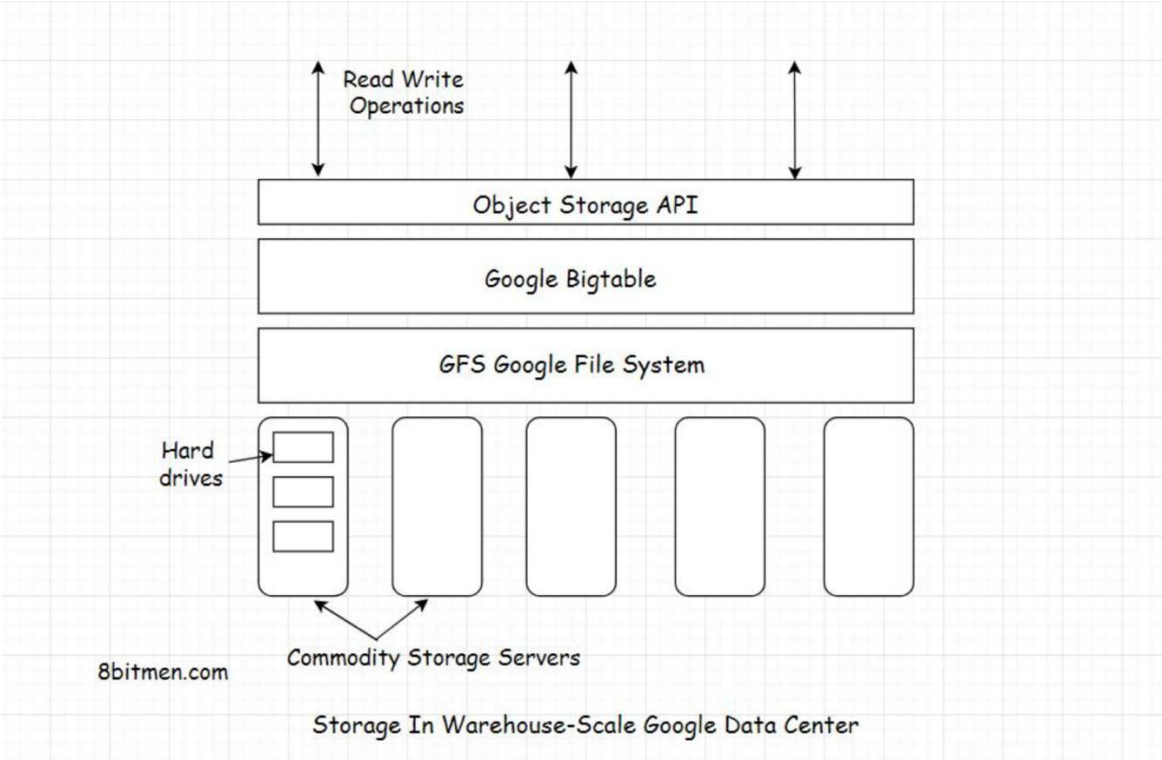


Figure 98: 'Storage in Warehouse-Scale Google Data Center by 8bitmen.com

[165] As described in Chapter 2, with Stewart Brand's Whole Earth Catalogue the presentation of material geared to the needs of the individual in the counter culture already foreshadowed the development of personalised technologies such as the computer (Turner 2006).

4.0 Y[OUR] history

Google is the world's most visited website (Zuboff 2015, 2018), indexing trillions of web pages daily. As explained in Chapter 5, the Caffeine update was a new indexing architecture that was able to continuously update instead of updating in batches, thereby providing fresher results. It also marked the transition from MapReduce to BigTable. Designed to scale into the petabyte range, BigTable uses the Google File System, which is a distributed file system that manages data across thousands of servers. This data processing infrastructure is of utmost importance as it 'deals' with exabytes of data everyday with its polyglot database architecture (relational and non-relational), primarily BigTable (see Appendix I). (Figure 98) Harking back to the address offices which held user data in public and secret registers, data is now retained and processed in large centres all over the world containing bits of the constantly evolving Data Dust (Chapter 8). Despite an increase in the accumulation of data points, the indications of the computational operations do not necessarily produce accurate 'data subjects'.

Rather, Goriunova points out that the construction of Digital Subjects differs from the conception of the classical modern subject', as they 'have the structures of computational actions, models, and socio-political cultures' and are connected indexically, which is a 'constant linking and referencing process' (2019:132).[166]

> It is now the discourse network of the Turing machine and its apparatuses – various models of databases—that contributes to the maintenance (rather than the full Kittlerian replacement) of the subject. The subject is produced not through an exercise in writing but through the relationality of data (Goriunova 2015).

With reference to Chapter 4, the Kittlerian networks of data inscription are now information models that are tabular and 'computationally aggregated into an order and composition (Goriunova 2019:133).

> Yes or no in a row; the table itself; the model—these are not spatial and temporal distributions of digital subjects, but forms of recording, aggregating, establishing relationality, and prediction, where each subject has its own spatio-temporal framework (ibid:134).

Although these conceptual database schemas dictate a model of the world, search engines reassemble information through algorithms (Burkhardt 2012:265).[167] With collaborative tagging systems, labeled resources and assigned tags of users not only suggest new, similar resources but also other users who have similar interests. By attributing preferences and interpreting actions such as search queries or clicking on results, individual users are 'assessed by comparison to a network of standardized behavioural expressions of others' (Alaimo and Kallinikos 2019:401). The organisation of these databases functions as a gated community, or, to revisit Chapter 5, a neighbourhood, and assigns these frameworked and databased Digital Subjects into categories or groups based on attributes of other users who are similar (Beer 2009;

[166] For a detailed explanation, please see *Indexing It All: The Subject in the Age of Documentation, Information, and Data* (Day 2014).

[167] The conceptual database schema is a symbolic formation or pre-logical structuring of reality, yet can this algorithmic production of relevance as an alternative way of symbolic formation be understood with and by computers? (Burkhardt 2012:265).

Feuz et al. 2011; Gillespie 2014; Chun 2016). However, similarities are not stored in the database, but can be derived from the database information (Burkhardt 2012:270).

Echoing my findings in Chapter 5 and my proposition of a 'lack' of authenticity regarding 'personalised' search results, with 'collaborative filtering' users are matched to similar users by Google's system of database organisation: classification and taxonomies of computable data. These 'seemingly accurate data-based techniques make personalized services look like an empirically grounded mediation through which users can discover their own, allegedly true, needs and predispositions' (Alaimo and Kallinikos 2019:409-410). However, during a 'complex journey of technologizing experience', the 'reductive' and hidden techniques in regard to personalisation 'undermine any genuine concern for persons as unique cultural individuals' (ibid:408). Where once standardised customisation enabled marketing and demographics of populations, nowadays the 'computable data footprint underlain by carefully crafted user models' is able to make 'users comparable or commensurable with one another' (ibid:401). This relationality of data is 'aggregated from singular points, rules, and models'…. and 'does not even need to exist or be proven right' (Goriunova 2019:134).

With the assigning of identities and subjectivities and constant aggregation that sorts, filters and compares, these database operations in turn rearrange 'singularities in the operation of multiplicity' (Goriunova 2015 cited by ibid). As exemplified by YOUs (Chapter 8), which are addressed by the search engine, and are, crucially, both singular and plural (Chun 2016), it is not the search habits of an individual, rather it is the actions of the collective because 'singular actions' have 'become indications of collective patterns rather than exceptions' (ibid:57). In her section *Habitual Connections*, Chun spells out how linkages are made between various sets of data as correlations and exposes the way 'data analytics are about habits' in that 'they "replace" causality with correlations between habits' (2016:57). These correlations decide what to illuminate—matching entities as inductive reasoning—and the more data processed, the more correlations appear (ibid). Moreover, these correlations are made not 'based solely on an individual's actions and history but rather the history and actions of others "like" him or her' (Chun 2016:104).

> Philosophical probability, in other words, expands beyond an individual's experience to draw from experiences of people "like you"; through data analytics, your history becomes Y[OUR] history (Chun 2016:56).[168]

5.0 Extraction of Surveillance Assets

Shifting back to the Novel Territories described in the Introduction, the main protagonist of Eggers novel, Mae, is underground in a cave-like structure with Kalden, one of the so-called 'Three-Wise Men' who run the company *The Circle*.

> They entered a large room, about the size of the basketball court, dimly lit but for a dozen spotlights trained on an enormous red metallic box, the size of a bus. Each side was smooth, polished, the whole thing surrounded by network of gleaming silver pipes forming an elaborate grid around it.
>
> "It looks like some kind of Donald Judd sculpture," Mae said.

> Kalden turned to her, His face alight. "I'm so glad you said that. He was a big inspiration to me. I love that thing he once said. 'Things that exist exist, and everything is on their side'.
>
> […]
>
> "He did this for the company?" she said, nodding at the massive red box.
>
> Kalden laughed, then looked at her, his interest in her not gone, but certainly in retreat. "No, no. He's been dead for decades. This was just inspired by his aesthetic. This is actually a machine. Or inside it is. It's a storage unit".
>
> He looked at Mae, expecting her to complete the thought.
>
> She couldn't.
>
> "This is Stewart," he finally said.
>
> Mae knew nothing about data storage, but had been under the general idea that storing such information could be done in a far smaller space.
>
> "All this for one person?" she asked.
>
> "Well, it's the storage of the raw data, and then the capacity to run all kinds of scenarios through it. Every bit of video is being mapped a hundred different ways. Everything Stewart sees is correlated with the rest of the video we have, and it helps map the world and everything in it. And of course, what you get through Stewart's cameras is exponentially more detailed and layered that any consumer device."
>
> "And why have it here, as opposed to stored in the cloud or in the desert somewhere?"
>
> "Well, some people like to scatter their ashes and some like to have a plot close to home, right?" (Eggers 2013:219-220).

In the novel, Stewart a.k.a. 'The Transparent Man' enacts 'sousveillance' by wearing a camera around his neck that is live-streamed to the server.[169] Constantly producing data that is correlated in millions if not billions of ways, he is simultaneously watched as he watches. With Stewart's data and that of others 'like him', *The Circle* increasingly expands its operations as more and more employees go 'transparent', enabling the 24/7 collection of their data and storage on its massive servers. However, this type of fictional 'becoming transparent' is also played out in real life, 'where transparency, as Antoinette Rouvroy has argued, refers not to user knowledge of the system but rather to user ignorance' (Chun 2016:58). It is an asymmetrical power relationship where the 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 2015:83).

Similar to the extraction of natural resources (humans, land and elements) in colonial times, this extraction process involves rounding up streams of search data with the 'logic of accumulation', what Zuboff deems the 'fourth fictional commodity'. She draws on the work of Karl Polanyi who, over 70 years ago, wrote in *The Great Transformation* how the '3 fictions' of market

[169] Whereas surveillance means watching from above (sur), sousveillance is from below (sous) and it is not the authorities but ordinary people doing the observing.

economies (life, nature and exchange) became transformed into 'commodity fictions', which can be bought and sold (ibid:85).

> The first was that human life can be subordinated to market dynamics and be reborn as "labor." Second, nature can be subordinated and reborn as "real estate." Third, that exchange can be reborn as "money"' (Polanyi cited by ibid).[170]

In the context of surveillance capitalism, this user data is now seen as 'the new oil' (Mayer-Schönberger and Cukier 2013, Taplin 2017:1), a 'resource' that must not only be extracted but refined as it fuels the information economy as well as search engines. 'After all, contemporary capitalism seems to increasingly rely upon data as a specific kind of raw material to be gathered, extracted and commercialized' (Srnicek cited by Beyes 2021:370). However, as Geoffrey Bowker points out, data is also never entirely 'raw' but 'cooked with care' during its collection, storage and transmission (Gitelman 2013:3).[171]

By collating and storing tremendous amounts of user data on Google server farms, databases were merged together, or organised with (big)datamining techniques operating according to a new statistical practice.

> Once they [data] are organised, through tabulation the analysis begins, which is referred to as "datamining", '(e.g. decision trees, cluster analysis, neural nets, text mining, anomaly detection and many others), what is distinctive about these methods is the specific *interested* perspective that guides the search for 'valuable' data and patterns (Zarsky 2003 cited by Weiskopf 2018:6).

With what is now called *Big Data* (Mayer-Schönberger and Cukier 2013), statistical validity is replaced by these 'digital procedures' that collect habitual search data and then 'identify correlations, analyze patterns, and produce additional information' (Esposito 2015:09).[172]

Over the course of the past 20 years, Google has continued to improve its 'personalised' search services and targeted ads by recycling and refining its treasure trove of data. Because of the pressure mounted from investors at Google, 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 (ibid). New methods were found and Google was prized for its ability to extract data, human experiences and interests, which users had intentionally opted to keep private. Applying advanced technical know-how, 'Google was able to access this behaviourial data—irrespective of users' intentions—even when users' intentions were made explicit' (Zuboff 2017). Google Search started to relate behaviour as a by-product of query activity and Zuboff's Behaviour Value Reinvestment Cycle has now been subordinated to a new kind of market process.

> Behavioural data are now expanded and commodified as "behavioural surplus". Users are the human natural resources from which the raw material for a new kind of manufacturing process is extracted and turned into "surveillance assets"' (Zuboff 2017). (Figure 98)

---

[170] As Zuboff notes, these fictions 'disregarded the fact that leaving the fate of soil and people to the market would be tantamount to annihilating them' (2015:85).

[171] 'Data are familiarly 'collected', 'entered,' 'compiled,' 'stored,' 'processed', 'mined' (Gitelman 2013:3).

[172] 'As stated by Pasquinelli, the two epistemic poles of pattern and anomaly are the two sides of the same coin of algorithmic governance. An unexpected anomaly can be detected only against the ground of a pattern regularity' (Joler 2016).

With Google's extraction activities creating this zero-cost game changer— a new class of 'surveillance assets' that provided a genuine market exchange, difficulties were implicit in this business model from the outset, along with ethical issues (Zuboff 2015).

> Critics of surveillance capitalism might characterize such assets as "stolen goods" or "contraband" as they were taken, not given, and do not produce… appropriate reciprocities (Zuboff 2015:81).

In other words, permission has not been extended, and whether users have 'given' their data is also debatable. However, data is never just singular nor is it 'given'.[173] 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"' (ibid:79). According to Zuboff, the 'prominent, perhaps decisive, characteristic of the emerging logic of accumulation under examination here' is that extraction is a 'one-way process', because there is no relationship —instead there is a monetisation of what is extracted (ibid).

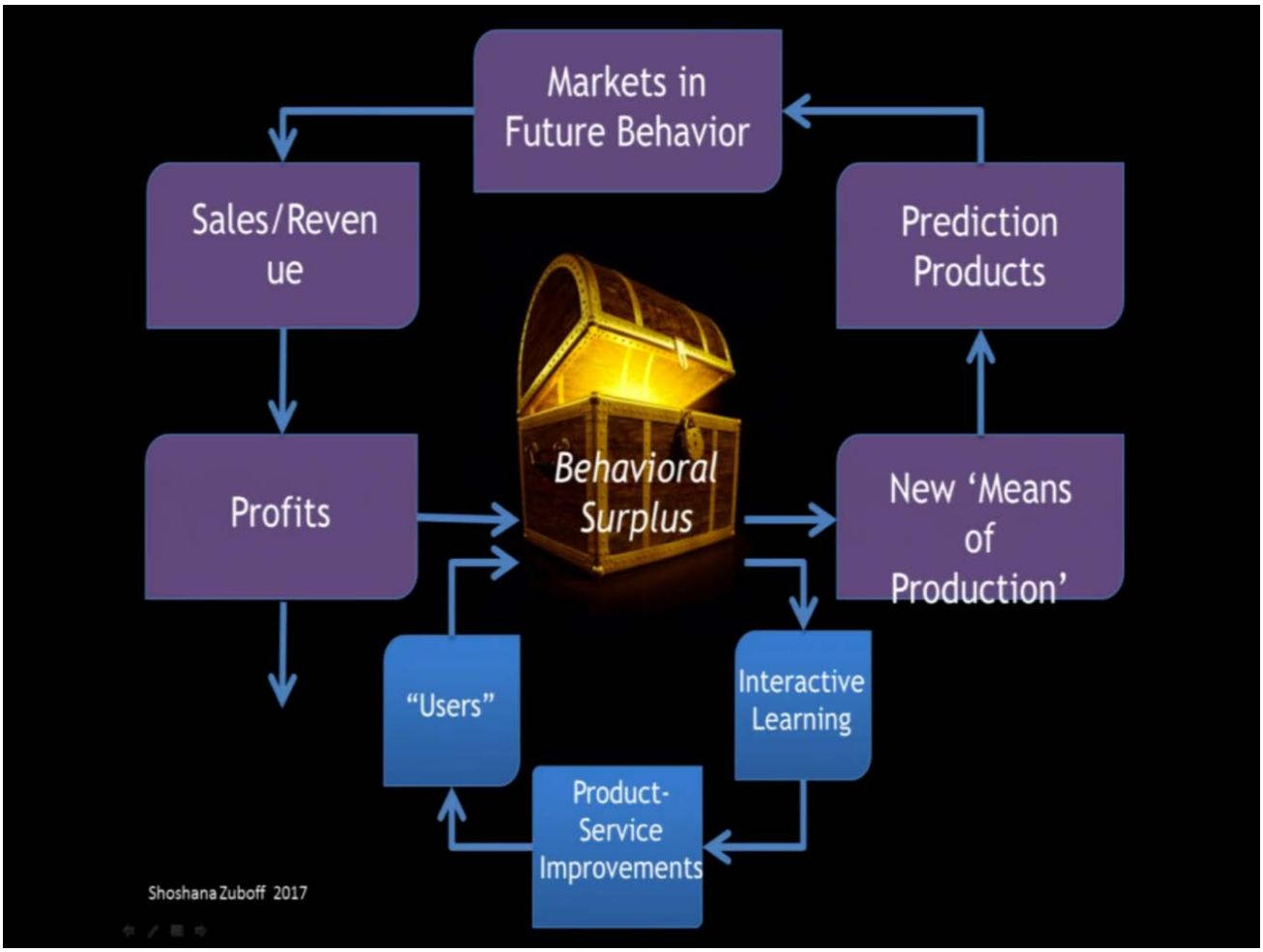


Figure 98: Shoshana Zuboff's diagram 'Behavioral Surplus' (2017)

[173] The etymology of word 'data' is actually the plural of the Latin word datum, meaning 'given'. Daniel Rosenberg provides a fascinating account of the shift from the Latin 'datum' to 'data' (2013).

6.0 Search engines as profiling machines

'Data gives you answers to your questions — just when you need them.
It helps you find the right words to say, in any language.
And gets you from A to B…to C, right on time.
It helps you discover that video that makes you laugh out loud — or your new favorite song.
And helps find everyone you care about in every photo you take.
It's personal. That's why we protect your data.
We do not sell your personal information to anyone.
Your security comes first in everything we do.
Every day, data makes our services work better for you.
That's why it's important that we keep it private and safe – and put you in control'.[174]

As shown in Chapter 5, Google Search is a platform that intervenes (Gillespie 2015) with search results by capitalising on users willingly sharing their data (Srnicek 2016). This makes the logic of accumulation and the extraction process of 'behavioural surplus' possible because users have entered into a 'data contract'. In recent years, public pressure by privacy advocates and journalistic exposure of hacks and breaches have forced corporate 'transparency', with major players such as Google 'promising' that they do not sell user data. This 'visibility management' (Flyverbom 2016) dates from 2015 and more recent policy updates have thrown 'searchlights' on how Google 'uses' YOUsers data. The question then needs to shift from: 'How do users benefit from Google?' to 'How does Google benefit from its users?' (Van Dijck 2010:583).

In his book *Profiling Machines: Mapping the Personal Information Economy,* Greg Elmer addressed consumer marketing with pre-electronic tools and contemporary feedback technology such as cookies in the digital economy, along with their implications (2003). Building upon Elmer, José van Dijck contextualises the contemporary condition of 'search engines as profiling machines', where every search leaves 'traces of its sender' such as an IP address along with 'query key words, activity log, date and time, search history, etc.'(2010:583). With the tracking of specific users' web behaviour, search engines also create a 'personalised' database (second index) on each user (Stalder and Mayer 2009), where categories are delineated and constructed as database schemas and tag clouds, which produce 'data profiles' (Bernhard 2017).[175]

In *Corporate Surveillance in Everyday Life*, technology and data activist Wolfie Christl elucidates how one face of this 'pervasive digital tracking', or profiling, is advertising. This comprehensive 'socially manipulative information technology' is facilitated by Intellectual Property Rights—thousands of Google patents (Cirio 2018). (Figure 99) People are evaluated with scoring systems through various lenses and devices, from browsers on computers capturing web searches to smartphones that 'now monitor and evaluate how people surf the web and use their mobile devices' (ibid:8). This 'behaviour surplus' from search data is not restricted to advertisers as in previous decades but circulated between many interested third parties who buy, sell and trade the behavioural data of users and diverse user groups. Even if Google does not technically 'sell' its treasure trove of user data, it is important to note that exchanges do, however, occur which in turn affect the profiling of users:

[174] Google published this text in 2015 yet it is no longer available at this URL: https://privacy.google.com/
[175] Nowadays corporations such as Google can promise advertisers that they can tailor 'the look and feel of their ads to unparalleled variability to individuals' resulting in the fundamental business model of digital cultures (Bernhard 2017:33).

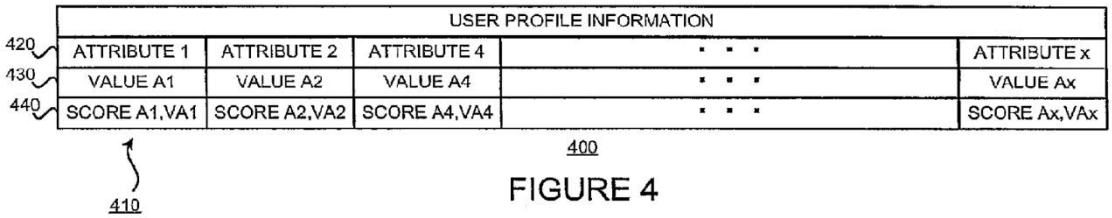


FIGURE 4

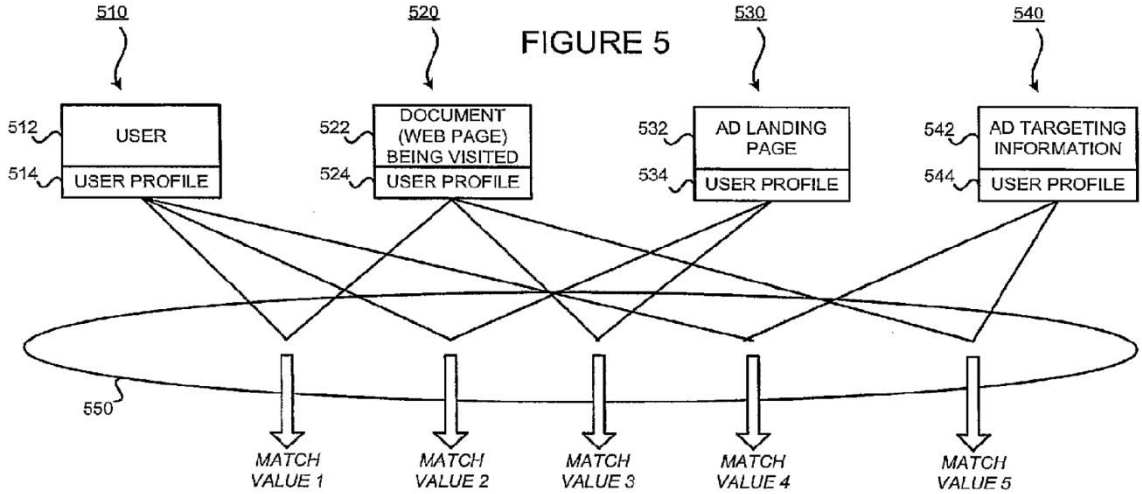


FIGURE 5

Figure 99: Patent for 'Generating user information for use in target advertising'.[176]

> [I]n contrast to popular belief, they do not directly, for the most part, *sell* and *share* their detailed digital consumer profiles to third parties, at least not in the form of unified dossiers. Instead, the 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 2017:11).

In this way it becomes clear that Google Search 'is a personal information economy where the standard exchange is service for profile' (Elmer 2004; Lyon 2007 cited by Rogers 2007) and this connectivity and exchange between users, advertisers and now data brokers is part of the 'service/dataprofile/ advertising complex', as elucidated by Lovink and Tkacz (2015:15).

This 'behavior surplus' (Zuboff 2015) or 'data assets' (Christl 2017) interests data brokers who not only partner with Google but merge offline identity data from previous decades with online purchase histories, termed 'onboarding'. Where once behavioural targeting schemes used analogous technology by collating data to define 'audience segments', nowadays users are matched with similar profiles. Because each individual has been profiled in the database,

[176] Patent: https://patents.google.com/patent/AU2009222546B2/en. Filing Date: 2009-10-02.
Paolo Cirio's artwork *Sociality* 'documented over twenty-thousand patents of socially manipulative information technology'. Cirio collected the patent images and data by hacking the Google Patents search engine. Then he rated the patents and created thousands of compositions with images of flowcharts and titles of inventions, which were published on the project's website. Subsequently, he invited the public to share, flag and ban the technologies designed to monitor and manipulate social behaviors. https://www.sociality.today

Google is able to track users and correlate previous collated data, along with new services and interactions. Simultaneously, real-time updates are constant and incorporated with Google willing to 'share this data with advertisers and many other companies and allow them to utilize, combine, and link it with digital profiles from other sources' (Christl 2017:16). With reference to 'collaborative filtering' (Chapter 5) and 'Y[OUR] History' above, the traded digital profiles by 'data brokerage' marketers are constructed as 'so-called "segments" or "audiences", "grouping" individuals with certain characteristics and behaviours' (Christl 2018). (Figure 100).

Figure 100: How Google constructs customer match audiences and similar audiences (Christl 2018)

> In the case of profiles, correlations in the data linked to an individual are used in a model that refers to a group of people, from which something can be inferred about the individuals whose data points might be missing. Such individuals are neither the people whose data is crunched, nor exactly specific people at all. They might not exist at all, yet might still be *likely* candidates to fit a profile (Van Otterlo 2013:44 cited by (Goriunova 2019:134).

As demonstrated by the Google diagram above illustrating the sorting and matching of users to 'customer match audiences', YOUs are grouped together into 'wes' (Chun 2019:76). Besides deleting inexact matches, the 'customer match audience' becomes part of 'Google Properties' and additionally, is correlated to 'similar audiences' within the YouTube and Gmail databases.[177] Arranged and organised by 'specific computational models' (or profiles), users are actually not humans or 'data subjects' but 'data aggregates', which the technology generates 'i.e., users are lists of such engineered operations such as rating, listening, liking, etc.' (Alaimo and Kallinikos 2019:409). However, these profiles or computational models or 'audiences' are not 'static lists'.

[177] This power of 'metadata' is the 'backbone' of search machines and 'constitute the heart of automated search and further advance Google's position in the race for knowledge production' (Van Dijck 2010).

> They are dynamic real-time feeds about groups of people with certain behaviours, continuously updated with data across contexts, and used to make automated decisions on these people according to complex sets of predefined rules (Christl 2018).

The extracted 'behavior surplus' is now taken up into an ever-changing process that is not stable but in flux and what becomes of importance 'is the algorithmic interpretation at the moments data can be used, sold, or otherwise acted upon' (Goriunova 2019:132).

7.0 Behaviour Surplus Creep

With the monitoring of user's browsing histories, nowadays search data is seamlessly tied into buying patterns across other technology platforms that enrich users' profiles, feeding diverse industries with data. 'Google gains information about our individual and collective search behaviour' and is 'able to collect and connect endless streams of behavioural patterns and reconnect them in ways we can only begin to imagine' (Van Dijck 2010:584). Through the process of disassembling and reassembling data, such as search queries, users are (re)identified through behavioural data or what Zuboff terms 'surveillance assets'. This is the second face of Christl's profiling as 'pervasive digital tracking'—the shift from the information society towards the surveillance society (2017).

Much like the fact that Google search is not a new phenomenon, as demonstrated by Anton Tantner's scholarship on the address offices in Chapter 1, these 'Black Utopias' also carried out surveillance on their employees as well as the people who used their services. The collection of addresses was a means to identify people and make them findable and this data was not only used for policing but profiling. The public took notice and was concerned about what happened to their data because their histories were contained within the pages of the register, which enabled the construction of their reputations based on past behaviour. Later on, technologies used in policing work and behavioural psychology produced profiles (Bernhard 2017).

> Long before the emergence of Big Data, criminal investigators composed profiles of "unknown suspects", psychologists created profiles of people with specific personality disorders, marketing managers created profiles for classifying potential customers according to their potential worth (Weiskopf 2018:11).

In the twentieth century, Gary Marx described how with predominately offline profiling individuals were compared to others, based on extensive State categorisation and 'people-processing' (2001). This profiling was derived from the gathering of manifold data sets, such as the social security number, and then merging gathered data from contexts unrelated to its original purpose, what Marx deemed 'surveillance creep' (2001). Nowadays 'function creep' is a term that describes this invasion of privacy, when technology and systems go beyond their original intended purpose and, in this case, the recyclability of data. 'Surveillance creep' (Marx 1999), 'function creep' or what Christl calls 'mission creep' occurs when

> information about people's behaviors, social relationships, and most private moments is increasingly used in contexts or for purposes completely different from those in which it was recorded –for instance, to make automated decisions in crucial areas such as finance, insurance, healthcare, employment, and law enforcement (2017:10).

Although 'online advertisers and intelligence contractors' collect information on users along with numerous players in the data ecosystem, platforms such as Google dominate (ibid:11). In the Introduction, the 'googlization of everything' *(and why we should worry)* exposed how with Google's 'creep' into other major industries, 'advertising, software applications, geographic services, e-mail, publishing, and Web commerce itself', all services were connected and it became a hegemonic force in the information society (Vaidhyanathan 2012:20). Ready to be called up from servers, proprietary behavioural data reflects Google's 'provenance of aggregated profiling techniques', every time a user searches with one of its engines (Search, Scholar, Images, Maps, News, YouTube, Orkut, etc.) (Van Dijck 2010:584). These 'electronically mediated contexts', defined by technological platforms such as Google also 'shape the ways information about us is tracked, gathered, analyzed, and disseminated' (Nissenbaum 2015:6). Echoing Zuboff's *informating* as well as *automating*, these

> algorithms are able to use all these data for a variety of secondary uses largely independent of the intent or the original context…processing them to find correlations and patterns with calculations that the human mind could not realize nor understand, but which become informative (Esposito 2013:16).

In an era of digital mass storage, 'decoupled' data is now 'torn loose' from the context from which it was originally gathered, thereby creating 'runaway profiles' (Pasquale 2015:30). By fusing data from disparate sources, data brokers 'provide services that allow other companies to recognize, link, and match people across different corporate databases' (Christl 2017:8). In this way, the 'secondary use of data also makes it possible to gain information relevant for the profiling and surveillance of citizens' (ibid:16, Mayer-Schönberger and Cukier 2013:103).

> Critical investigations by journalists, and ironically also by Experian's AdTruth, suggest that Google does not only use IP addresses, browser fingerprints, the way users type, move their mouse, or use their touchscreen "before, during, and after" a reCaptcha interaction, but also several cookies set by Google's services (ibid:38).[178]

Christl's image shows that by combining offline data from previous decades, browsing histories, IP addresses, cookie IDs and device fingerprinting, a data profile is complete. (Figure 101). As shown above with 'mission creep', once the profiles have been removed from their original source or context, the patterns and permutations of personal preferences are endless. This makes its extremely difficult, if not impossible to understand 'on what kinds of decisions, assumptions and prerogatives the constructed object has been formed' (Weiskopf 2018:20).

With this present-day data brokerage (Christl 2017), 'data subjects' are given credit scores by third party marketers, repackaged as digital objects that 'are always updatable and constantly in the making' (Alaimo and Kallinikos 2019:410). Using software, companies are able to establish relationships between these 'digitals objects', which are Chapter 8's Data Dust (shadows,

[178] When users visit websites, 'cookies' or a text file similar to an ID number is added to the user's browser and in this way they can be tracked across the World Wide Web as they search. 'There are different applications of data gathered by cookies and some of them do influence functionality (localisation and user specific settings), but the widest use of cookies is advertising. More precisely, profiling based on users' habits. Many cookies gather data about the referrer URL address, i.e. the URL from which the user came to the website containing the cookies' (Share Labs 2016). At the time of writing (March 2020), cookies are used by about 44.5% of all websites and 'persistent cookies' are used by 24,8% of all websites, including google.com and youtube. Available here: https://w3techs.com/technologies/details/ce-persistentcookies

doubles, derivatives), thereby producing 'surveillance assets'. Consumers are identified and sorted into groups having 'shared characteristics', what Christl describes as 'segments', along with being assigned 'scores' that reflect the 'likelihood that an individual exhibits certain characteristics or predicted behaviors' (2017:15). In order to monitor individuals, hidden connections are determined which might present a threat to society (ibid:36), yet these correlations affect decision-making processes. The result is that employers, universities, banks, insurance companies, law enforcement, etc. address potential risk factors by making 'behavioural predictions' on these 'digital objects'. Therefore, algorithmic judgments not only hone in on 'data subjects' to sell commercial products through targeted digital advertising but 'to imprison, medically treat, or discriminate against individuals' […] which in contemporary society 'become urgent, political matters' (Goriunova 2019:130).

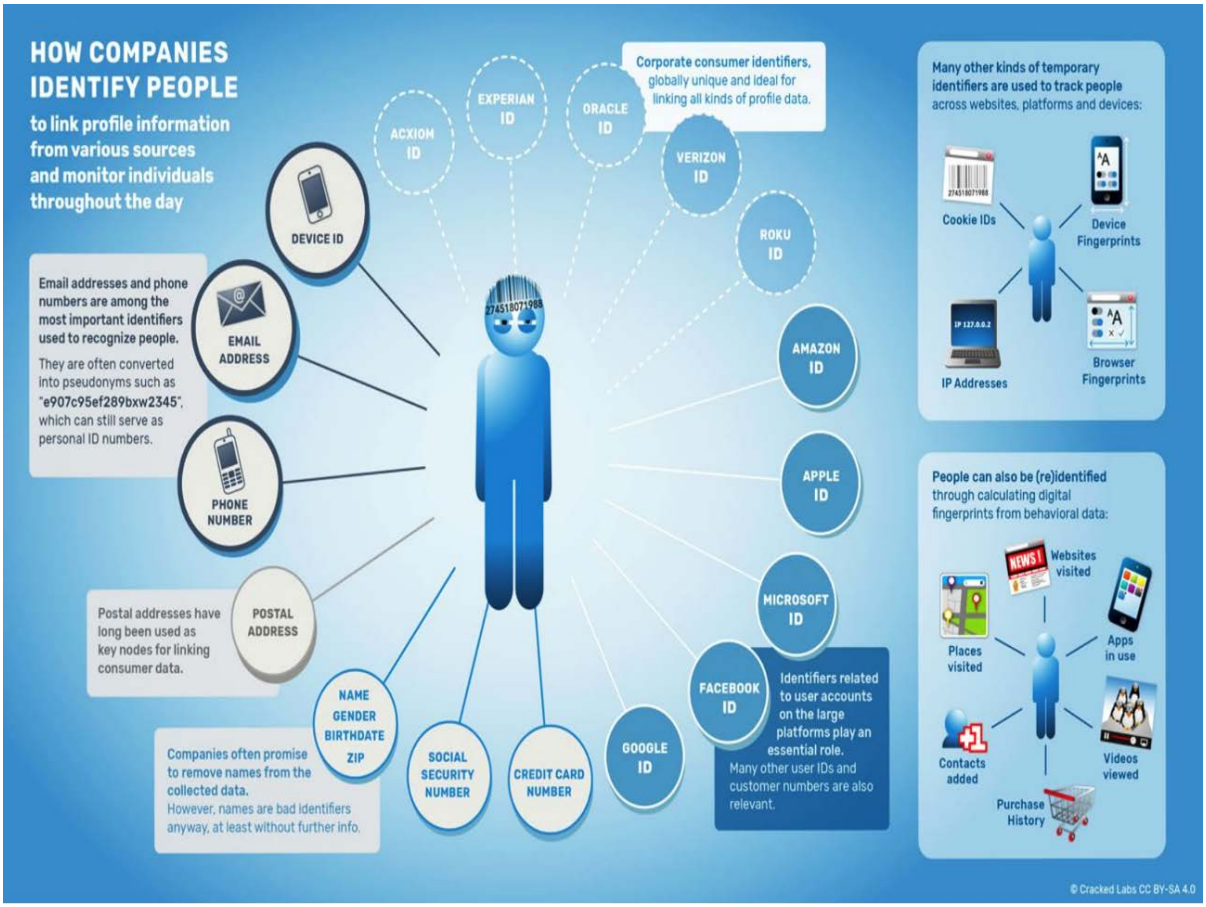


Figure 101: 'How Companies Identify People' (Cracked Labs 2017)

However, these black-boxed algorithmic decisions can also generate what Lisa Blackman terms 'haunted data', which can be 'redacted, removed, accumulate, leave traces and also disappear' (2016:10).[179] The extraction of data and signals results in the aggressive stealing of individual subjectivities (Zuboff 2015:79), which conditions the subject and leads to the monetisation *and* subjectification of those profiled. As with the Interpellated Subject (Chapter 8), through the habit of Google Search, users are constantly addressed or 'hailed' by this algorithmic iteration,

[179] 'Haunted data' practices reflect the 'dark net of big data' in regard to users interactions with search engines and their implications—the construction of profiles that are actually abstractions, spectres or ghostly demarcations, which can be created in one point in time but appear elsewhere.

thereby interacting with search results that also provide them with 'an image of themselves' as 'prior orderings' (Weiskopf 2018:13).

> If, then, today, in the guise of detailed databases (PROFILES) that circulate in the corporate cyberspace and determine what we effectively are for the big Other of the power structure—that is, how our symbolic identity is constructed—and we are in this sense 'interpellated' by institutions even without being aware of it, one should nevertheless insist that this 'objective interpellation' actually affects my subjectivity only by means of the fact that *I myself am well aware of how, outside the grasp of my knowledge, databases* (PROFILES) *circulate which determine my symbolic identity in the eyes of the social 'big Other'*. My very awareness of the fact that 'the truth is out there', that files (PROFILES) on me circulate which, even if they are factually 'inaccurate', none the less performatively determine my socio-symbolic status, is what gives rise to the specific proto-paranoiac mode of subjectivization characteristic of today's subject: it constitutes me as a subject inherently related to and hassled by an elusive piece of database (PROFILES) in which, beyond my reach, 'my fate is writ large' (Zizek 2001:260 additions by ibid:21).

Instead of habitually searching as individuals, with encoded profiling, advertising and surveillance value is then ascribed to these Subjectivities of Search, where 'individual actions coalesce bodies into monstrously connected chimeras' (Chun 2018:76). The 'spectres' of subjectivities are repackaged, bundled, transformed into commodities and sold for the specific needs of the companies who deal in profiles and future behaviours. Moreover, with these algorithmic profiles undermining reality as 'ghostly demarcations' (Weiskopf 2018:12), users not only reflect upon but monitor their behaviour or 'surveil' themselves. Although appearing 'neutral', without critical interpretation these 'carefully crafted fictions' (Kitchin 2017:17) can come to haunt corporal [data] subjects (Weiskopf 2018:12).

8.0 Prediction products

> That's the weird thing about search engines, it's like striking oil in a world that hadn't invented internal combustion. Too much raw material, nobody knew what to do with it. You see my competitors, they were fixated on sucking it up and monetising it via shopping and social media. They thought that search engines were a map of what people were thinking but actually they were a map of how people are thinking. Impulse, response, fluid, imperfect, patterned, chaotic.[180]

As conveyed in the beginning of this chapter, in 1981 Shoshana Zuboff tellingly pointed out the fundamental duality made possible by the new IT tools of capture, with information technology alone having the capacity to 'automate and to informate' (2015:76). With 'informating', what was previously not commensurable became 'textualised' or codified as data, with 'signals' revealing human behaviour (Zuboff 2015). Captured predominantly unseen by devices 24/7, users' queries and clicking online as 'grammars of action' (Agre 1994) are part of this unceasing flow of information. Electronic devices are now connected to infrastructures and their everyday functions automated, as 'ubiquitous computing' (Weisner 1998) makes the communication of information pervasive.

[180] Spoken by the protagonist Nathan, Ava's creator in *Ex Machina* (Written and directed by Alex Garland, 2014).

Although Brin and Page initially did not set out to combine advertisements with search, in the beginning of the 2000s they began to connect user's queries to advertisements, which appeared on the margins of the webpage, as explained in Chapter 5. Zuboff's Behavioral Value Reinvestmant Cycle demonstrated how Google recycled search data to improve its search services (2017). However, Google's Hal Varian realised early on that user data is ubiquitous and cheap and with the atomistic capture of search data, Googlenomics expedited 'predicting the present' or in business jargon, 'nowcasting'.

> Selling ads doesn't generate only profits; it also generates torrents of data about users' tastes and habits, data that Google then sifts and processes in order to predict future consumer behaviour, find ways to improve its products, and sell more ads. This is the heart and soul of 'Googlenomics'. It's a system of constant self-analysis: a data-fuelled feedback loop that defines not only Google's future but also the future of anyone who does business online (Levy 2009).

Search is now the most common activity of today's online world (Lewandowski 2016) and 'ubiquitous googling' has become the means to find information and seek knowledge in the 21st century. For the past 20 years, users worldwide have been contributing to Google's 'database of intentions' (Chapter 3 and above) with their queries along with their most intimate desires.

> In the case of Google, though not an archive of the specific intent organized in the interest of a particular concern, it functions as one of the most ubiquitous and powerful record keepers of digital engagement. It records our searches or inquiries, our curiosities and thoughts (Noble 2018:126).

These keywords and phrases indicate not just *what* people are thinking but *how* they are thinking.

One former Google researcher, Seth Stephens-Davidowitz, studied 'anonymous' Google data during 2013-2017, collecting and analysing what people divulge to a search engine. Although trained as economist, in *Everybody Lies: What the Internet Can Tell Us About Who We Really Are,* Stephens-Davidowitz argues that while 'everybody lies' offline, when they engage with Google search to find information they also express their true feelings, insecurities, anger, hate and fears.

> Google was invented so that people could learn about the world, not so researchers could learn about people, but it turns out the trails we leave as we seek knowledge on the internet are tremendously revealing (Stephens-Davidowitz 2017b).

Stephens-Davidowitz believes that search queries are a type of 'digital truth serum' and that these datasets can not only offer answers but improve lives. Based on his comprehensive analysis, Stephens-Davidowitz asserts that 'I am now convinced that Google searches are the most important dataset ever collected on the human psyche' (2017b).[181]

Although Google's treasure trove of user data has continued to grow, Varian revealed that keyword terms alone cannot measure affect because single-word terminologies of search queries

[181] Stephens-Davidowitz was able to obtain certain data from Google Trends, which is now no longer possible. Personal meeting on October 29, 2019.

are categorised as 'revealed preferences' and do not provide enough 'information' as data. Therefore, by also extracting 'behavioural surplus', Google's 'logic of accumulation' facilitated an expanding business of surveillance capitalism where 'users were no longer ends in themselves, instead they became a means to other [commercial] ends' (Zuboff 2017). As more actors joined the Google market, profits and 'exchange was not with users but with other companies—customers who learned how to make money from low-risk bets on a user's future behavior' (ibid). Nowadays, along with the real-time collation of search queries, Google's behavioural surplus, a.k.a. surveillance, assets are refined as prediction products, which are now 'sold into a new kind of marketplace—a market that trades exclusively in future behaviour' (ibid).

Besides the financial aspects, targeted digital advertisement alters users' behaviour, consciously or unconsciously, as it creates a demand in order to generate revenue—exerting behavioural control over users. Search engines as profiling machines (Van Dijck 2010) conjure up abstractions that are created for the purpose of diagnosis or prediction yet these configurations 'are not a reflection of a given identity, but a projection of possible future behavior' (Weiskopf 2018:12). With the 'corporate surveillance of everyday lives', computer systems aggregate data from different entities and synthesize information in order to identify patterns of behavior and predictive assessment (Christl 2017). 'Based on data-driven predictive analytics, personalization, measurement, and testing, these marketers aim to influence behavior at scale' (ibid:5). Made possible with the proliferation of 'Big Data', now 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).

> The light of Big Data creates big shadows through its very mechanism of capture, which shapes the reality it allegedly mirrors by depending on past data to 'pass on' data (ibid:59).

However, *Big Other* (Google) and 'Big Data' break with the past—whereas 'capitalism once profited from products, then profited from services, it now profits from surveillance' (Zuboff 2015, 2017, 2018). Google realised that as user behaviour is influenced and modified —thereby organised—it increasingly improves the quality of the prediction, which in turn perpetuates the cycle.

> But it is more than merely Google's particular domination of the user that is perpetuated in the way commodified user behavior acts upon us. As Astrid Mager describes, the "capitalist spirit gets embedded in search algorithms" by way of the impetus users give to the continuation of search and its advertising-based metrics (Jarrett 2014:19-20).

The habits of Chun's 'always searching, never finding' Neoliberal Subjects (Chapter 8) are tracked, captured and altered by technological infrastructures where a distinction no longer exists between them and the external world. Analogous to Henry Ford's production model where he increased wages so that his employees could afford the commodities they produced, '[p]opulations are the sources from which data extraction proceeds and the ultimate targets of the utilities such data produce' (Zuboff 2015:79).[182]

[182] Expressed otherwise, where once General Motors discovered Managerial Capitalism, nowadays Google has discovered Surveillance Capitalism.

With Google's unbridled collation of search data as a 'continuous experimentation' of untamed learning opportunities from 'data subjects', 'reality' is now transformed, reborn and marketed as 'behaviour' (Zuboff 2015:79). As AI driven technologies and the IoT (Internet of Things) distribute computed resources, Web 3.0 'will give rise to intelligent systems based on tracking, interpreting and predicting intuitive behaviour of human actors' (Van Dijck 2010:587). The application of machine learning algorithms, such as RankBrain (Chapter 5), along with 'smart devices' that answer queries, will continue to influence 21st century production, whilst incorporating users' searches and 'modifying behavior to produce new varieties of commodification, monetization, and control' (Zuboff 2015:85). Analogous to the Black Utopias with their policing and surveillance that made it possible for data to be taken up in the register, with 'surveillance capitalism' Google decides what data is accumulated and how it is organised (Zuboff 2015:86). Concomitantly it organises (us)ers in the process, as this chapter has attempted to elucidate. The next chapter addresses some of the implications and consequences of surveillance capitalism, i.e. the 'automated ubiquitous architecture of *Big Other*' (Zuboff 2015:86).

Chapter 10:

Agency: Ad, State, Computational

With recourse to the Introduction and why the politics of search engines (still) matter, this chapter maps some of the implications and consequences of Google's organisation of (us)ers by surveillance capitalism. It expands upon Chapter 9 towards findings that attempt to further Zuboff's analysis, structured by three agencies: Ad, State and Computational.

1.0 Ad Agency

> The future is 'dataism'—the idea that if you have enough data, on a person, especially biometric data, and if you have enough computing power, you can understand that person better than that person understands himself or herself. And then you can control this person, manipulate them and make decisions for them. It's not a person, but it's a corporation or an entity that we created but now it is controlling us. It is shaping our society, our views, our decisions. In 20 years, the advice will be, forget about your heart, what does it know? Ask Google. Google knows you better (Harari 2017).

1.1 In Google We Trust

For a thousand years, until the 4th century when it was outlawed, people (Theopropoi) from all over the ancient world made pilgrimages to Delphi to consult the 'oracle'. As shown in Chapter 1, Renaudot's Parisian bureau d'adresse also performed an oracular function as 'issues to be addressed' (questions à résoudre) and by 1820 the 'inquiry and information comptoir' in Vienna had acquired a reputation for predicting the future (Tantner 2014). Similar to the Delphic oracle, clients of the address offices paid for their consultations and found answers by consulting the register. These ledgers eliminated the middleperson and were eventually replaced by newspapers that in turn have been usurped by search engines, which answer users' queries. When searching for answers or verification nowadays people often say, 'google it'—another way of stating 'ask the oracle' for guidance or responses to burning questions.

As elucidated in Chapter 9, Google is now in the business of delivering the future, with its prediction products offering a form of highly calibrated 'programmatic' advertising whilst marketing personalisation based on its enormous amounts of user data. As explained in Chapter 5, users reciprocally trust Google to deliver quality search results, which they believe reflect their interests, which are contingent on their search histories and locative data (Feuz et al 2011). Google's chief economist Hal Varian imparts that 'people have come to expect personalized search results and ads' and that users are willing to divulge their data because they get something in return (2014:28). He rationalises the exchange as a reciprocal relationship, just like users share information with their doctors, lawyers and accountants, whom they trust (Zuboff 2015:83). However, as demonstrated in Chapter 9, 'surveillance capitalism is the precise opposite of the trust-based relationships to which Varian refers' (ibid). Google's business model in recent years has shifted to include not only 'intangible assets' such as proprietary algorithms but also the behavioural surplus of 'freely given' user data contained within its databases. According to Evgeny Morozov, the secret of Silicon Valley has now been revealed to the public:

the data that users (freely) supply to digital platforms has a greater economic value than the costs of the platforms' services (2017).

Besides the behavioural modification of users through *Big Other's* 'prediction products', Varian's stance 'adds a new dimension to both hegemonic ideals in that now this "God view" can be fully explicated, specified, and known, eliminating all uncertainty' (Zuboff 2015:82). Chapter 9's *Big Other* is an architecture 'existing somewhere between nature and God' with Varian declaring 'that Google wants to do even more. Instead of having to ask Google questions, it should 'know what you want and tell you before you ask the question' (Zuboff 2015:81). With Google's former motto 'don't be evil' still resonating, the public rhetoric of the contemporary condition is to 'trust the benevolent nature of corporate oversight' (Bilic 2017) as an 'object of faith' (Halavais 2009:1-2; O'Neil 2016:29). As Tantner keenly notes, every invention of 'new media' generates a utopia of omniscience and it was Google's co-founder Sergey Brin who suggested that 'the perfect search engine would be the mind of God' (Peters 2015:318). With the *Big Other* now accorded divine attributes, the wisdom of the oracle or human ruler—'a place protected by a metaphysical limit to knowledge'—is replaced by the 'sovereignty of data processing', which is no longer a 'cosmology' of the gods but rather 'technology' (Beyes and Pias 2019:101).

The 'materiality of technical logical output' determines search results, which are 'situated in an authoritative mechanism that is trusted by the public: Google' (Noble 2018:32). However, studies show that users cannot distinguish between organic SERPs and advertisements and that 'users' trust in Google may be misplaced' (Lewandowski 2017:22).[183] With 'paid search', Google collects money for ads to be displayed when certain terms are searched, yet in 2012, of the 83% who state that they use Google, '[m]ost people surveyed could not tell the difference between paid advertising and genuine [organic] results' (Noble 2018:35). Even if users believe they know the difference, there is a 'blurred line' between the carefully calculated presentation of organic results and ads (Lewandowski 2017:22). Instead they view the content they see (including ads) as trustworthy (ibid). The consequences include users still predominantly clicking on the highest ranked links above the 'fold' (Introna 2016; Lewandowski 2017),'nor are they often looking past the first page or so of search engine results, as a general rule' (Noble 2018:85), which is one of my motivations for this research.

1.2 Being number one

> Google's search results page has changed over the years, becoming much more ad-heavy. Ads in recent years claim more space at the top of the results page with various features that present specialized results for specific topics, like hotels or places, often with photos or maps. The results in some of these features are paid advertisements. As Google has placed more ads and verticals at the top of the page, organic search results have shrunk (Grind et al. 2019).

Returning to the Prologue, the fictional president in the television series *House of Cards* describes how his opponent could control the election with his 'powerful gun'—the search engine *Pollyhop* that turns searches into votes. As shown in Chapter 3 with Googlearchy (Hindman 2009), 'there is a sociopolitical value in highlighting the populism of the criteria the

[183] Users were asked to identify organic search results and ads on five different screenshots, yet 'only 1.3 percent of participants' were successful in doing so correctly (Lewandowski et al 2017).

algorithm uses' (Gillespie 2014:182). Certain links have more 'relevance' and therefore majority and authority, with the ranking of information reflecting the 'political, social, and cultural values of the society that search engine companies operate within' (Noble 2018:45,148). With the increased employment of search engines by users in the preceding two decades, society now places a 'premium on being number one, and search-result rankings live in this de facto system of authority' (ibid:32). Rather than delivering 'objective, consistent and transparent' results (ibid:45), information is now prioritised by Google Search as it reciprocally shapes culture. 'Search does not merely resent pages but structures knowledge, and the results retrieved in a commercial search engine create their own particular material reality' (ibid:148).

In *Algorithms of Oppression, How Search Engines Reinforce Racism*, Safiya Noble unearths the 'corporate logic' behind the platform, which delivers 'algorithmically crafted web search that offers up racism and sexism as the first results' through paid online advertising (ibid:5). Moreover, Noble poses the question of what happens when the 'majority rules' and controls minorities by 'the way they're represented in the search engine?' (ibid:16). To return to Chapter 9 and Stephens-Davidowitz's 'digital truth serum' hypothesis (2017), the past twenty years users worldwide have been feeding Google's index not only their thoughts, questions and desires, but also their anger, biases and prejudices. Connecting back to the Prologue, Dylann Roof's federal hate crime was influenced by his search results regarding the killing of the unarmed Trayvon Martin by neighbourhood watch volunteer George Zimmerman in 2012 and this was 'the event that truly awakened him' (Hersher 2017). In his manifesto he states that there were 'pages and pages' of 'brutal black on White murders' and that the media had blown up Martin's case while these were ignored.[184] (Figure 102) Roof typed into Google the following query: 'black on white crime' and 'the first website' he came to was the Council of Conservative Citizens, a white supremacist organization instead of leading to any experts, libraries or articles about the history of race in the United States (Noble 2018:115).

Roof Manifesto Excerpt

> The event that truly awakened me was the Trayvon Martin case. I kept hearing and seeing his name, and eventually I decided to look him up. I read the Wikipedia article and right away I was unable to understand what the big deal was. It was obvious that Zimmerman was in the right. But more importantly this prompted me to type in the words "black on White crime" into Google, and I have never been the same since that day. The first website I came to was the Council of Conservative Citizens. There were pages upon pages of these brutal black on White murders. I was in disbelief. At this moment I realized that something was very wrong. How could the news be blowing up the Trayvon Martin case while hundreds of these black on White murders got ignored?

Figure 102: Dylann Roof's manifesto (2015)

The consequences of Roof's (re)search were far-reaching—the first result is deemed the 'most credible and trustworthy information available', in turn reinforcing that 'Google queries can reaffirm one's existing ideological beliefs or, as was the case with Dylann Roof, help mold an extremist' (Noble 2018b). Through what Noble terms the 'gaze of the search engine', she argues that search results 'oversimplify complex phenomena' (2018:71,116). 'Algorithms are, and will continue to be, contextually relevant and loaded with power' (ibid:171) yet they do not 'read'

[184] Roof's manifesto: http://archive.is/KeAK3#selection-9.571-9.993

the content but 'point'.[185] Rather than receiving answers that reflect 'counterpositions' or a 'disclaimer or framework contextualizing what we get', algorithms that 'rank and prioritize for profits compromise our ability to engage with complicated ideas' (ibid:118). Because Google is an advertising company that serves the interests of its advertisers above that of information, Roof 'trusted' the first result, although it remains unclear how 'autocomplete' —what each individual user is shown by Google Search as they enter keywords—affected his search results and to what degree Roof was personalised.

Therefore the 'relevance' and 'information quality' of the retrieval of information should be brought into question because search platforms such as Google are 'not unique to the individual searcher' (ibid:150). 'For now, this level of personal-identity personalization has less impact on the variation in results than is generally believed by the public' (ibid:55). Noble's position concurs with my results in Chapter 5 and 9 concerning the collaborative filtering of users into other similar users in Google's database. As Wendy Chun further points out, '[n]etworks preempt and predict by reading all singular actions as indications of larger collective habitual patterns, based not on our individual actions but rather the actions of others' (Chun 2018:75). Although this 'homophily' reflects how 'individuals "stick" together, and "wes" emerge', it also 'maps hate as love' (ibid:76). By enforcing groupings of people as 'co-relations' instead of correlations, homophily circulates emotions as a form of capital that also breeds 'publics of hate' along with offering opportunities to game the system. In this way, attention-based search algorithms construct strategic behavioural audiences that often lead to confirmation bias as a type of 'information pollution through platform vandalism' **(**Albright 2016 cited by Cadwalladr 2016a).

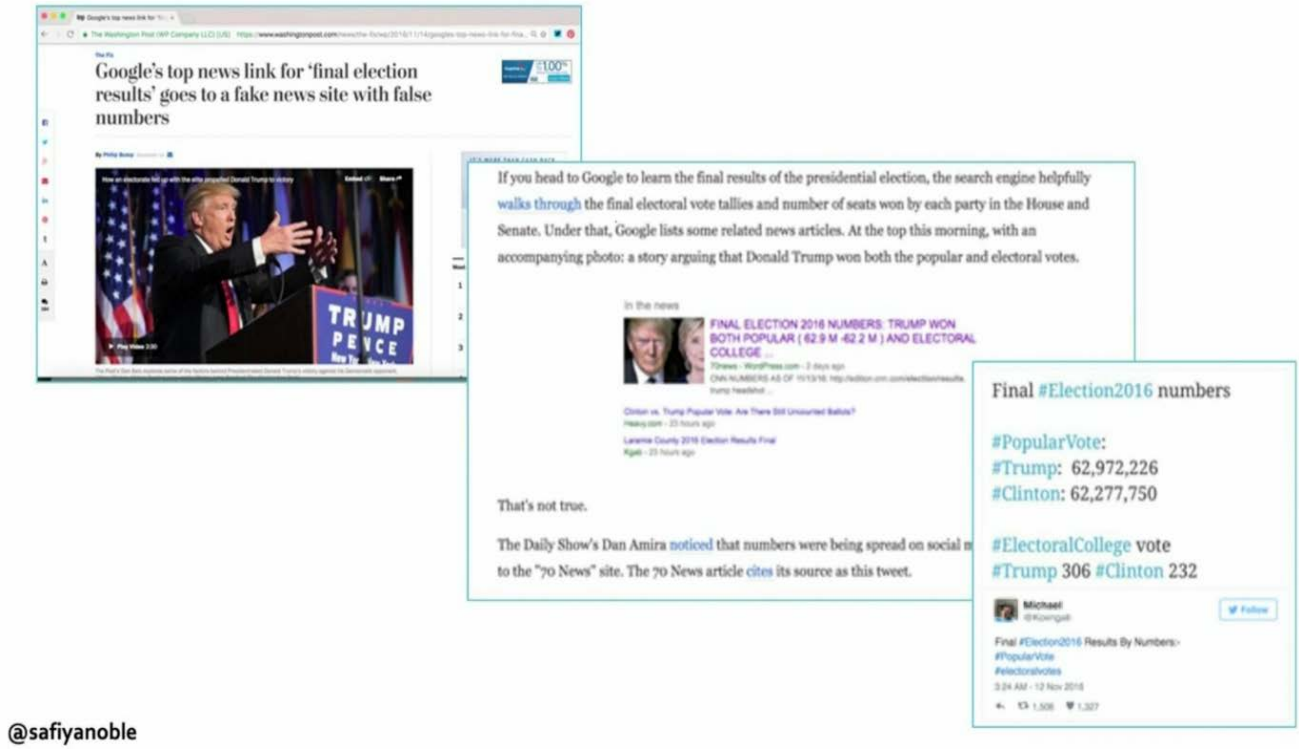


Figure 103: Google's fake news headlines with results of US presidential election 2016 (Noble 2020)

[185] They have 'absolutely no semantic understanding of the meaning or context of the results they deliver' (Vaidhyanthan 2011:22).

1.3 Infrastructuralization of platforms and the platformization of infrastructures

> Google reverse-engineers the web and extracts its intelligence. Blessed are they who know how to read infrastructure (Peters 2015:328).

As mentioned in the Prologue, in the lead up to the 2016 US presidential elections Beqa Latsabidze discovered that by applying certain 'hyperlinking' strategies to direct traffic and increase online advertising revenue through affect, more people would be induced to click on campaign paraphernalia for Trump than Clinton. Latsabidze, a computer student from Tbilisi, ran the 'traffic arbitrage' website Departed (amongst others) thereby generating income (a few cents for each read or click by a user) earned from Google's AdSense advertising network. He later admitted, 'For me, this [clickbaiting] is all about income, nothing more' (Kolbert 2017). Likewise, according to personal testimonies, 'hacker entrepreneurs' organised 'clickbaits' in Macedonia in order to redirect Trump links instead of Clinton links, which would earn them more money with AdSense (Subramanian 2017). Eventually media campaign organisations and third-party actors revealed that there were many factors involved in this hack of what is now, since July 24, 2018, called Google Ads: online targeting based on user profiles, social media 'dark posts' and malevolent actors located in different places in the world distributing 'fake news'. (Figure 103) Organised by these invisible malevolent human actors, 'clickbait and manufactured "news" from all over the world clouded accurate reporting of facts on the presidential candidates' (Noble 2018:42).

These hidden infrastructures promote the interests of paid advertisers, or to be even clearer, 'Google creates advertising algorithms, not information algorithms' (Noble 2018:38). Nonetheless, many users still assume that 'Google is a search engine, rather than a multi[b]illion dollar corporation making large profits from devising personalised advertising schemes' (Stadler and Mayer 2009:99). Google advertising revenue reached 37,93 billion dollars in Q4 of 2019 and CEO Sundar Pichai admitted that 'half of advertisers' Search Spend is now from automated bidding' (Elias 2020).[186] As 'the starting point for billions of dollars of commerce', every minute users ubiquitously google around 3.8 million queries, 'prompting its algorithms to spit out results' (Grind et al. 2019) yet a whole panoply of actions are unfolding behind the scenes (Chapter 5). Thousands of third-party companies compete in silent auctions for the privilege of showing their ad, now called programmatic advertising (AdX, previously known as DoubleClick Ad exchange) that only automated software can handle. However, revenue-producing behavioural advertising is now questionable, because users are not as personalised as they think (as Chapter 5, 9 and above have shown). Moreover, as recent research has demonstrated, targeted advertising doesn't really reach enough desirable audiences effectively and the increase of digital ad fraud, or clickbaiting, is contributing to the 'ad bubble' (Hwang 2020).

The economic culture of Google advertising incorporates 'wider processes of capitalisation' that 'mobilise code and data analytics to compose immanent infrastructures' (Langley and Leyshon 2016:15). This platform of capitalism (Srnicek 2016) drives technological development while at the same time technology configures advertisement markets. As stated in the Introduction with regard to why the 'politics of search engines [still] matter' (Introna and Nissenbaum 2000), it is crucial to emphasise that Google is not a search engine operating as an infrastructure of public good but a commercial service. Chapter 5 showed how Google Search currently serves as a

[186] 'Return on Advertising Spend (ROAS) is a marketing metric that measures the efficacy of a digital advertising campaign'. https://www.bigcommerce.com/ecommerce-answers/what-is-roas-calculating-return-on-ad-spend/

platform that 'intervenes' (Gillespie 2015), keeping users engaged with advertisements as it organises them through the protocol of TCP/IP address within the digital infrastructure—the seven layers of the networked system called the internet. Google is now an example of 'this combined infrastructuralization of platforms and the platformization of infrastructures' (Plantin et al. 2017:301).

Looking back on Chapter 1, the former address offices were platforms of early capitalism yet also infrastructures that organised and distributed information. Additionally, the names of the 'address offices' changed as the centuries progressed, shifting from 'register office', to bureau d'adresse to message house (Berichthaus) and then to the more ominous sounding 'intelligence office' with its connotations of surveillance, before becoming 'question offices' and eventually newspapers. Now the 'media a priori' (Peters 2015) has shifted to search engines. Just as the evolution of the offices' nomenclature reflected their dynamic functions and the services they advertised, it also projected their reputation onto the general public. During the month of August in 2015 Google's metonymia resulted in it quietly rebranding itself as a 'research' company, with the 'Googlization of everything' becoming 'Alphabetized'. However, unlike the former address offices, Google, or rather Alphabet, has yet to disappear. Transitioning from a research-oriented search engine in 1998 (Chapter 3) into a 'multinational advertising company' (Noble 2018:50), 'Google became the company it is today by revolutionizing advertising, not by delivering good search results' (Haeselin 2017:6).

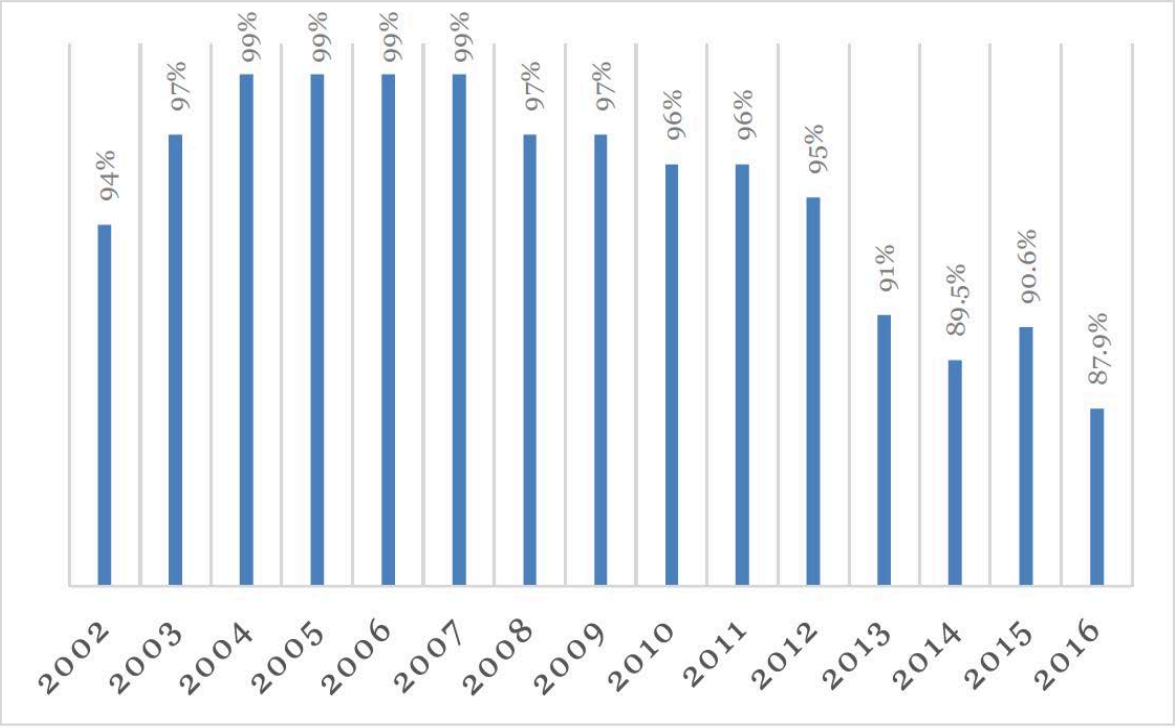


Figure 3: Share of advertising revenue within total revenue (2002-2016). Source: author's analysis based on Form 10-K (2004-2016)

Figure 104: Share of Google's advertising revenue within total revenue (Bilic 2016)

Thus, over time, Google capitalised on the 'informational rationality of generating value from advertising and audience labour' (Smythe 1981; Jhally and Livant 1986 cited by Bilic 2017:8). Eventually its 'database of intentions' (Battelle 2006) or 'vast archive' became Google's 'key monetizable resource, as its contents are sold to advertisers to generate the bulk of the company's revenues' (Jarrett 2014:17). Advertisers wish to place ads on Google because of its dominance in the search engine market, which is more than two thirds in the U.S. and well over 90 % in most European countries (Lewandowski 2017:10). Earning most (86%) of its revenue from advertising, 134.81 billion dollars in 2019 (Srinivasan 2020), it is therefore not a 'search engine'—Google Search is just their most lucrative service (Figure 104). However, with the merger of 'search' and 'research', the advertising company is presently focused on maximising its profit through the creative use of the treasure chest of user data (behavioural surplus) it has extracted and refined through artificial intelligence (AI) such as RankBrain (Chapter 5). Ostensibly the aim is to continue to spend the surplus created by advertising revenue on investments across 'research' industries ranging from abc.xyz (i.e. Deep Mind, Sidewalk Labs, Waymo).

> One could only hope that Alphabet would at least update Google's original mission to make it more accurate: to organise the world's information and make it universally accessible and useful – in selling ads (Morozov 2017).

1.4 Amnesia

As illuminated above, past information creates future events and the Google search engine 'is a convenient target' to elucidate how dominant 'media a priori' as inscription systems have brought about 'existential consequences for memory and human history' (ibid:315). As proposed in Chapter 4 with *hupomnemata*, humans 'act in cognitive collectivities, in symbiosis with external memory systems' (Donald, 1991:382 cited by Parker and Cooper 2016:242), such as the alphabet, writing, calendars, clocks, maps and now computers. This is exemplified by the search subjectivity Cyberorganization (Chapter 8), where the hybrid is the mind of the human mapped onto these 'cognitive architectures' (ibid). Moreover, 'mnemotechnics' such as search engines concern questions of 'how to tag and give order to a mass of materials' (Peters 2015:318). These omnipotent infrastructures facilitate not only communication transfer but the organisation and storage of user data, which is Google's business (ibid). 'Just as writing manages the inhospitable materials of language and voice, so Google deals with the intractable problem of memory's order' (ibid).

Increasingly, these 'external storage media' or 'mnemotechnics' play an important role in altering the 'arts of [human]memory' (Peters 2015:318). Because of Google's 'logic of accumulation' there are 'fairly robust long-term data trails' and the habituation of users to their own connections occurs due to the 'relatively solid longitudinal data set' that keeps track of users' actions over time (Chun 2016:57). Described as the 'politics and practice' of memory as storage, 'traceability has entailed the massive rehabituation of individuals into authenticated users through the expansion and contraction of privacy' (Chun 2016:57-58). As explained in Chapter 9, the tracking and organisation of (us)ers through surveillance along with the storage techniques (databases) of 'big data' enables correlation instead of causality in an era of never-forgetting. 'Rouvroy also contends that by recording everything and relying on nonhuman forms of perception, such systems deny humans the ability to forget and thus to create new norms' (ibid:58). This adaptation not only gives rise to new norms but new forms—data visualisations

constructed from information—that which has been given form from data—for representation, documentation and interpretation.

Serving as memory aids or hupomnemata, the dilemmas surrounding search engines as technological artefacts are manifold, because they have become a normative part of our experience with digital media and computers (Halavais 2009). When querying for information, instead of (us)ers drawing on tacit knowledge or memory or visiting libraries to read reference books such as encyclopeadias, consulting search engines is now a daily habit (ibid). As put forth in Chapter 9, with 'ubiquitous googling' users offer the search engine keywords for things they desire to know, thereby expressing their 'intentions' (Battelle 2006, Halavais 2009, Jarrett 2014, Noble 2018). By organising not only (us)ers but information over time, these media of inscription incite questions of value, validation and trust. Moreover, 'mnemotechnics' are memory tools and important social inventions—artefacts which operate as prescriptive technologies designed for compliance (Franklin 1989).

> Contemporary society is not only changing in terms of what is remembered and forgotten (e.g. digital instead of physical artefacts) but, more importantly, how society remembers and forgets (e.g. search engines instead of catalogues) and, ultimately, social reality is constructed (Berger et al. 1966) cited by (Marton and Kallinokos 2017:8).

2.0 State Agency

The Circle has been devouring all competitors for years, correct? It only makes the company stronger. Already 90% of the world's searches go through the Circle. Without competitors this will increase. Soon it will be nearly 100 percent. Now, you and I both know that if you can control the flow of information, you can control everything. You can control most of what anyone sees and knows (Eggers 2013:482).

2.1. Anti-Trust

As of February 2021, Google still has a monopoly on search, in South America 97,5%, Africa 97%, Oceania 94%, Europe 93%, Asia 91% and North America 89%, with a market share worldwide of around 92%.[187]Although Google reorganised into a holding company in 2015 (Alphabet), ostensibly as an attempt to prevent anti-trust actions and to defer prosecution, in 2020 there have been a slew of lawsuits filed in the U.S. Historical US precedents do exist. Actions taken by former US presidents (Roosevelt 1901, Taft 1911, Kennedy 1962) resulted in facilitating more competition, AT&T was broken up in 1982, yet in 1998 a government antitrust case was unsuccessful in preventing Microsoft from forcing users to install its own browser instead of Netscape Navigator.

> Search engines prescribe a specific mode of being, as normative constructions that have consequences for 'what we can do and say' and thereby need to be regulated because the design not only shapes 'ethical and political precepts' but places them [the powers that shape the technical platforms of our mediated lives] beyond the pale of normative judgment (Nissenbaum 2015:17).

[187] According to Stat Counter: http://gs.statcounter.com/search-engine-market-share

Platforms such as Google Search, however, are difficult to regulate because judges argue that it's about the corporation providing low prices for consumers instead of taking into consideration how many competitors there are. Contemporary 'Silicon Valley capitalism' embodies the hedonistic 'creative' and flex lifestyle with a Google 'too big too fail' mantra inherited from the banking crisis of 2008 (Morozov 2017). Seemingly the most common approach on how to 'fix the tech monopolies' is not to regulate them but to actually allow them a 'permanent monopoly and in exchange we will get them to curb their behaviour' (Doctorow 2019). However, this does not seem to be working. As a result, 'Google search is one of the most under-examined areas of consumer protection policy, and regulation has been far less successful in the United States than in the European Union' (Noble 2018:29).

In June 2017, the EU Commissioner Margrethe Vestager fined Google €2.42 billion for abusing its dominance as a search engine by giving an illegal advantage to Google's own comparison shopping service. In July 2018, the Commissioner fined Google €4.34 billion for illegal practices regarding Android mobile devices to strengthen the dominance of Google's search engine. In March 2019, the Commissioner fined Google €1.49 billion for abusive practices in online advertising. (Figure 105) Fines for anti-competition and anti-trust activities are seemingly not affecting their business model, with shareholders shrugging their shoulders at the meagre penalties. Although Google has appealed, discussions surrounding antitrust issues with Silicon Valley giants are increasingly prevalent and people have now become aware of the problem, yet for some the challenge lies not only in regulation (Hindman 2018, Pasquale 2016) but visibility.

Figure 105: Margrethe Vestager's Twitter post July 18, 2018

Another key issue is about their power and control and how they impede their competition. '[S]omething far less visible to the public' (Noble 2018:35) is how Google 'prioritizes its own properties in keyword searches' (ibid:162) such as books.google.com, as demonstrated by my results in Chapter 5.

> Google's monopoly status, coupled with its algorithmic practices of biasing information toward the interests of the neoliberal capital and social elites in the United States, has resulted in a provision of information that purports to be credible but is actually a reflection of advertising interests (ibid:36).

‘Information monopolies such as Google’ are emboldened with their own self-promotion at the top of the fold with search results, favouring multinational corporations at the expense of ‘less profitable advertising clients’ (ibid:24). A recent investigation exposed these types of practices and showed that Google engineers do intervene by adjusting search results to favour ‘prominent business over smaller ones’ (Grind et al. 2019).

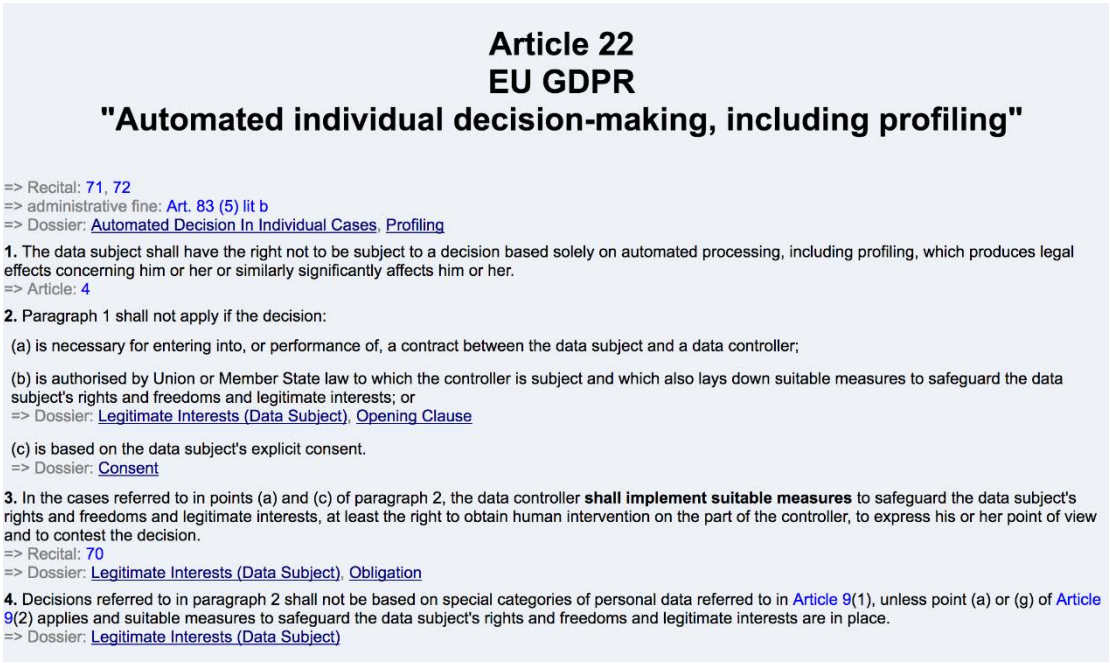
> Google engineers said it is widely acknowledged within the company that search is a zero-sum game: A change that helps lift one result inevitably pushes down another, often with considerable impact on the businesses involved (ibid).

Diversification is also not one of Google’s strongest qualities—rather in recent years it has engaged in takeovers of companies that would be in competition with it or innovate relative to its business practices. ‘We quickly forget all of Google’s failed ventures and overlook that Google’s big successes were acquisitions’ (Hoofdnagle 2018:2). As Google’s profits have grown, concomitantly there have been demands by lawmakers to turn back deals such as the acquisition of DoubleClick for 3.1 billion dollars as mentioned in Chapter 5. With this takeover (which was approved by EU regulators), not only did Google increase revenue but it enabled tracking and collecting data by storing ‘cookies’ on users’ computers. Although DoubleClick was Google’s primary source of revenue, it downplayed the idea that data sets would be merged and this continued until 2017, when it did just that. By instating complicated terms of services, over the previous 20 years users have agreed to the collection of data that is simultaneously combined with data from other contexts, with this ‘mission creep’ increasing the depth and detail of an individual’s profile, as described in Chapter 9.

2.2 GDPR and Contextual Integrity

To return to the Prologue, on February 2, 2016 Mr Schrems’ actions brought about a new framework, the EU–US Privacy Shield, which went into effect concerning the Safe Harbour ruling and the control of data as a prime commodity. European ‘data subjects’ will now not only have the right to question companies’ decisions, which must comply with deadlines, answer to limitations and oversight, but have the right to appeal, free of charge (Savin 2018:295). As of 2020 the European Digital Services Act is in the making and would force Big Tech companies such as Google to share customer data files with smaller competitors. Additionally, certain regulation has been introduced in California and increasingly in other US states with regard to users’ privacy rights along with the European General Data Protection Regulation (GDPR) that was enacted on May 25, 2018. The GDPR protects the privacy of the ‘data subject’ (Chapter 8) by forcing companies to obtain users’ permission to collect certain types of data and transfer personal data outside of the EU. Although the GDPR attempts to make the policies and terms of service clearer in order to understand consent agreements, users are now asked to update their preferences when they visit a website in relation to ‘functionality’ and ‘statistics’ and to adjust their degrees of ‘personalisation’.

Additionally, some websites actually list all of their advertising affiliates but because there are thousands, users would have to spend a large amount of time clicking and researching their privacy policies. However, the launch of a ‘global privacy control’ in October 2020 by a conglomerate of privacy-supporting tech companies and publishers instates a universal ‘opt-out’ setting in the browser or device, attempting for once, to legislate ahead of technology.



Article 22
EU GDPR
"Automated individual decision-making, including profiling"

=> Recital: 71, 72
=> administrative fine: Art. 83 (5) lit b
=> Dossier: Automated Decision In Individual Cases, Profiling

1. The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.
=> Article: 4

2. Paragraph 1 shall not apply if the decision:

(a) is necessary for entering into, or performance of, a contract between the data subject and a data controller;

(b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or
=> Dossier: Legitimate Interests (Data Subject), Opening Clause

(c) is based on the data subject's explicit consent.
=> Dossier: Consent

3. In the cases referred to in points (a) and (c) of paragraph 2, the data controller **shall implement suitable measures** to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.
=> Recital: 70
=> Dossier: Legitimate Interests (Data Subject), Obligation

4. Decisions referred to in paragraph 2 shall not be based on special categories of personal data referred to in Article 9(1), unless point (a) or (g) of Article 9(2) applies and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place.
=> Dossier: Legitimate Interests (Data Subject)

Figure 106: Article 22 of the GDPR

Supposedly, if companies such as Google violate GDPR regulations they would have to pay up to 20 million dollars or 4% of their global annual revenue, which, for 2018, was approximately 136.8 billion dollars. Moreover, Art. 22, 'Automated individual decision-making, including profiling', enables data subjects to exercise their right to obtain human intervention in order to contest the algorithmic decision-making. (Figure 106) Yet the chronology of being informed and the right to explanation along with exceptions (crime-fighting) weaken the legislation (Tiku 2018). Furthermore, not only does legislation enforce the right to privacy of 'data subjects' but crucial privacy issues lie within contexts and the 'infrastructures and the organizations that deploy them' (Forte et al. 2017:2). Although there are 'inappropriate flows' of personal information that transgress 'informational norms', Helen Nissenbaum's 'contextual integrity' stipulates how those revelations are managed (ibid).[188] There are various conditions, agreements or 'principles' regarding how information is transmitted i.e. 'in confidence,' 'with third-party authorization,' 'as required by law', 'bought,''sold,' 'reciprocal,' and 'authenticated' (Nissenbaum 2015:9-10). Contextual integrity is successful when these informational norms are maintained yet when people are shocked or feel that their privacy has been compromised contextual integrity has then been violated.

As 'constituents of differentiated social space', contexts serve as 'organizing principles for expectations of privacy' (ibid:8). Nissenbaum's contextual integrity for privacy research connects back to Gary Marx's sociologically defined contexts for concealment and revelation that encompass the concept of privacy, each contingent on space, place, sphere or setting (1999), as described in Chapter 7. '[P]rivacy, as appropriate information flows, serves not merely the interests of individual information subjects, but also contextual, social ends and values' (Nissenbaum 2015:12). These 'moral social and political values' include fairness and ethical values that are crucial in addressing 'democracy, unfair discrimination, informational harm,

[188] Nissenbaum's text, *Respecting Context to Protect Privacy: Why Meaning Matters*, draws on US legislation that the Obama administration supported called a *Privacy Bill of Rights* (2012) containing seven principles. The third, entitled 'Respect for Context' questions when privacy is crucial and that the collection of data is only used for that which it has been collected or whether it can induce change or 'devolves to business as usual' (2015:3).

equal treatment, reputation, and civil liberties' (ibid:3). Privacy is thus embedded in the social sphere as contextual integrity, which embodies the socio-technical flows of information. If only because, '[c]ontextual integrity reveals the systematic dependencies of social values on appropriate information flows, once-and-for-all challenging the fallacy of privacy as valuable for individuals alone' (ibid:16). With this, Nissenbaum echoes not only Shoshana Zuboff's plea for the contractual and entering into new forms of law-making through regulation but also that change—breaking away from 'business as usual'—stems from the social.

Changing technologies involved in these processes or 'media a priori' such as search engines demand updated legislation. As outlined in Chapter 1 with regard to the policing and surveillance activities of the address offices (Black Utopia), despite the collection of data in public and the profiles constructed in the registers clients continued to expect that their privacy would be maintained. Communication systems and forms of address, whether the 'intelligence forces' use of the offices' registers or the monitoring of Olmstead's telephone (Chapter 6), reflect the way in which data is gathered by institutions and governmental agencies. Today's computers, IP protocols and Google's rapacious appetite for data—not only routing information but content (Appendix H), have usurped the telephone and the 'pen register' on which data legislation has historically been based,

> To this day (underscored by the Snowden revelations) we are living with the consequences of legislation that attempted to define duties of phone companies, and the varied types and degrees of access they (and others) would have to the new forms of data generated by the telephonic medium, from pen register data to content of phone calls (ibid:16).

New technologies produce new categories as well as new actors, information and transmission principles, along with redefining users' 'expectation of privacy' when using anonymising technologies such as Tor (Chapter 7). Thus, to adapt to new media, there needs to be new kinds of legislation in place.

2.3 Rule 41, NSA's PRISM, Keyword search warrants

As discussed in Chapter 7, with 'Search and Seizure', the distinction between the private and public spheres also surfaces with regard to obtaining warrants for certain contexts including computers, the internet, ISPs and the Tor browser. Although the Fourth Amendment demands that warrants need to describe the place to be searched, in previously defined physical 'jurisdictions' on US soil, Rule 41 is an all-encompassing single search warrant that enables the FBI and other agencies to carry out 'remote computer searches' globally, searching computers, devices and phones to include a wider sweep of surveillance. With Rule 41, governmental agencies are now able to 'remotely search' and 'seize' digital information that has been intentionally concealed with anonymising software (such as Tor or TAILS) by accessing and then copying information on any connected device, along with capturing any IP address of a computer connected to the internet.

As shown in the Prologue in the imaginary *House of Cards,* the search engine *Pollyhop* can turn searches into votes for the challenger but the incumbent President Underwood has a more powerful weapon—the NSA, the world's largest surveillance organisation. Snowden's revelations (2013) exposed the NSA PRISM programme, which, according to a Powerpoint

slide, has been able to directly access the servers of tech companies such as Google since 2009, collecting communications such as search histories. (Figure 107) Redefining 'electronic surveillance' by including 'the content of communications and not just the metadata', the NSA supposedly obtained intelligence information without a warrant, instead with the 'participation of US internet firms, claiming "access is 100% dependent on ISP provisioning"'(Greenwald and MacAksill 2013). As time went on it became apparent that 'the data was turned over to the US FBI by the companies under lawful access procedures and then shared with the NSA' (Deibert 2020). The 'nightmare' of the so-called transparency society mentioned in the Prologue defines how companies such as Google's *Big Other* maintain corporate secrets by 'remaining intransparent themselves' (Beyes and Pias 2019:86) whilst 'surveilling and capturing' (Agre 1999) enormous amounts of data on individuals, as shown in Chapter 9, which is then shared with governments.

> Although Agre separates capture and surveillance systems, it has become clear that not only can the two not be separated; but also that their initial separation was based on a fictional distinction between state and private corporations. Snowden's revelation about the extent of cooperation between the U.S. NSA (National Security Agency) and corporations, such as Google.com and Facebook.com, reveals this nicely. Again, the protests by these corporations against the U.S. government ring false, not only due to the history of their cooperation with the state, but more importantly because the value of the NSA's data stems from corporations insisting on real names and unique markers. *Surveillance has become a co-production* (Chun 2016:121 emphasis mine).

Figure 107: NSA Powerpoint released by Snowden. https://nsa.gov1.info/dni/prism.html

Although Tor obfuscates the IP address of 'anonymous' users through layers of encryption, one of the consequences is that the IP addresses of Tor users can be collected and correlated with other data, with much effort, tech-savvy and resources, in order to deanonymise users. Who are these anonymous users that could be considered suspicious persons? In October 2020 it came to

light that in response to a 'keyword [search] warrant', Google discloses to the police people's IP address who searched certain terms, which enabled investigators to match this information with phone numbers that in turn enabled them to pinpoint the suspect's location, obtained by 'geofence warrants' (Ng 2020).[189] As explained in Chapter 7, usually probable cause is needed for a warrant. However, 'keyword search warrants' are targeting groups of people and these 'reverse search warrants' have been challenged as they infringe upon civil rights protections and breach the Fourth Amendment (ibid). Keenly, Zuboff asks who [other than Google] is learning from the global data flow that is collected, how is it accumulated and what if there is no oversight and 'authority fails'? (2015:77). Therefore, besides updated privacy legislation and its enforcement, state agencies also need to be held accountable when they do not uphold constitutions or laws.

2.4 Right to be Forgotten

Nowadays users rely on search engines as a substitute for what was once publicly funded information sources, 'libraries, teachers, books, histories and experience' (Noble 2018:116). The societal implications of the construction of knowledge by commercial search engines should not be underestimated as they also 'function as a type of personal record and as records of communities, albeit unstable ones' (ibid:35). The visibility of websites' results 'wield tremendous power' and therefore the responsibility for information lies with 'Google and other large monopolies in the information and technology sector' because search engine results are records of 'human activity' (ibid:122-123). Moreover, the complexities of privacy and profiling in relation to Google Search is based on two groups—those being searched and those searching (Tene 2007). Chapter 9 explained how user's habit of 'ubiquitous googling' facilitated years of 'Google's meticulous collection of each user's search queries and their retention in search logs' that in turn is 'yielding increasingly detailed profiles ripe with personal information' (Tene 2007:4).

The flip side of the coin is those being searched and what can be discovered about them online. As danah boyd keenly pointed out in Chapter 6, the observer armed with a search engine has greater control over the social situation (results) than the person presented. Referencing the Prologue, the *Right to be Forgotten* addresses the former by the actions of the latter as Google indexes the web as a commercial entity that controls the record, regulating and constructing not only privacy but identity ownership (Noble 2018:129). Since 2014 it is possible for EU users to request that damaging or false information be removed from search results through legislation. In this way the *Right to be Forgotten*, 'to become anonymous', is an 'incredibly important mechanism for thinking through whether instances of misrepresentation can be impeded or stopped' (ibid:123). However, historically Google does not wish to comply with 'requests to alter their content moderation practices' unless they are legally forced to do so; instead they respond by placing 'responsibility for altering search results on information providers and seekers' (Mulligan and Griffin 2018:559).[190]

---

[189] 'Google received 15 times more geofence warrant requests in 2018 compared with 2017, and five times more in 2019 than 2018 […]The company declined to disclose how many keyword warrants it's received in the last three years' (Ng 2020).

[190] 'Google says it has received 845,501 "right to be forgotten" requests in the past five years, leading to the removal of 45% of the 3.3m links referred to in the requests. Although the content itself remains online, it cannot be found through online searches of the individual's name' (Marsh 2019).

Individuals or 'data subjects' can continuously request their 'right to erasure' (Article 17 GDPR) yet 'more attention needs to be paid to information that is collected and archived by Google that is not visible to the public' (Noble 2018:129). Although Google can be forced to delete 'objectionable hits' it becomes 'equally troubling, because the company provided search results without informing searchers that information was being deleted' (ibid:45). In her text *Algorithmic Memory and the Right to Be Forgotten on the Web*, Elena Esposito expresses concern as to the elimination of unwanted search results from Google, as their storage in databases remains problematic:

> This is a great opportunity but also the most serious threat to the freedom of self-determination of individuals and to the possibility of an open future. Information can be forgotten in the indexing in accordance with the right to be forgotten, while data continue to be remembered and used by the algorithms to produce different information (Amoore/Piotuck 2015:355 cited by Esposito 2013:17).

3.0 Computational Agency

'Someday there'll be search engines you can just input a little spritz of anything and voilà, nowhere to run to, nowhere to hide, the whole story will be there on the screen before you can scratch your head in amazement' (Pynchon 2013:209).

3.1 Scopophilia

As related in the Prologue, in 2016 Fancy Bear queried 'company's competence' to see if their hack had been already been indexed by Google Search. Google's index is the largest in the world and their 'organizing by indexing' webpage proclaims that its search index

> contains hundreds of billions of webpages and is well over 100,000,000 gigabytes in size. It's like the index in the back of a book—with an entry for every word seen on every webpage we index. When we index a webpage, we add it to the entries for all of the words it contains (2019).[191]

In order for users to share links and decide what's valuable, websites (some of which are the users' own) first need to exist online. Applying an analogue analogy pertaining not only to search but research, '[a]n unindexed Internet site is in the same limbo as a misshelved library book' (Peters 2015:329). As explained in the Introduction and Chapter 2, Google 'shapes the web' by indexing websites, which need visibility so that they can engage with consumers. However, individuals also need to accrue value within their attention economy, as shown in Chapter 3, thus 'to exist is to be indexed by a search engine' (Halavais 2009:105). Simply put, increasingly 'the record, then, plays a significant ontological role in the recognition of the self by existing, or not, in an archived body of information' (Noble 2018:126).

In 1971 Herbert Simon declared that in an era of an abundance of information, there is a dearth of attention, which is controlled by a scarcity model. However, today attention also means recognition and high ranking in Google search results, because '[e]very author is then faced with the question of how to draw enough attention to their pages to allow them to be seen by large audiences' (Halavais 2009:67-68). It is Google's 'impact factor' or what is often referred

[191] Available here: https://www.google.com/search/howsearchworks/crawling-indexing.

to by the SEO industry as 'Google Juice' that commensurates 'the strength of a site or document's position in the web as defined by incoming web links'(Peters 2015:328). This is 'identical' to Garfield's SCI (ibid), as demonstrated in Chapter 3 with regard to hodology organised by sociometry and bibliometrics (Rieder 2012). PageRank displayed the '*authoritative*, or, to connect back to citation analysis, the *canonical*' (Rieder 2012) yet with economies of attention (network surplus value), ranking shows a *meritocracy* or, through advertising and traffic *(economic forces),* 'the map of a *plutocracy'.* To repeat: 'The search engine as a visibility engine subjects both to the self-reinforcing dynamic of *cumulative advantage*' (ibid).

Therefore, in the present online attention economy, 'the constantly updated map' of the 'web cosmos' that Google's spiders weave fulfils the dream of completion: 'I am tagged, therefore I am' (Peters 2015:329). Peters furthermore expounds upon the merger of research and search with this quote referencing the 'reading logic of PageRank' that 'mimics academic prestige systems':

> Professors love to read; and even more, they love to write. Even more than writing, they love to publish. But even more than publishing, they love to be read. Better than being read is being cited. Even better than being cited, however, is being cited by someone important. And how do you know who is important? By citations, of course: an important scholar gets cited a lot. Scholars who are cited by other scholars a lot confer greater authority when they cite another scholar: they channel the power of their inlinks (Peters 2015:328).

This contemporary attention economy with Google Search is an extension of Bentham's definition, where publicity has come to mean 'press attention' (Lepore 2013). As demonstrated in Chapter 9, with regard to surveillance capitalism, Google's powerful record-keeping apparatus captures (us)ers 'grammars of actions' (Agre 1994) with their 'ubiquitous googling': 'I search therefore I am' (Hillis et al 2013:19). However, this reproduction of the 'liquid modern' includes recognising oneself in search results as a 'consumerist introversion that is simultaneously and paradoxically a form of extroversion, a desire for publicness' made possible by the 'willing involvement of consumers in their own surveillance' (Baumann and Lyon 2013:107). In her book, *Configuring the Networked Self*, Judith Cohen draws on Jodi Dean's argument that 'the credo of openness drives a political economy of "communicative capitalism" organized around the tension between secrets and publicity' (2012:147). Furthermore, Cohen elucidates the complexities and consequences of surveillance and subjectivity on the user, through certain 'norms of transparency' that are 'deployed to legitimate and reward practices of self-exposure' (2012:135). This culture of self-exposure relates to voyeurism and agency with respect to 'coveillance', what she calls the 'Culture of the Spectacle' (ibid).

Both the 'rise of both the public eye (the eye of the citizen, and of the reporter) and the private eye (the eye of the detective)' (Lepore 2013), self-surveillance, along with the monitoring of others, has nowadays became commonplace. Although people are protecting their privacy they participate in the attention economy, as with Chapter 8's Homoeconomicus, 'watching themselves, and one another, refracted, endlessly, through a prism of absurd design' (ibid). Part of a continuation of Western culture, it is the 'love of being seen' or 'scopophilia' that 'merges with the growing ubiquity of surveillance practices, with several striking effects' (Baumann and Lyon 2013:108).

> Hegel famously defined freedom as necessity learned and recognized … The passion for putting oneself on record is a foremost, perhaps the most blatant example of that Hegelian rule in our times in which the updated and adjusted version of Descartes' cogito is "I am seen (watched, noted, recorded) therefore I am" (ibid:110).

3.2 Gatekeeping & Calculated Publics

> [Algorithms] are arguably the most powerful lines of computer code in the global economy, controlling how much of the world accesses information found on the internet (Grind et al. 2019).

Chapter 5 (Authorship) elucidated how the SEO industry's 'reverse-engineering' of algorithms in order to optimize results for their customers is an inherent part of the 'political economy of search and traditional advertisers' (Noble 2018:49).'[D]riven by economic imperatives' Google is one of the main actors in this political economy with its monopolistic position that includes influencing 'broad swaps of society to see it as the creator and keeper of information culture online' (ibid:86). By deciding the hierarchy of search results and exerting control over 'online data generation and distribution' (Stone 2018:3), 'Googlearchy' becomes 'another form of American imperialism that manifests itself as a "gatekeeper" on the web' (ibid:86).[192] Armed with proprietary algorithms, Google now dictates what gets disseminated, the price of information acquisition as well as exerting bias and discrimination, such as with solicitations.

> The unequal treatment at the hands of these gatekeepers extends far beyond résumés. Our livelihoods increasingly depend on our ability to make our case to machines. The clearest example of this is Google (O'Neil 2016:114).

As shown in Chapter 3 and 5, search algorithms determine 'relevance' and the expertise involved is programmed into the computational process, 'but it is judgment that is then submerged and automated' (Gillespie 2014:178). Borrowing the term 'actants' from 'actor-network-theory, algorithms are computational agents that are not alive, but that act with agency in the world' (Tufeksci 2015:206).[193] Drawing on the work of Ivan Illich from 1973, there are 'convivial' technologies that as they become more powerful, tend towards becoming 'non-convivial' tools, or expressed otherwise '[a]s the power of machines increases, the role of persons more and more decreases to that of mere consumers' (1973:17). Echoing Zuboff in Chapter 9 as to how 'convivial' tools 'learn' human behaviour, as these technologies 'acquire more information about consumption habits' they simultaneously 'may begin to exert non-convivial, autonomous technological action' (Tufekci 2015:205).

This 'computational agency' raises new concerns, challenges and consequences, some of which pertain to the intermediation of news content and civic information.

---

[192] 'I make this claim on the basis on the previously detailed research of Elad Segev on the political economy of Google' (Noble 2018:86).

[193] Zeynep Tufeksci differentiates between those algorithms that are not just mathematical, producing 'correct' answers by 'alphabetically sorting a database of names or calculating the average sales per employee', and those that are 'utilized as subjective decision makers' (2015:206).

> It [Google] desperately does not want to be seen as a media company, as a content provider, as a news and information medium that should be governed by the same rules that apply to other media. But this is exactly what it is (Cadwalladr 2016).

Although Google proclaims that it is not a media company, or states it is not a publisher, it prefers certain content [its own] and by ranking it, exhibits editorial judgment similar to the editors of newspapers, even if this decision-making is carried out by algorithms. Intellectual material as well as all forms of media, including news, is limited by such filtering. With this editorial judgment, Google preferences certain content over others, which is increasingly 'algorithmically' executed. Furthermore, the procedures of algorithms are 'unavoidably selective' by what they show and what they hide, thereby 'emphasizing some information and discarding others, and the choices may become consequential' (Gillespie 2014:191).

> There is the distinct possibility of error, bias, manipulation, laziness, commercial or political influence, or systemic failures. The selection process can always be an opportunity to curate for reasons other than relevance: for propriety, for commercial or institutional self-interest, or for political gain (ibid).

Figure 108: 'Move to Gibraltar' search spike. Image source: Twitter Google Trends from 24 June 2016

To return to the Prologue again, when people searched online for the results of the UK referendum in 2016, Google Trends captured the raw data and offered a glimpse into *how* people are thinking at the moment they type words into a searchbox. In his article, *#trendingistrending: when algorithms become culture*, media theorist Tarleton Gillespie

proposes that 'trending algorithms' offer not only a glimpse into what many citizens online are searching for, but 'trending' is 'an oblique category' that standardised 'measures rarely unpack' (2016:60). (Figure 108) Unbeknownst to the user, search algorithms feed back results based on 'some undefined population of users who have found these sites relevant, and have left their assessment in traces like incoming links and past clicks' (ibid:67). Like a focused lens 'Trends data' collects real-time (and non-real time) searches as an anonymised unbiased sample, which is categorised based on topics and then aggregated, or grouped together (Rogers 2016). In this way they are 'hieroglyphs', without numbers or subscribers or an amount of likes whilst they don't even 'claim to represent a verifiable fact', nor do they have to reflect a correct reality; however, they cannot be incorrect (Gillespie 2016:60).

> But they do aspire to say something about public attention, beyond the user-selected community of friends or followers; they say something — perhaps implicitly, perhaps incorrectly— about cultural relevance, or at least we are invited to read them that way. They crystallize popular activity into something legible, then feed it back to us, often at the very moment of further activity (ibid).

'Trending algorithms' that captured users' searches on the referendum vote in the UK unite publics who share similar interests 'rather than fracturing them', with Gillespie noting that simultaneously they 'privilege some publics over others' and can be biased (ibid). Moreover, as shown in Chapters 5,8 and 9, these trending algorithms construct a 'we' and YOUs based on similar interests, location or thousands of other categories, making explicit what Gillespie deems 'calculated publics' (ibid:67). With their cycles of anticipation, the production of calculated publics—the algorithmic presentation of publics back to themselves—'shape a public's sense of itself' yet who is being left out in the measurement and who is being 'calculated'? (Gillespie 2014:189).

> And perhaps most important, how do these technologies, now not just technologies of evaluation but of representation, help to constitute and codify the publics they claim to measure, publics that would not otherwise exist except that the algorithm called them into existence? (ibid).

In the year of 'post-truth' (2016), during the lead up to the UK referendum, Google Trends became an elucidation of what *The Filter Bubble* predicted in 2011. One of the civic consequences of filters and personalisation includes the dangerous way it shapes agendas in traditional media (Hindman 2018) through ranking, therefore reaching only specific audiences or calculated publics. What is made visible and what is seen by certain groups of users and not by others has questioned the very idea of access to information, a shared understanding of 'truth' in the public sphere and what this means for democracy.

3.3 Arbiter of Truth

> Public objections are grounded in well-founded imaginaries of search engines not merely as providers of relevant information, but at least with respect to human rights atrocities such as the holocaust, as stewards of authoritative historical truth (Mulligan and Griffin 2018:558).

When UK residents were searching in the lead up to the referendum in 2016, it is still not clear if their queries were influenced by 'autocomplete', which offers suggestions when a user types a query, or if they reflect a corpus of the exact same questions. As mentioned previously Dylann Roof's searched with 'black on white crime' in 2015 and, in 2016, journalist Carolyn Calladawr received 'did the holocaust happen' when she queried 'did the hol'. Perhaps because of the implications surrounding Roof's killing spree and holocaust denial, Google no longer calls autocomplete 'suggestions' but 'predictions'. Additionally, Google has continued to update its policy, claiming its efforts will improve the overall system of autocomplete and that its 'guiding principle' is that it 'should not shock users with unexpected or unwanted predictions' (Sullivan 2018).[194] Therefore, in recent years Google has been adjusting its autocomplete algorithms, stating that it no longer allows predictions concerning 'sensitive and disparaging remarks' (Grind et al. 2019), however its moderation continues.

As demonstrated previously, users believe in the veracity of algorithmically determined and filtered results by instantiating trust in Google to offer them the path to truth through its organising of information, which is viewed as 'depoliticized and neutral' (Noble 2018:118, 25). Marketed to the public as a neutral technological tool, Google Search has become 'normative' for users, engendering the assumption that 'algorithms can only bring multiple benefits to humanity' (Bilic 2017:4). In spite of the promise of objectivity, 'algorithms are made and remade in every instance of their use because every click, every query, changes the tool incrementally' (Gillespie 2014:173). Although Google promotes itself as a neutral purveyor of information (Introna and Nissenbaum 2000, Vaidhyanathan 2011, Castells 2013, Gillespie 2010, 2012 and 2014, Crawford and Gillespie 2014, Noble and Roberts 2015, Pasquale 2015, Tripodi 2017) and likes to come across as an objective platform, with its commercial search activities it 'prioritizes results predicated on a variety of factors that are anything but objective or value free' (Noble 2018:65). As shown by Chapter 5 with the advertisement actuation of Platform Capitalism and Surveillance Capitalism in Chapter 9, computational transactions orchestrate how 'platforms intervene' (Gillespie 2015). However, 'Google is not 'just' a platform; it frames, shapes and distorts how we see the world' (Cadwalladr 2016).

Chapter 3 elucidated search engine bias from the late 1990s until 2009 (Van Couvering 2010) and in the last decade there has been increasing and widespread concern. Often bias is implicit in algorithmic modelling along with questions about what data the algorithms have been trained on (O'Neil 2016), yet there are those who believe that Google is the 'only way to get "unbiased" information' (Bilic 2017:4). Whereas journalists and librarians are expected and entrusted by the public to 'fact-check and curate information', the legitimacy of websites is [now] taken for granted' (Noble 2012:41). Additionally, in regard to the dissemination of online

[194] 'Google admits that they can't catch everything racist, instead they ask the user to submit free feedback by clicking on the "Report inappropriate predictions" link that now appears in the lower right hand corner, below the search box' (Sullivan 2018).

(mis)information, Safiya Noble asks: 'How does a group resolve the ways that the public engages with Google as if it is the arbiter of truth? (2018:125). Some technology critics such as Evgeny Morozov suggested 'having search engines identify and label suspect sources as "compromised"' yet digital sociologist Noortje Marres retaliates that 'in setting up the search engine as arbiter [of truth], Morozov's proposal effectively places these powerful digital platforms themselves beyond the reach of "bias critique"' (2015:656).

Furthermore, Cathy O'Neil argues that it's becoming increasingly harder for the public to obtain 'good' information because algorithms, or what she deems 'weapons of math destruction', are being used to determine how users obtain information about the information itself (2016). Fact-checking 'neutrality' rests on a 'basic misunderstanding of how Google curates results' which in turn, 'can fortify people's existing ideological beliefs, even if they are using Google to "verify the facts" or challenge their convictions' (Tripodi 2017:38).

> A 2016 internal investigation at Google showed between a 10th of a percent and a quarter of a percent of search queries were returning misinformation of some kind, according to one Google executive who works on search. It was a small number percentage-wise, but given the huge volume of Google searches it would amount to nearly two billion searches a year (Grind et al. 2019).

Algorithms are both socially constructed and productive—in the way they organise reality by 'producing as they select, filter and frame information and create facticity' (Weiskopf 2016:9). Even 'organic' (non-commercial) algorithmically determined search results are based on Google criteria. Having recourse to 'Why the politics of search engines [still] matter' by Introna and Nissenbaum (2000) in the Introduction, politics matter now perhaps more than ever because of these 'curation' activities.

> Public controversies over the results returned by search engines to politically and morally charged queries evidence the growing importance, and politics of corporations' content moderation activities (Mulligan and Griffin 2018:558).

Truth is now determined by proprietary and obfuscated decision-making algorithms and their implications in the realm of politics, culture, social injustice and dis/misinformation, can no longer be denied.

3.4 RankBrain's machine seeing

As explained in Chapter 5, Google's black-boxed algorithmic curation is not transparent and 'algorithmic manipulation is neither public, nor visible, nor easily discernible' (Tufeksci 2015:216). Instead of straightforward procedures, proprietary algorithmic design dictates 'the combination of their labyrinthine components, processes and opacity, and the subjective nature of the decisions reached' (ibid:206). Yet there are 'quests for accountability and transparency' (ibid:208-9) along with calls for regulation to understand how algorithms work and if they actually do what they were programmed to do. This 'infatuation with transparency as a compensatory gesture' (Beyes and Pias 2019:87) was already argued by Chun in *Programmed Visions* (2004), where the more we are shown the less we can see in relation to ubiquitous [search] interfaces.

> As our machines increasingly read and write without us, as our machines become more and more unreadable,…we the so-called users are offered more to see, more to read. The computer—that most nonvisual and nontransparent device—has paradoxically fostered "visual culture" and "transparency" (Chun 2004:27 cited by Beyes and Pias 2019:87).

Progressively, computers and technology 'complicate the field of visuality' and systems of knowledge/power have become more visual, along with the 'points of slippage or instability of epistemic authority' (Berger 1972 cited by Cox 2016). In his article 'Ways of Machine Seeing', the software studies theorist Geoff Cox draws on the 1970s BBC series 'Ways of Seeing' in which the critic John Berger explained to the audience: 'The relation between what we see and what we know is never settled' (Berger 1972). Berger's warning is to 'remind the viewer' that the technical 'arrangement' of elements are recombined for specific purposes and 'to remain sceptical' of these organisatory structures or configurations. 'Algorithms are ideological only inasmuch as they are part of larger infrastructures and assemblages' and this brings into question exactly 'under what conditions' does the unstable relationship 'between seeing and knowing' exist (Cox 2016). It is not a question of 'whether machines can see or not' but rather that 'we should discuss how machines have changed the nature of seeing and hence our knowledge of the world' (ibid). Drawing on the scholarship of Adrian MacKenzie (2015:431), Cox articulates the effects, which also harken back to Chapter 8's Subjectivities of Search and Zuboff's 'behavioural surplus' (Chapter 9):

> There is a sense in which the world begins to be reproduced through computational models and algorithmic logic, changing what and how we see, think and even behave. Subjects are produced in relation to what algorithms understand about our intentions, gestures, behaviours, opinions, or desires, through aggregating massive amounts of data (data mining) and machine learning (the predictive practices of data mining) (Cox 2016).

According to the SEO company MOZ, in a 'pre-RankBrain' era the SEO industry was able to map the traditional signals that determined ranking and there have been constant updates in recent years, as demonstrated in Chapter 5. Concerning the 'Authorship' of the black box, the machine-learning RankBrain has been implemented for unfamiliar searches, with Google now applying AI to decide how and which information should be retrieved and displayed. Nowadays there is enough computational power to handle large amounts of data (O'Neil 2016:76), along with RankBrain 'learning' from previous queries and interpreting why users click on certain results. With complex mathematical formulas (algorithms) playing an important role in the automation process, machine-learning technology now disrupts the human ontologies and taxonomies of keywords that previously structured both queries and the corresponding search results.

Contemporary consternation continues with SEO industry consultants declaring that 'the algorithms remain borderline indecipherable' and that it's 'black magic' (Grind et al 2019). Even Google's algorithmic programmers and experts state that since the introduction of RankBrain they don't know how the search algorithm works, which performs as if it 'acts' without the assistance of humans.[195] Moreover, Google employees are forced to sign non-disclosure agreements (NDA) that prevent them from discussing present and past work, which other programmers and the public could learn from. Algorithms contained within these black boxes make decisions and neither observers (Glanville 2009) nor 'those who programmed it nor

[195] Conversation with former Google data scientist. Stephens-Davidowitz. October 30, 2019.

those who entered the data knew of the output or could have predicted its particularities' (Esposito 2015:09). This reinforcement exemplifies the Cartesian method of rationalism, or a deterministic approach where analysis is carried out by a series of black boxes embedded with functions or algorithms. Computational processes 'are intertwining with wider economic, political and cultural forces' (Soon 2016:73) and with regard to the 'ethical and juridical attribution of decisions and their agential consequences', autonomous decision-making 'is transferred to increasingly intransparent machines' (Beyes and Pias 2019:100).

3.5 Habit of Automaticity

> As I dug further and deeper down into the sedimentary layers of media cultures, it was Kittler's understanding of how language is not something that we speak but rather language speaks us, whether that be the typewriter, the computer, the software (hardware) and now perhaps neural-based networks—they impose new regimes of sensation and use to which we have to accommodate ourselves in order to be functioning subjects. We are secondary to such systems. Besides agency, this has to do with power… (Parikka 2012:70).[196]

As explained in Chapter 5, algorithms are not only complex, opaque and but proprietary, with Google's exercising its black boxed (in)visibility management (Flyverbom et al. 2016) through a series of patents that keep their exact workings hidden, what Jonathan Zittrain deems 'security through obscurity' (Grind et al. 2019). Returning to Chapter 6, privacy and secrecy are still intertwined but with surveillance capitalism, they are 'moments in a sequence', with secrecy 'an effect of privacy, which is its cause' (Zuboff 2015:82). Chapter 9's Black Utopia showed how '[s]urveillance capitalists have extensive privacy rights and therefore many opportunities for secrets' (ibid:83). As described in State Agency above, surveillance capitalists 'skillfully exploited a lag in social evolution' by getting ahead of legislation for users' privacy with their 'surveill for profit' abilities (ibid). Google can invoke its privacy rights 'as legitimation for maintaining the obscurity of surveillance operations', yet with the 'work of surveillance' users' privacy rights are eroded (ibid:82).

As elucidated in Chapter 9 and Ad Agency, this facility to control users' privacy rights is contingent on the 'logic of accumulation' of their data, with new contractual arrangements perpetuating Zuboff's *Big Other*. The embeddedness of Big Data in the social generates computer-mediated economic transactions that 'empty the contract of uncertainty', thereby 'eliminating the need for and possibility to develop—trust' (ibid:81). In surveillance capitalism's economy of computational agency, human authority has been mitigated by other forces, namely 'technique', what Zuboff refers to as

> "the material dimension of power," in which impersonal systems of discipline and control produce certain knowledge of human behavior independent of consent (Zuboff 1988 cited by ibid).

[196] Media theorist Jussi Parrika refers back to the writings of Kittler, as conveyed in the methods chapter regarding inscriptive media: Power is no longer circulated and reproduced solely through spatial places and institutions – such as the clinic or the prison, as Foucault analysed – or practices of language, but takes place in the switches and relays, software and hardware, protocols and circuits of which our technical media systems are made (2012:70).

Instead of contracts, governance and the rule of law, 'anticipatory conformity' is 'a new kind of sovereign power' (ibid:86) which, referring back to Chapter 3 and Chapter 5, exceeds Bentham's panopticon of surveilling, monitoring and controlling human subjects. 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 new kind of invisible hand' (ibid:82).

With anticipatory conformity people are coerced to give up choice of whether or not to relinquish their data to an 'invasive power' and instead adhere to contractual obligations, not wanting to be involved in lawsuits or other unpleasantries.

> In a world of Big Other, without avenues of escape, the agency implied in the work of anticipation is gradually submerged into a new kind of automaticity— a lived experience of pure stimulus-response (ibid).

Nowadays, *the habit of search* has been usurped by the *habit of automaticity*, where there is no escape from *Big Other* when searching with Google. Conformity is no longer a 20th century fear of not belonging, loosing oneself in a collective or a relationship—rather it hinges on a loss of choice: '[c]onformity now disappears into the mechanical order of things and bodies, not as action but as result, not cause but effect' (ibid). By purporting an 'inevitability doctrine' that Google knows YOUs better than (us)ers do themselves and that users are helpless (Zuboff 2017), its technological programming targets human agency by deleting resistance and creativity from the text of human possibility.

> False consciousness is no longer produced by the hidden facts of class and their relation to production, but rather by the hidden facts of commoditized behavior modification. If power was once identified with the ownership of the means of production, it is now identified with ownership of the means of behavioral modification (Zuboff 2015:82).

3.6 Information as Becoming: Post-history

In an interview with the *Financial Times* in 2014 Larry Page expressed his concern about people not being satisfied in society, nor having enough ambition as well as raising the following questions, 'how do we organise people? or how do we motivate people?' and even, 'how do we organise our democracies?' (Waters 2014). As an answer to Page's second question, by organising the world's information and making it accessible, Trusted Users engage with Google Search as a habit of automaticity whilst concomitantly making their data useful to surveillance capitalism. In regard to Page's third question, platforms such as Google have gone beyond merely 'navigational media' (Van Couvering 2010) to become infrastructures that facilitate daily communication technologies through 'commands, address and data' (Kittler 1999). The social implications and consequences of these 'intervening platforms' (Gillespie 2015) are increasingly more noticeable. Where once the information platforms of the address offices, then publishers and newspapers, later on radio and television were the primary organising media of the public sphere, nowadays Google's search algorithms are shaping culture by enacting 'gatekeeping', 'calculated publics', 'clickbaiting', 'algorithms of oppression' and 'autocomplete' as 'arbiters of truth'.

They are also simultaneously 'becoming culture' (Gillespie 2016), thereby legitimising an online public sphere that is manipulated by algorithms.[197] Additionally, by addressing how these 'technical affordances of digital intermediaries shape public discourse' (Plantin et al. 2017:296), the 'tensions' between human agency and these infrastructures becomes apparent.

> As these algorithms nestle into people's daily lives and mundane information practices, users shape and rearticulate the algorithms they encounter; and algorithms impinge on how people seek information, how they perceive and think about the contours of knowledge, and how they understand themselves in and through public discourse (Gillespie 2014:183).

Keeping in mind that search results are 'a reflection of the content across the web,' it is crucial to 'acknowledge that searchers' queries shape and direct the mirror' (Mulligan and Griffin 2018:573). With users' 'ubiquitous googling' as 'digital truth serum' (Chapter 9), search engines such as Google are 'a window into our own desires, which can have an impact on the values of society' (Halavais 2009 cited by Noble 2018:25).

As an answer to the first question, digital technologies 'fundamentally alter the ways we collect, circulate, and make sense of information' (Flyverbom 2017:99).[198] Nowadays the ways in which 'media organise,' (Martin 2003) such as search engines, 'determine our situation' (Beyes et al. 2019) and define the practices of human organisational life because 'Google is a medium, and media have ontological effects' (Peters 2015:329). Therefore '[t]he lesson to be learned from the 'big data revolution' discussion incited by Mayer-Schöneberger and Cukier (2013) is that 'it would instigate no less than a change in human beings' (Beyes 2017:2). With Google capturing users' IP address (and other identity markers) whilst extracting data on users ad infinitum, digital technologies reflect an organisational a priori, resulting in (us)ers being unconsciously organised by the habit of searching online (ibid).

Thus, Google Search is not merely an abstract logic that sorts and filters information but a 'symbiotic process that both informs and is informed in part by users' (Noble 2018:25) as its algorithms condition (us)ers reciprocally.

> There is a case to be made that the working logics of these algorithms not only shape user practices, but also lead users to internalize their norms and priorities (Gillespie 2014:187).

With this human algorithmic interaction, users are recursively transformed, as they enter queries about their interests, thoughts, secrets and values into the search box. Along with accepting the search results as truths, algorithms are 'informing us who we are' (Weiskopf 2018:17), which have life-changing effects, like deciding 'what we should desire or hope for, including who we should become' (Lyon 2014: 7). As users habitually consult new media [search engines] they become more like them, and '[t]hrough habits users become their machines' (Chun 2016:1).

---

[197] In the 19th century Søren Kierkegaard portrayed the phantom of the 'public' as a network of relations that flattened into a grouping within society (1846:59). Public is a creation of the 'press' and it is only the glue or 'instrument' that holds these phantoms together who 'never are and never can be united in an actual situation or organization' (ibid:60).

[198] Marshall McLuhan argued in 1964 that the medium acts as a conduit of information but that it is not innocent because a medium affects how users obtain information as it simultaneously shapes them.

Echoing Yuval Noah Harari's 'dataism' —the proposal that Google knows us better than we know ourselves— from the beginning of this chapter, the media historian George Dyson proffers that 'Google defines what we think' (Dyson 2012:308 cited by Beyes and Pias 2019:94). Having recourse to the *Ex Machina* quote in Chapter 9, search engines reflect not *what* people think but *how*—one of the consequences of the ontological and organisational properties of Google Search on (us)ers. The cognitive architectures of computers connected to vast databases parked on server farms capture the thoughts of 'how' human agents are thinking and how letters, words, sentences, utterances and numbers can be perpetually recombined (Parker and Cooper 2016).

> Viewed thus, information has no destination; it doesn't go anywhere since it has nowhere to go; it simply keeps on moving […] (ibid:243)

As shown in Chapter 8, this Cyberorganization (Parker 2000), the organisation of searching subjects or (us)ers, occurs through continuous human-computer interaction and feedback— 'information as becoming' (Parker and Cooper 2016:246) as predicted in Tiqqun's *The Cybernetic Hypothesis* (2001).

Returning recursively back to the black box of Chapter 5, Timon Beyes and Claus Pias advance what they describe as a new temporal order, or 'chronotope', which begins with cybernetics, as mapped out by Norbert Wiener. '[T]he 'advent of digital computers—along with concepts such as feedback, self-regulating systems, and prediction—had initiated a fundamental rearrangement of temporal structures' (Wiener 1961:60-94 cited by Beyes and Pias 2019:92). Wiener contrasted Cartesian closed-systems by promoting the complexity of cybernetic open-systems models with a major innovation—prediction, which 'lies at the heart of cybernetics' (ibid). Comprised of 'self-regulating systems' and 'digital computers', this 'new temporal order' is what Beyes and Pias call a 'chronotope of prediction' (2019:94). As elucidated in Chapter 9, with surveillance capitalism and its behavioural surplus, 'real time' capture of users' 'database of intentions' is fed back to them in a never-ending loop of prediction products. With the 'gift of feedback' (Ernst 2013:94), the temporalities of the 'chronotope of prediction' mark the shift of memory, how society remembers and forgets and technological artefacts (technics) change the user's perception of time and space.[199]

Humanity is not the subject of history with technology as the object; humans are intertwined within a dynamic process of temporalities and via 'technics' (Steigler 1994), such as search engines, access to the past and future is possible, as shown with 'mnemotechnics' (Peters 2015) and State Agency's 'Right to be Forgotten' legislation. Furthermore, as shown with 'function, mission and surveillance creep' (Chapter 9) data can be recombined to form various profiles, comprised of Chapter 8's Data Dust, thereby applying 'historical information to make a prediction about the future' rather than being based on observable behaviour (O'Neil 2016). Elena Esposito argues that algorithms, which ignore original contexts of data, 'produce forms of profiling that severely constrain the openness of the future of people' (2013:17). The consequence is that 'it is no longer possible to escape from ourselves; rather, we are incessantly confronted with ourselves and with our own surprising predictability' (Beyes and Pias 2019:94). Whereas once there existed the possibility of searching online without digital profiling, with

[199] The shift in the 'microtemporality' of the archive purported cybernetics as an economy in which 'information is neither matter nor energy', rather the 'new archive in this cybernetic being was the gift of feedback' (Ernst 2013:94).

surveillance capitalism ‘the problem of preserving the self-determination of individuals as possibility of reinvention’ (Solove 2007a; 2007b; 2011 cited by Esposito 2013:17) has disappeared.

> If, according to Flusser, a bond exists between cybernetic machines that interconnect by means of feedback, that behave adaptively, that process interferences independently, and that allow, by means of what today is called “big data,” the data traces of subjects to be conflated with the prediction of forms of subjectivation—then the relation between what is and what ought to be collapses and thus, with it, the modern concept of an open future. Like other thinkers before and after him, Flusser referred to this condition as “post-history” (ibid:93-94).

The next chapter elucidates some forms of resistance to Google Search’s black-boxed politics, through various degrees of user agency as ‘Emancipatory Utopias’.

As far as agency is concerned, the law holds that things and media are strictly passive (Vismann 2013:86).

Chapter 11:

Emancipatory Utopias: User Agency

1.0 Protocological Sphere

Chapter 9 showed how Google search organises (us)ers through the Black Utopia's of surveillance capitalism and Chapter 10 demonstrated its implications and consequences in a 'post-history' era. Although Google's organising of (us)ers is black-boxed, this lack of appearance through 'intransparency' and the arcane (Beyes and Pias 2019) marks a politics of (dis)appearance that could provide alternatives.[200] As explained in Chapter 8, Alexander Galloway declared that the 'generative contradiction' regarding how protocol determines control after decentralization consists of an inverse logic—on the one hand the black box that is corporate and secretive, in contrast to the 'black bloc'—reactionary technology and how people programme it (2010).

> I suggest that to live in the age of protocol requires political tactics drawn from within the protocological sphere (Galloway 2004:151).

Protocol is thus a technology of inclusion and resistance to power that cannot be organised from the outside.

This final chapter discusses the implications and consequences of reimagining search with Tor through various types of user agency. Gleaned from Chapter 8's 'Black Bloc', I interweave these effects with historic human agency, what Tantner, in discussing certain activities of the address offices in Chapter 1, deems 'Emancipatory Utopias'. These activities made information universally accessible and useful as well as providing the means for enabling user privacy or even, by way of secret registers, anonymity. Human agency also exists in the present, demonstrating how not only search raters but users have and can intervene, just as algorithms do. Instead of illuminating the black box, users can apply various tactics through programming and code or 'Agencies of Anonymity', to circumvent data capture as they (re)search. (Figure 88) Trusting others in the collective, these commons exhibit qualities of Emancipatory Utopias by building non-proprietary indices and enabling access to information through 'shadow libraries'. Additionally, I return to the address office and their clients' privacy concerns, anonymity and profiling regarding search, along with expanding upon questions of research—how to reimagine search through notions of serendipity and the art of fire.

1.1 Anonymity Machines

As shown in Chapter 1, Renaudot was of the opinion that access to information should not depend on nepotism within personal relationship networks (Tantner 2015:277-278). Besides swearing oaths, servants were not allowed to gossip nor convey private intimacies to outsiders, or to eventually pass on details to the bureau d'adresse, 'Intelligence offices' or, later, even the

[200] 'For if the modern temporal order has in fact become problematic or has even collapsed entirely, the challenge would then consist in no longer conceptualizing digital cultures with the categories of transparency, participation, and the public sphere but rather in terms of a fundamental intransparency—in terms, that is, of the arcane' (Beyes and Pias 2019:94).

police (Tantner 2014:134). Often the two parties exchanging information didn't know each other, only knowing the names of servants or a 'traditional intermediary' who could name a third place or address for follow-up questions or to carry out the request (Tantner 2015:66).[201] Therefore one of the primary functions of the offices was obtaining information, with the register facilitating anonymous job search activities. Additionally, some clients didn't wish to divulge their names because they were unsure what the offices might disclose and what might subsequently happen with their data in the public register. Because of this concern, Tantner proposes that Renaudot invented a second register to keep secrets safe and maintain privacy, what he calls an 'anonymity machine' (ibid:67).

Creating contexts of anonymity continued. Already in 1988 cryptoanarchists were endeavouring to construct a 'terra nova' in which they could be anonymous and where, as mentioned in Chapter 7 and 8, a certain degree of 'Technical Elitism' was required to navigate this realm. 'Cyberspace', as it was known, was a place of freedom-oriented communication systems, where users adopted pseudonyms to hide their identities and where encryption enabled degrees of anonymity. Cognizant that they could be monitored even though they didn't have CCTV cameras, smartphones or Google, these cypherpunks believed that anonymity is intimately tied to human potential and necessary for an open society in the electronic age (Bartlett 2014). Chapter 6 described how in the 1980s, David Chaum was already aware that metadata (who talks to whom) was just as important as content. Over the past 40 years, as technologies of identification and surveillance developed so has the means to circumvent and prevent certain types of data capture with Chaum's encryption perpetuating anonymity and improving user privacy. With the help of encryption or 'new crypto protocols' (Marx 2001), the personas and identities of Chapter 7's 'Anonymous Users' engendered various degrees of anonymity.

Nowadays information and data can be communicated using Phil Zimmerman's PGP, which simultaneously generates a public and a private key in a 'web of trust'. The user has to 'trust' that the PGP is valid with 'trust signatures', which are cryptographic 'identity certification', with information remaining encrypted and visible only to certain parties. Analogous to the 'anonymity machines' of the former address offices, PGP keys are also needed to gain access to certain 'onion' addresses' on the Dark Net when using the Tor browser. Moreover, the Tor browser peels off layers of encryption as data packets travel worldwide, facilitating anonymous communication and the sharing of sensitive information, without profiling and identity discrimination while protecting users in repressive regimes. Although there are those who have argued against online anonymity in regard to accountability (Davenport 2002), or Suler's formulation of the Online Disinhibition Effect by which users might evade responsibility (Chapter 8), as it becomes easier to impose the biopolitics of Google's surveillance capitalism on the population, the right to anonymity has become an extraordinarily important aspect of non-predictability.

> These rights to become anonymous include our rights to become who we want to be, with a sense of future, rather than to be locked into the traces and totalizing effect of a personal history that dictates, through the record, a matter of truth about who we are and potentially can become (Noble 2018:126).

[201] This is analogous to how anonymous users apply the Tor browser to meet up at certain onion addresses or with Robert Gehl's DWSN research as shown in Chapter 7.

1.2 Right to Privacy to Privacy Rights

Long before Google, as demonstrated by Anton Tantner's account of the 'first search engines' and above, there were public concerns regarding privacy and anonymity. How anonymity and privacy was conceived of in previous centuries, either by the address offices or Brandeis and Warren's formulation of a 'right to privacy' (1890) bears no resemblance to the manifold methods by which interaction with technology online affects users today. As Chapter 1 and 5 have shown, where once citizens could be located from the registers of the address offices, now the IP address identifies the online user with the hupomnemata of computers and Google Search. The daily 'adoption of information and network technologies' such as 'search tools online' that were 'previously impossible' or 'possibly unimaginable' in the past have become ubiquitous [googling], with users unwillingly divulging data to corporations (Nissenbaum 2015:11).

Formerly searching online held no repercussions yet as participation changes so too do privacy concerns and regulation (Forte et al. 2017:9). As shown in Chapter 10, users are not fully informed of what the terms of service entail when they agree to them. This entanglement between Google Search and the 'right to privacy' is epitomised by I've got nothing to hide (Chapter 8). In 2009, when Google's 'personalisation' first came to public attention, former Google CEO Eric Schmidt's response was that Google Search saves information and could share this with authorities, and 'if you have nothing to hide then why worry'? However, user concern about personalisation was already prevalent in 2012, as shown by a Pew study where

> 73% reported that they would *not be okay* with the search engines keeping track of searches and using that information to personalize future search results. Participants reported that they feel this to be an invasion of privacy (Noble 2018:54).

It becomes increasingly more difficult to establish boundaries of privacy as technology infiltrates all aspects of peoples' lives, from intimate spaces in the home to online participation and every social situation differs.[202] However users can apply an Online Persona (Chapter 8), or with Pseudonymity, as Gary Marx explains with his own moniker 'Georg Simmel', to have privacy in certain contexts and situations it is better to avoid using one's real identity (1999, Nissenbaum 2010). Yet in Chapter 6 the 'Nymwars' described how corporations advocate for an internet where all users are identifiable and that users must use 'real names' to bolster safety and legality, but this profiling also enacts behavioural control over the user (Chapter 9). In turn, certain users exhibit 'resignation' that explains the 'privacy paradox' which is a 'perfectly rational response to a situation in which human beings have very little agency' (Senior 2019). People are nowadays overwhelmed and have limited control over their data, or in the words of boyd, 'public by default, private through effort' from her article *Privacy, Publicity, and Visibility* (2010) that concerns contemporary online challenges. Users feel defeated because their data has already been collected in the 'dark net' of Big Data by governments and corporations or learning new tools is just too difficult and that they don't have the time.

[202] There are four organisational practices which are crucial for information privacy as identified by Smith: 'collection and storage of large amounts of personal data, unauthorized secondary use of personal data, errors in collected data, and inappropriate access to personal data' (Forte et al. 2017:3).

> For most people, that effort—to change how they search, how they buy stuff, how they connect with others and absorb news—is just too great (Senior 2019).

Furthermore, the amount of data a user divulges can differ from context to context as can the degrees of agency, as demonstrated by Agencies of Anonymity (Chapter 8) and Nissenbaum's 'contextual integrity' in State Agency (Chapter 10). When participating in the networked digital world, one can choose to disclose certain information, thereby raising 'issues of privacy and risk, but in some contexts privacy negotiations are trivial or routinized to the point of being unnoticeable, whereas in other contexts they require extensive attention and effort' (Forte et al. 2017:2).

Moreover, the concept of privacy often doesn't encompass the concept of 'creep' (Senior 2019). Defined as surveillance creep by Gary Marx in 1999 and then later as function or mission creep (Christl 2017), data is extrapolated from divergent contexts and then recycled to build profiles and organise (us)ers (Chapter 9). With this surveillance capitalism,

> [search] subjectivities travel a hidden path to aggregation and decontextualization, despite the fact that they are produced as intimate and immediate, tied to individual projects and contexts (Nissembaum, 2011 cited by Zuboff 2015:79).

Instead of being cast as a criminal by enacting the slogans from *The Circle*, 'Privacy is Theft' and 'Secrets are Lies', certain users do have something to hide—thoughts, health concerns, secrets and 'real' identities as citizens, not just as passive consumers. In veneration of the cypherpunks' legacy and in spite of 'perceptions of safety', with certain Tor users there is a 'belief that privacy should be a default in electronic communications' (Forte et al. 2017:8). Therefore the 'right to privacy' is not about users deceiving law enforcement, as many governments and corporations promote in the media, but pushing back on how user agency is eroded in various contexts.

> People feel as though their privacy has been violated when their agency has been undermined or when information about a particular social context has been obscured in ways that subvert people's ability to make an informed decision about what to reveal (boyd 2012:31).

In contrast to the 'habit of automaticity' represented by Google Search with users' 'anticipatory conformity' (Chapter 10), when one chooses to share something or not one exercises one's right to privacy, which produces choice (Zuboff 2015:82-83). In this manner, privacy is part of a decision-making process that places an individual 'on the spectrum between secrecy and transparency in each situation', or expressed differently, '[p]rivacy rights thus confer decision rights' (ibid).[203]

1.3 Users Intervene

If 'platforms intervene' (Gillespie 2015), why can't users?

[203] Privacy in the cyberspace era was enabled by having a certain skillset (tech-savvy). Now it is considered a privilege, either by paying extra money for subscriptions or expertise, or being able to afford social distance through wealth. But privacy is actually an inalienable and fundamental right, the basis for human dignity as stated in Article 12 of the Universal Declaration of Human Rights (1948) by the UN.

1.3.1 Rater army

> Google's business model is built around the idea that it's a neutral platform. That its magic algorithm waves its magic wand and delivers magic results without the sullying intervention of any human (Cadwalladr 2016a).

The address offices in Chapter 1 organised information manually, deciding how and where it would appear and in which register as well as what search result would be provided to its clients. As described in Chapters 5,9 and 10, online 'platforms intervene' (Gillespie 2015), with Google deciding what to display as results, based on hidden criteria. Although Google states, 'We do not use human curation to collect or arrange the results on a page', since 2004 the corporation has undertaken 'manual actions' intervening against specific websites that were 'abusing the algorithm' (Grind et al. 2019), such as by SEO gaming, paying websites to link to it (spamming) or 'clickbait' link farms. The implications from these actions include not only blacklisting websites (Tor, Chapter 7) and down-ranking competitors in regard to anti-trust practices, but also removing undesirable content from 'Right to be Forgotten' requests, preventing holocaust denial with autocomplete predictions to mitigating political bias gatekeeping or misinformation headlines as an Arbiter of Truth (Chapter 10). These daily changes affect millions of website owners worldwide, often without the owners' knowledge or explanations as to how or why these alterations occurred.

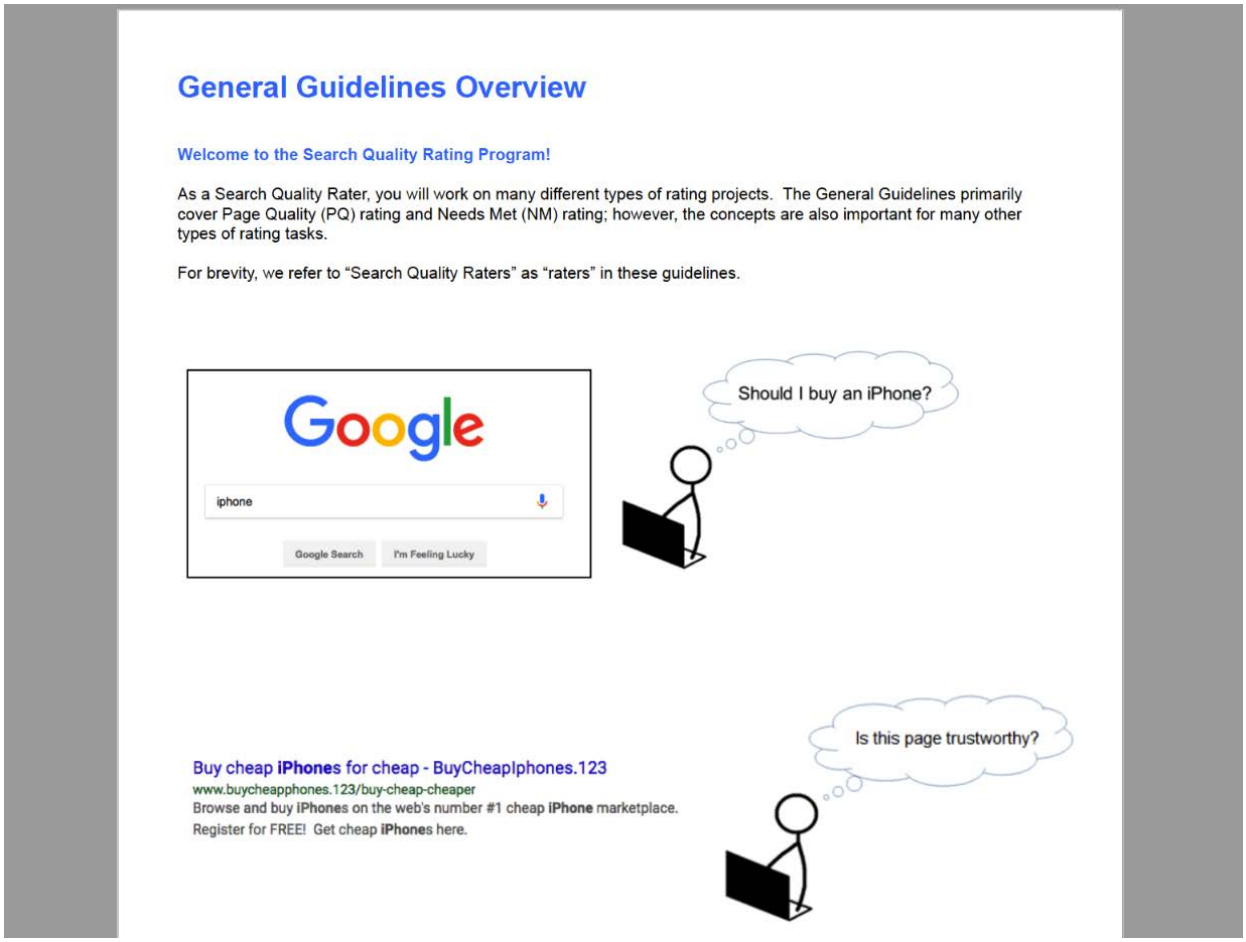


Figure 109: First page of Google's 'Search Quality Evaluator Guidelines' 2020

Behind the curtain in Mountain View there exists a 'grueling process' where a so-called 'launch committee' consisting of employees (engineers, programmers, executives) argue about what decisions to take, which is similar to defending a thesis and often contentious (ibid). Recent investigations have demonstrated how Google has 're-engineered and interfered with search results' more than it previously admitted with decisions 'made by the world's most accomplished and highest-paid engineers, whose job is to turn the dials within millions of lines of complex code' (ibid). Moreover, Google has guidelines for changing its ranking algorithms and there is also an 'army of more than 10,000 contract workers, who work from home and get paid by the hour to evaluate search results' (ibid). These contractors 'assess the quality of the algorithms' rankings' by first reading the 'Search Quality Evaluator Guidelines', then rating search results with Google providing feedback 'to convey what it considered to be the correct ranking of results' with the assessors incorporating this feedback accordingly (ibid). (Figure 109)

In spite of signing non-disclosure agreements, certain former search evaluators have divulged behind-the-scenes curation methods and interventions within Google that exert 'editorial control over what it shows users' (ibid). It is not only individual ratings—rater responses are evaluated collectively, which in turn can shift a site higher or lower in the ranking (ibid). Although one former Google executive declared that it has been 'very convenient for us to say that the algorithms make all the decisions' (ibid), increasingly Google intervenes manually as well as algorithmically concerning contentious content.

> In March 2017, Google updated the guidelines it gives contractors who evaluate search results, instructing them for the first time to give low-quality ratings to sites "created with the sole purpose of promoting hate or violence against a group of people"—something that would help adjust Google algorithms to lower those sites in search (ibid).

Therefore, Google now removes 'harassment, bullying, threats, inappropriate sexualization, or predictions that expose private or sensitive information' (ibid) and even requests that users worldwide submit free feedback by clicking on the 'Report inappropriate predictions' when searching. Yet as 'Google's culture of publicly resisting demands to change results' lessens, increasingly larger implications and consequences continue to loom regarding the US first amendment and freedom of expression (ibid) .

1.3.2 Tech-savvy

> Halavais suggests that every user of a search engine should know how the system works, how information is collected, aggregated, and accessed. To achieve this vision, the public would have to have a high degree of computer programming literacy to engage deeply in the design and output of search (Noble 2018:25).

Returning to Chapter 9 and Agre's surveillance and capture, Chun notes, 'with and against Agre—that capture systems and their logic of tracking offer possibilities for user actions that do not necessarily disempower, because they also offer a means of imagining and inhabiting the capture' (2016:61). Repurposing surveillance and capture systems through the 'temporality of networks' is 'contra Agre' but is also the context of where they 'intersect as processes of intervention' (ibid). It has now become a question of how to intervene within technological (as well as legislative) infrastructures with a 'détournement' of the very machinations that have been designed for compliance (Franklin 1989). Moreover, instead of perceiving users as being

computationally compliant, there could be a 'rethinking the division between users and programmers' (Parikka 2012:81).

Two of the Agencies of Anonymity from Chapter 8, Track Me Not and Ad Nauseum are 'interventions' that prohibit personalisation and block ads. These activist cum artistic projects apply the tactic of Obfuscation that 'has been broadly applied—in search, location tracking, private communication, anonymity—and has thus been recognized as an important element of the privacy engineer's toolbox' (Howe and Nissenbaum 2017:57). With the 'ad-blocking' industry developing over recent years increasingly more users are applying privacy-enhancing tools. In 2014 only 9,5% of users applied adblocking tools and by 2016 this number had doubled. Google is very much aware of the increase and supposedly invests millions each year to be on the 'acceptable ads list' from *Adblock Plus*. Additionally, danah boyd 's conclusions from *It's Complicated* (2014) exhibit how teenagers' use of social media exemplifies that they do care about privacy, only that they *obfuscate* through the flooding of posts and updates. Ostensibly, those who might be the most susceptible to adverse effects of technology are also the most technically savvy.

Kittler's claim 'there is no software' emphasised the 'proprietary logic that shuts off the machine from the end user through the protected mode', what is often termed 'app culture' (Zittrain 2008). Although Kittler declared, 'software obscures hardware', such as Google Search with GUIs (Graphic User Interfaces) and 'protection software' that blocks user 'access to the operating system's kernel and input/output channels' (1995:58), he also stressed that every user should be literate in at least one computer language. Additionally, Soon points out that it is code (commands, languages, protocols and datatypes), which enables the processes of actualisation and materialisation that produces agency (2018). According to Galloway, code is 'the only language that is executable':

> This is why one must invert the logic of Marx's famous mandate to "descend into the hidden abode of production." In other words, and to repeat: It is no longer a question of illuminating the black box by decoding it, but rather that of functionalizing the black box by programming it (Galloway 2010:6-7).

Here it is not just circumnavigating the hidden infrastructures that comprise the World Wide Web and internet by being able to intervene through a degree of 'tech-savvy' and with programmers' coding.

The internet is fraught with protocols that enable hierarchies yet is decentralised, as expressed by Galloway in Chapter 8. In his section on 'Hacking' in *Protocol*, Galloway articulates how resistance is being forever altered with the inclusion of the enemy in the distributed network, yet the protological enables resistance through decentralisation and code. 'Protocol gains its authority from another place, from technology itself and how people program it' (Galloway 2004:121). Hackers not only reverse engineer technologies to find exploits, they also 'hack' in order to instantiate the positive, providing the means to create action and disruption, if necessary.[204] To them, digital technology is inherently 'general-purpose stuff', i.e., it begs to be 'redesigned in ever new, creative ways and to be put to ever-unexpected uses' (Zandbergen 2014). More users are increasingly developing their Technical Elitism by using TAILS, both

[204] I use the term Hacker in the positive sense of the word. For a brief overview of the history of this complicated term, see Gabrielle Coleman's research (2015).

Agencies of Anonymity from Chapter 8, instead of having their data fed into Google's 'marvellous clouds' (Peters 2015). The Algorithmic Anonymous User, a.k.a. *Random Dark Net Shopper*, embodies this agency with the freedom to anonymously search and shop the Dark Net just as a human could with a digital pseudonym using Tor in combination with TAILS.

1.4 Trust

As with the end of *The Personalised Subject*, with Google Search Authorship, the exact workings of PageRank cum RankBrain remain a proprietary secret, thus why shouldn't the Anonymous User also become stealth and anonymous when searching by applying Tor and TAILS? Chapter 3 elucidated how the computer replaces the ledger as the contemporary hupomnema and Chapter 9 showed how Google's 'Database of Intentions' records users' most intimate thoughts and secrets. Stephens-Davidowitz's 'digital truth serum' of users' wants, desires and questions when they query the oracle Google, is echoed by Edward Snowden in the Prologue as well as DuckDuckGo founder Gabriel Weinberg:

> [T]his seems evident but it wasn't obvious to me when I started, that *search is literally the most private data on the internet because you just type in your most private thoughts into your search engine, your medical, financial, any kind of problem you have* (2019 emphasis mine).

In response to Google's 'personalisation' and 'surveillance capitalism', by choosing to use Tor instead of being grouped with others 'like me' as 'Data Dust' within Google's 'collaborative collective', I reimagined search and 'intervened' by joining an Emancipatory Utopia. As elucidated in the effects described in Chapter 8 on 'collaborative collectives', rather than embodying the Trusted User of Google's personalisation, I am assigned as a Tor 'anonymous user'. I have agency and I am part of the social fabric of the Tor collective, which I trust. With the decentralised mesh of proxy servers, 'trust depends on others trusting' (Peci and Introna 2019:43), instilling the belief that Tor relay and exit node operators are well intentioned and that researchers and governments are not carrying out exploits or relaying IP addresses to authorities for correlation with ISPs.

Moreover, a question raised by Peci and Introna in regard to Bitcoin as a decentralised technology could also be asked of Tor: 'how is trust possible in a technological reality, governed by algorithmic code, that might be in some instances unfamiliar to us?' (2019:43). Not only does a belief exist that code is neutral (Chapter 10), there is also the 'wisdom of the crowd', peer to peer review systems and what Nicholas Luhmann (1979) referred to as 'trust in trust' (ibid:49). 'Trust is placed in the trust of other people, rooted in a common political project' (ibid). As explicated in Chapters 5 and 9, the internet links people to each other with TCP/IP protocols as they engage in networked projects that are social because of the prevalence of mass-communication. The reason that this type of voluntary engagement exists is that 'they are held together by common protocols of communication and based on trust among their participants' (Stalder 2010).

As more engaged users share similar privacy concerns, a critical mass is constantly evolving through interventions into solutions: Tor, data regulation, adblocking enabling software and individual tunnels of privacy or VPNs (virtual private network). By changing the user's IP address VPNs encrypt all their internet traffic from a user's computer, whereas Tor only encrypts requests when using the browser. There is an uptake in usage because VPNs are

ostensibly faster, however certain VPN providers can see user activity, often collecting user data and some VPN software has 'leaks'.[205] However, VPNs are not a p2p collective whereas Tor is a anonymity network exemplifying a culture of solidarity that is 'about participation and not about representation' (Stalder 2013:54). One result of the 'Snowden effect' is that 'between 2m to 2.5m people use Tor worldwide every day' (Zaunseder 2018). Moreover, as shown in Chapter 7, the more people who use Tor the harder it is to crack its anonymity. This is also a factor with K-anonymity (Chapter 8), which alters released database information based on scaling, with a greater number of candidates facilitating more anonymous data. Most crucially, rather than only focusing on individual privacy concerns with VPNs, a range of user agencies are addressing methods of belonging that shift from the individual to the collective.

> Contemporary privacy studies are trying to reconfigure our perception of privacy away from a grounding in individual, liberal rights and reformulate the value of privacy in terms of social and political benefit (Munkholm 2019).

1.5 Collective Commons

Although the Google Search platform functions as an infrastructure (Chapter 10), Tor is an infrastructure for individual *and* collective agency that 'relies on voluntary co-operation to enhance the use value of a shared resource' (Stalder 2013:28). In his book *Digital Solidarity* (2013), Felix Stalder describes these forms of 'communing' that reflect the shift from concerns of the individual to that of the network. 'Commons' here means 'long-term social and material processes', where the 'relevant choices to be made are collective not individual ones' (ibid:33). Production is not just commercial, but social because people produce things together and have a 'use-value' for them, emphasising that 'no commons exist in a social void' (ibid). As described in Chapters 6 and 7, Tor is a p2p commons that offers anonymity to users searching online who desire to be anonymous, from those protecting their lives to those searching the Dark Net to those preventing data collection. Yet Tor often incites critique because it cannot be 'hacked' by governments. However, this hostility towards the commons is not new.

Larger systems, which are usually oriented towards market exchanges or state control, are more often than not hostile towards the practices of the commons (Tantner 2015:78-79). According to his critics, Renaudot (Chapter 1) wanted to collect all the charters, overturn all order in the kingdom, encourage freedom and disseminate knowledge to everyone and, along these lines, the bureau d'adresse thereby meant shared property, or even worse, the 'heresy of the Anabaptists'(ibid). Renaudot's 'Emancipatory Utopia' organised information so that clients could offer and find services, even anonymously, yet this extended even further by making knowledge accessible and useful to the larger public. In 1633 Renaudot's bureau d'adresse also functioned as a kind of 'scientific academy', with lectures advertised at the office and in ad journals, which were well attended by the public (ibid). For Renaudot it was primarily about the exchange of knowledge and diverse courses were held from various disciplines—arts or sciences or genealogical information. The lectures were also summarised and published in book form, anonymously, echoing the pamphleteers and anonymous uploads on the Dark Net described in Chapter 7.

[205] Njalla is a trustworthy VPN, formerly called IPREDator that also is a privacy hosting service that even accepts cash in the post to Sweden anonymously for subscription fees. https://njal.la/blog/ipredator/

Recalling Renaudot's belief in access to information without nepotism and also for the poor or needy, the contemporary condition of '(re)searching' online, for people without personal networks, academic access or the funds to purchase books, is often fulfilled by those making information 'free'. Analogous to the book lending libraries, reading groups and book clubs of Anton Martin's 'question office', p2p 'networked' distribution models enable the publishing of scholarly material as 'shadow libraries' (Thylstrup 2019). These practices of 'file-sharing' facilitate the dissemination of knowledge where one does not have to pay.[206] Rather one must scan, upload and contribute as part of the exchange system or know the address (hyperlink) in order to find them.[207] Yet are these shadow libraries any 'more illegal than Google's controversial book scanning project?'(ibid:79). As shown by my results in Chapter 5, Google not only now scans sections of books without permission but promotes and prioritises its own products and services (books.google.com) as results when 'searching' with specific keywords or carrying out 'research'.

In her book *The Politics of Mass Digitalization,* Nanna Bonde Thylstrup argues that these shadow libraries 'belong to the same infrapolitical ecosystem as Google Books and Europeana' and although they might be considered competition, they are also 'strengthening them' (ibid:81). '[V]isitors' interactions with shadow libraries are also marked by parasitical relations with Google, which often mediates literature searches, thus entangling Google and shadow libraries in a parasitical relationship where one feeds off the others and vice versa' (ibid). Moreover, as shown above with the section on Trust, co-dependent relationships exist and enable shadow libraries because they are key infrastructural elements, 'what James C. Scott calls the "infrapolitics of subordinate groups" by offering 'unhindered access' (ibid:83). Or, working within the shadows as methods of opacity, Open Access publishing of 'research' as 'liquid books' by Gary Hall and Claire Birchall becomes a type of commoning. Rather than the networking and citation cartel of Garfield's SCI (Chapter 3), academic publishing is transformed by re-using other's ideas in a shared and collective authorship, sometimes pseudoanonymously, as demonstrated by Tiqqun's *The Cybernetic Hypothesis*.

In the 1950's and 1960's Hannah Arendt addressed the complexity of the commons versus individual agency in the predigital era when she expressed concern about being 'overly determined by external ascriptions and heteronomous forms of subjectivication' (Matzner 2018). Now in an era of digitalisation, the effects of subjectivication (Chapter 8) are made possible by the Big [haunted] Data of surveillance capitalism (Chapter 9). Yet, there are increasingly more alternatives and moments of 'Emancipatory Utopias' that provide the means to circumvent this type of (subjective) behaviour control.

> The desire to capture and to draw connections, maps, and models can open a future that defies what is captured, but only if we dwell in the disconnect between map and action, model and future—only if we *occupy the collective chimera* we are offered and become characters, not marionettes, in the ongoing drama inadequately called Big Data (Chun 2016:62 emphasis mine).

[206] A former Yugoslavian 'Whole Earth Catalogue' the Galaksija enabled file transfer well before the advent of the World Wide Web, a pre-internet pirating protocol. Available here: https://www.jacobinmag.com/2020/08/computer-yugoslavia-galaksija-voja-antonic?
[207] sci-hub.org, memoryoftheworld.org, aaaaarg.org and monoskop.org

Whether one is searching for research within the p2p file-sharing cooperatives of 'shadow libraries' or joining the p2p Tor collective in order to search anonymously and simultaneously increase the anonymity of the network, individual users engage with the commons. The 'antidote' then is perhaps 'grounding the value of privacy in plurality rather than individualism' and by this, it becomes not the 'autonomous individual' but the 'socially situated political subject' (Matzner 2018) who has agency.

1.6 Redistribution of the sensible

As shown throughout this thesis, users freely give away not only their labour but also personal data when adopting Google's personalised services. This is what Joseph Turow refers to as the 'hidden curriculum' of training 'personalised subjects'—giving up data for customised results because the 'whole idea of data collection is to make it sound natural' (2017). Shoshana Zuboff argues that with surveillance capitalism searching subjects cannot just demand privacy, or self-regulate as that doesn't work because 'operations were secret and opaque by design, legible only to a narrow priesthood' (2017). Free from detection, free from sanction, free from regulation and 'free from any meaningful mechanisms of consent' such as those 'associated with democratic oversight expressed in law and regulation', surveillance capitalism is intended to 'produce user ignorance' (ibid). In addition, the ownership question of data is constantly being debated—companies state that they provide 'free services' in exchange for user data, and users respond that their data needs to be something they regulate, or that it is an extension of themselves, or their property, returning to a question posed in the Prologue.[208]

Amplifying self-determination, the European DECODE project intends to utilise data as a common good, where users decide rules, regulate the environment, have geopolitical independence and contribute to the well-being of the places they live in (Bria 2017).[209] However, Zuboff contends that the capacity for user self-determination and personal autonomy is accumulated in this regime of private administration and capital (2018). With this in mind, Claire Birchall defines the term *shareveillance*, a portmanteau of 'share' and 'veillance', as a subjectivity that is 'antipoliticized' through neoliberalism's consumer society of 'free choice', with the sharing of data in exchange for services foreclosing politics (ibid). Unlike the Neoliberal Subject (Chapter 8), Birchall emphasises that this subjectivity has a choice whether to share—or not—and her argument is to push back on the subjectifying elements that enable a 'shareveillant subject' in the first place. Critique does not necessarily lie with strengthening privacy legislation, which has become a lesser political concept—as it does not 'redistribute the sensible' (ibid).[210] For the theorist Jacques Rancière, the 'distribution of the sensible' is the politics of aesthetics, perceiving and altering the visible/invisible as well as the political rearrangement of inclusion/exclusion, an interrelationship of who can be seen, heard and take part in this new apparatus, dispositive or regime. Analogous to the Homoeconomicus (Chapter 8), Birchall inserts the digital into Rancière's regime, stating that the 'one who simultaneously works with data and on whom the data works' is produced within an assemblage:

[208] However, companies now retort that they don't 'sell' data about users, rather they admit to selling 'their opinion' about user data, citing that this falls under the 1st Amendment of the US constitution, concerning freedom of speech.
[209] https://decodeproject.eu/
[210] 'The system of self-evident facts of sense perception that simultaneously discloses the existence of something in common and the delimitations that define the respective parts and positions within it' (Ranciere 2004:12).

> "Shareveillance" is intended to capture the condition of consuming shared data and producing data to be shared in ways that shape a subject who is at once surveillant and surveilled (2018:18).

Moreover, Birchall challenges the conditions of shareveillance that 'produce subjects from whom transparency is demanded' (ibid:80). To return to Chapter 8, although certain 'data subjects' are made visible, they are Interpellated Subjects with demands such as participation placed upon them, as with one of the fictitious Circle's mottos, 'sharing is caring' (Eggers 2013). Following the tension between secrecy and transparency already articulated by Beyes and Pias (2019) in the Prologue and Chapter 10, Birchall argues 'against the insidious logic of the value of transparency' (Southerton 2018) that is attached to individual concerns. Birchall seems to reject the 'agential subject', which validates an individual's claim to a 'right to privacy' and instead engenders politics collectively.

Citing Massimo de Angelis and Silvia Federici, Birchall emphasises that the idea of commons comes to the fore within an assemblage as an alternative practice of sharing, or 'non-shareveillant' sharing', which is not rooted in transparency and quantification but a 'quality of relations' (2017:50-51). Alternatives exit, as with Édouard Glissant's 'right to opacity'. In *Poetics of Relation* Glissant speaks of opacity as an 'unknowability', and although it is an alterity which is unquantifiable, it needs to be defended at all costs so that democracy can take hold. Opacity also reveals the limitations of visibility or, reminiscent of Brunton and Nissenbaum's Obfuscation (Chapter 7 and 8), counteracting the barbarism of visibility is an ontological proposition yet is also an ethical and aesthetic one (Blas 2016). What all of the above theorists and artists 'share' with the Tor Browser are modes of collective resistance (redistribution of the sensible, right to opacity, p2p obfuscation tactics), which embodies an individual's political relationality— as a 'redistribution of the sensible'.

1.7 World Brain

"When you have the maximum amount of viewers, I want you to read this statement." He handed Mae a piece of paper, on which he'd written, in crude all capitals, a list of assertions under the headline "The Rights of Humans in a Digital Age." Mae scanned it, catching passages: "We must all have the right to anonymity." "Not every human activity can be measured." "The ceaseless pursuit of data to quantify the value of any endeavor is catastrophic to true understanding." "The barrier between public and private must remain unbreachable." At the end she found one line, written in red ink: "We must all have the right to disappear."
"So you want me to read all this to the watchers?"
"Yes," Kalden said, his eyes wild"(Eggers 2013:485).

This excerpt from the end of Dave Egger's novel elucidates Kalden's (one of the three founding Wise Men) desperate attempt to keep *The Circle* from 'closing'. What fiction like Egger's gives the reader is a prediction of the future, in this case not an Emancipatory Utopia but a dystopia where everyone is required to be 100% transparent. As witnessed with the character Stewart's server of data in Chapter 9, the invisible hand of *Big Other* perpetually captures everyone's data on planet Earth and there is no escape—no privacy and especially no anonymity. In the last pages of this apocalyptic tale, the protagonist Mae contemplates collecting her friend Annie's 'thoughts she was thinking', as she lay in a coma.

'Why shouldn't she know them?' The world deserved nothing less and would not wait.' (Eggers 2013:491).

Making a comparison between territorial marauders of the past and present-day surveillance capitalists, Shoshana Zuboff elucidates how the Taino people on St. Croix, USVI were exploited and eventually eradicated by colonial empires (Christopher Columbus) searching for indigenous gold and silver, turning the people and natural resources into commodities (2017). These colonisers marked territories on parchment, as vague 'contracts' of sale or through the lines and cross-hatchings of Dutch cartography that rendered 'native' land part of the empire. Cartographies reflect the 'spatial domain' of intellectual exploration and the X and Y vectors of today usurp the longitudinal and latitudinal degrees of the colonial era. This action of mapping 'undiscovered' areas of the world and the quest for knowledge is still performed in the 21st century, translated into computer-generated models, network analysis infographics or data visualisations. Colonisation continues with the corporate extraction and transformation of natural resources, however the commodity of now is user data, which feeds the economic motor of surveillance capitalism.

> The largest map of the world, the largest databases, drawing the most attention from the greatest number of web users. Like a vacuum cleaner, the winner will suck up the minds of the world. And what better way to suck up the human mind, than by taking inspiration from the human being itself? (Degoutin and Wagon 2015).

To return to Chapter 9 and Polanyi's fictional commodities, it is now the 'logic of accumulation' of user data (Zuboff 2015) and its extraction process which, as with the oil industry, involves rounding up all the data streams that flow. The goal is to 'extract as much personal data from as many people in the world at the lowest possible price and to resell that data to as many companies as possible at the highest possible price' (Taplin 2017). Unlike oil, there is no need to worry about reserves as users constantly produce data, however 'there is plenty of dirt on data' (Peters 2015:333).[211] Sociologist William Davies provides an analogy between oil and privacy, which he considers a 'concealed natural resource that is progressively plundered for private profit, with increasingly harmful consequences for society at large' (2018). Just as indigenous peoples, environmentalists and citizens worldwide demand that the fossil fuel industry 'leave it in the ground,' the ultimate demand to be levelled at Silicon Valley should be that data be left 'in our heads' (Davies 2018).

During the past 20 years users willingly added websites to build Google's proprietary index, along with clicking on the hyperlinks of search results. Besides these interactions, users' queries have become the Memex of search, as explained in Chapter 9 with Google's 'database of intentions'.

> Clickstreams are the seeds that will grow into our culture's own memex—a new ecology of potential knowledge—and search will be the spade that turns the internet's soil (Batelle 2006:258).

To return to Chapter 1, the address offices carried out the administrative activities that comprised the protocols of search—organising information by gathering citizens' data, such as

[211] Those mining and analysing data thereby produce more data about the data already produced and the people who produced it.

their house address, then building an index of information and queries in the form of the register and subsequently delivering search results to clients. In Chapter 2, the *Mundaneum* (1910) was conceived as a 'mechanical, collective brain' yet Otlet's endeavour to organise the world's information by structuring knowledge through Western 'universal' categories was a colonial project (Constant 2020).[212] The godfather of biological warfare Vannevar Bush envisioned his Memex machine to connect 'associated linkages', or streams modelled on the human brain, which created roads into areas of knowledge and foreshadowed the hyperlink (1945). H.G. Wells conceived of the non-commercial, international and publicly accessible *World Brain: The Idea of a Permanent World Encyclopaedia* as an 'all-human cerebrum' or world information centre made possible by telecommunications and microfilm, yet it was only to be in English (1937).

Eugene Garfield cites Wells in the first sentence of his introduction to the Scientific Citation Index (SCI), which for him was a 'harbinger of things to come—a forerunner of the World Brain' (1964).[213] The video *World Brain* demonstrates how user data is captured globally from underwater cables, circulating it to data centres globally as it characterises corporations as

> giants with mysterious financial structures who collect data, hoping to gather the largest network. Data are their lifeblood. And the winner will be whichever has the largest library of cultural content (Degoutin and Wagon 2015).[214] (Figure 110)

Figure 110: still from *World Brain* (Degoutin and Wagon 2015)

[212] Constant has also intervened in Otlet's Wikipedia page, a world encyclopaedia itself, by including racist statements from Otlet during the course of his life (1888 and 1935). https://en.wikipedia.org/w/index.php?title=Paul_Otlet&diff=950711784&oldid=949099522&diffmode=source. Constant Newsletter May 1, 2020

[213] The founders of Google were most likely aware of these endeavours to 'organise the world's information and make it universally accessible and useful', only, except for Garfield, they didn't cite them.

[214] Already in 1964, students marching for free speech at Berkeley were acutely aware of dehumanisation, 'as if they were bits of abstract data,' referencing Hollerith punch cards that tabulated the 1890 US census (Turner 2006:2) and Achenwall's statistics. Instead of protest signs and fearful of how they would be treated by America's political leaders, '[o]ne student even pinned a sign to his chest that parroted the cards' user instructions: "I am a UC student. Please do not fold, bend, spindle or mutilate me"'(ibid).

Above all, it is the machine-learning RankBrain that perhaps has become the realisation of the 'World Brain', which is now being implemented in all search queries unfamiliar to Google, adding these to its ever-increasing index. Yet organising the world's information and making it universally accessible and useful to itself through invisible criteria that organises (us)ers and answers to one corporation, is in itself a contemporary colonial project.

> Page once dreamt of downloading the entire Internet as nothing but its links, and saw the best search engine as "a reference librarian with complete mastery of the entire corpus of human knowledge" (Peters 2015:328).

However, although Google's 'search method has provided a stunning answer to the enduring problem of how to find your way around the library or universe' (Peters 2015:325), its index remains incomplete. As elucidated in Chapter 6, approximately only 4% -11% of the internet has been indexed by Google as it omits all content that isn't publicly accessible on the Deep Web: password protected intranets, content not linked to other web pages and websites that require subscriptions (which are only partially indexed) and the Dark Net–Tor's Onion Services. Moreover, there are also all the other forms of knowledge, especially tacit or oral, which are not digitalised. Rather than advancing the 'neoliberal capitalist project' of keeping search mechanisms opaque, the public, workers, librarians and institutions can together 'reconceptualize the design of the indexes of the web' in order to not only 'radically shift our ability to contextualize information' (Noble 2018:133), but to offer alternative search results.

1.8 The art of searching

'Everywhere *surfing* has already replaced the older *sports*' (Deleuze 1992).

As stated in the Introduction, Introna and Nissenbaum (2000) put forth 'why the politics of search engines matters', which is what inspired me to embark on this PhD. Following Foucault, I did not wish to be governed at a certain cost, in this case as *The Personalised Subject*. These prescriptive technologies that 'command' are designed for compliance (Franklin 1989), or in the words of media theorist Jussi Parikka, 'we do not program anymore, but are programmed as merely users/consumers of media' (2012:81). Moreover, the hidden consequences of these control processes are 'exceedingly effective and efficient but they come with an enormous social mortgage': being conditioned to 'accept orthodoxy as normal and only one way of doing it' (Franklin 1989). Thus by using media such as Google Search and its 'personalisation', I participate, thereby acknowledging I am organised, yet I question whether search engines could to be designed otherwise. According to Bijker and Law,

> the processes that shape our technologies go right to the heart of the way in which we live and organize our societies. (. . .) Understanding them would allow us to see that our technologies do not necessarily have to be the way they actually are (1992:4 cited by Mager 2012:771).

In Chapter 1, Tantner elucidated how the pre-history of search engines, or address offices, organised information as well as society and that this was a 'creation of something certain out of randomness', citing Leibniz (2015:153). The office then was not only an 'Emancipatory Utopia' for finding information but a place where searches resulted in fortuitous encounters, perhaps unanticipated and therefore serendipitous. Although the word 'serendipity' didn't arrive in

English until 1754 when it was coined by Horace Walpole, the phenomenon it defined, a chance encounter when searching, is an integral part of the process of exploration.[215] Coming across things unexpectedly later became a way to articulate how search engines could possibly work, clicking on hyperlink after hyperlink, or 'surfing'.

> Serendipity was inherent to the initial metaphor for traversing the web: surfing. The metaphor suggested that while you may be moving through the information, there was room to turn around, or take detours, and that the topography of the information encouraged these actions (Halavais 2009:53).

As demonstrated in Chapter 7, I attempted to 'reimagine' search and achieve Anonymity by applying the Tor browser and TAILS. Besides its concealment of the user's IP address, with privacy searching the p2p anonymity network (Tor) also offered unpredictable or 'serendipitous surfing', facilitating a 'right to disappear' or to obtain Unreachability (Chapter 8), whilst navigating a continuous online attention economy.

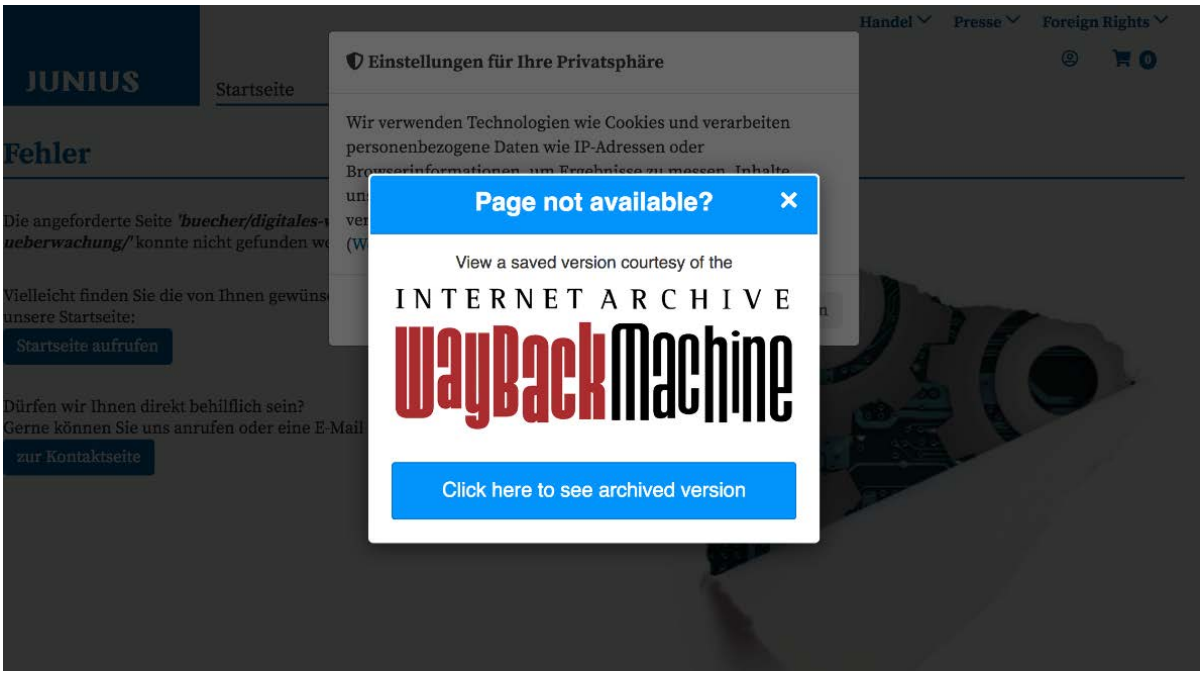


Figure 111: WayBackMachine's index of the web

Harking back to Jean Amour Polly and her 'Surfing the Internet' guidebook (1992), Brewster Kahle also imagined early on that users could 'surf' the web when working on WAIS (Wide Are Information Service) in 1990 and when starting Alexa (1996), a web traffic analysis website, with Bruce Gilliat. Kahle and Gilliat went on to found the Internet Archive in 1996, which captures not only the present web but records it as snapshots of time and which has expanded to encompass web pages, books and texts, audio recordings, videos, images and software programmes. The Internet Archive has already accrued a massive index with its *WayBackMachine* saving more than 327 billion web pages, which are in the public domain and can be searched and accessed by anyone with an internet connection. (Figure 111) I also benefited from a screenshot from the *WayBackMachine* as 'mnemotechnics'—records of human memory—of what was published on a website in regard to 'researchers' braggadocio concerning

[215] Such as the book on the library shelf next to the book one was searching for.

their ability to collect Tor IP addresses, before it was deleted (Chapter 7). Most recently, in a contemporary 'post-truth' society, the *WayBackMachine* is being adopted by the general public as well as researchers and investigative journalists to document the past in ever-increasing disputes of (dis)information.[216]

According to Alexa.com, since 2013 Google.com is the most visited website in the world.[217] In spite of this constant online attention economy and one of its consequences—scopophilia, the agency of deletion also exists, such as with 'Right to Be Forgotten' legislation (Chapter 10) and various user interventions. Human programming also enables the user to 'disappear' by adding 'robots.txt' to websites so that they are not taken up into search results, or a 'noindex' meta tag in the page's HTML code, or 'donotlink.it' that prevents higher ranking in search engine results, along with referrer removal that shows where the link comes from.[218] However, the art of remembering, forgetting and eliminating is not only implicit to search but research. Tagging, giving order through organisation, is an inherent part of 'mnemotechnics'—yet also a creative act steeped in editing, filtering and discarding (Peters 2015:318). For some it is an act of cultural barbarism, throwing out paper documents made from trees or deleting digital files, whilst for others it becomes a means to 'slash and burn' words, paragraphs and chapters.

> The art of searching, like that of fire, consists largely in elimination. Knowledge is not the gathering but the throwing away of information. Learning involves discarding as much as gaining— students must abandon fear, habit, and attitude (Peters 2015:318).

If, as Peters states, with the art of searching, knowledge is about eliminating information and, like fire, a tactic of curation, then instead of PageRank's 'random surfer' with an .84 probability predicting where I would go surfing, the 'd' for 'damping factor' has arisen from the ashes (Chapter 3).[219] I decided to go off the Page and changed my habit—by choosing another search engine.

---

[216] Since March 2020 the Internet Archive has been involved in a lawsuit by various publishers for 'mass copyright infringement' because during the Covid-19 lockdown it made millions of books accessible to the public with its National Emergency Library that 'suspended limits on the number of patrons who could borrow its digital books simultaneously' (Bustillos 2020). The lawsuit brings into question what is known as 'Controlled Digital Lending' (CDL)—'lending digitized books within parameters of existing copyright laws' (ibid). "What libraries do, is they buy, preserve, and lend. What this lawsuit is about—they're saying the libraries *cannot buy*, they *cannot preserve*, and they *cannot lend*" (Kahle cited by ibid). It closed this library on June 16 but the publishers wish to close the Open Library permanently.

[217] Alexa is now owned by Amazon and continues to provide web crawls for the Internet Archive.

[218] https://donotlink.it/

[219] The d damping factor is the probability that they eventually get bored and start on another page somewhere… with the 'probability that the random surfer visits a page is its PageRank' (Brin and Page 1998:109). Or see the patent description: https://www.google.com/patents/US6285999

Conclusion

The objective of this thesis was to find out *how Google Search organises (us)ers?* and each chapter responded to this central question through a series of sub-research questions.

I began by investigating *how search happened in the past,* drawing on Tantner's book *The first search engines* (2015), which asserts that during the 17th-19th centuries, the 'bureau d'adresse', 'Intelligence Office' and 'Question Offices' within European cities were pre-modern search engines (Chapter 1). 'Human crawlers'—servants—gathered information for these 'address offices' and then created organisational and storage technologies of information in 'Protokollbücher' or registers, including the residential addresses of clients. With the register filtering queries and results, these 'hupomnemata' or memory aids (Chapter 4) served as an interface between clients and those in search of items, services or employment. Subsequently the media shifted to printed matter, posters in public squares and journals with paid advertisements, which could be distributed throughout cities and the countryside. Certain offices engaged not only in information gathering but also policing and surveillance activities that profiled clients, what Tantner deems 'Black Utopias', which are the antecedents of surveillance capitalism and databased profiles (Chapter 9). Yet 'anonymity machines', or secret registers were created along with visions and realisations of 'Emancipatory Utopias' that offered free information to the public and operated as cultural places of learning (Chapter 11). Eventually these 'whimsical' offices became publishers and in the 19th century governmental institutions and veritable newspapers replaced their functions.

As with the transition of bound office registers to distributed newspaper listings, people's house addresses became analogous to the call numbers of books which allowed them to be found by location in libraries, as well as their metadata: index catalogues. 'Index cards' are often perceived as the pre-history of computers (Tantner 2015) and Otlet and La Fontaine's *Mundaneum* (1913), with its Western system of classification, comprised thousands of them in order to answer queries from the public as a 'collective brain'. Other 20th century human endeavours imagined *organising the world's information* (Chapter 2), which is itself a colonial project (Chapter 11). Wells' *World Encyclopaedia* (1936) conceived of gathering together global knowledge, Bush's *Memex* (1945) envisioned 'associative linkages' of stored information and Brand's *Whole Earth Catalogue* (1968) foreshadowed individualism with its personalised tools of 'counterculture'. The World Wide Web and hyperlinking ushered in the online era where academics developed early search engine technologies. By the end of the 1990s commercial portals from media conglomerates dominated the sector and in 1998 two academics as Stanford launched Google.

The first word of this thesis, *Re:search* fuses two concepts: Garfield's Scientific Citation Index (SCI) (1964) that measured scientific publications through linkages with an 'impact factor' along with their authority (research), which in turn served as an inspiration for PageRank (search). *A media archaeology of citation* (Chapter 3) demonstrates *how Google search works*, first by explaining what is an algorithm and then what is in an algorithm—how PageRank as a computational artefact is both a link analysis and an IR (Information Retrieval) system (Rieder 2013). This hermeneutical treatise on PageRank explicates how 'research' and 'search' merge, based on the SCI (bibliometry), eigenvector centrality (sociometry) and graph theory, along with Brin and Page's figure of the 'random surfer' (1998). During the first years of this millennium,

through ‘Googlarchy’ (Hindman et al. 2003) and later on with the ‘algorithmization of the hyperlink’ (Helmond 2013), PageRank transitioned from a navigational tool to the currency of the web by creating an attention economy of relevance based on the reputation of ‘backlinks’. This ‘citation notation’ became the organisational paradigm for the way seekers found information online and how that information is prioritised and organised for them. With the outsourcing of labour—clicking on relative hyperlinks to determine value, Brin and Page’s ‘trusted users’ created an attention economy of traffic and built Google’s ‘database of intentions’ (Chapter 9). Over the years ‘trusted users’ delivered content, generating ‘network surplus value’ (Pasquinelli 2009) for Google, as well as exchanging their data for free search results. With the collection of an IP address, location information and search queries, Google recirculated this data to provide ‘personalised’ search results to individual users.

As a researcher living, working and searching as a ‘personalised subject’, *Reflection on methods* (Chapter 4) begins with my ‘critical ethnography of the self’ and being ‘commanded’ by Google’s prescriptive technologies. Using my office as a site of data collection, I designed and carried out an ‘experiment in living’ (Marres 2012). I compared inscriptive modes of ‘address’ by searching on two computers: a ‘personalised’ Apple with Google Search that captured my IP address and a ‘hacker approved’ PC with the anonymity browser Tor (The Onion Router), which hides my IP address. Whilst conducting ‘interviews’ with algorithms—invisible interlocutors—I collected data on myself and then gave form to the data with my method, ‘data visualisation as transcription’, as a type of ‘critical cartography’. *Re:search - Terms of Art* is a series of data visualisations that compares the search results (URLs) of Google (personalised) versus Tor (anonymised) by searching with the same chosen ‘keywords’. This ‘methods assemblage’ (Law 2004) is positioned in relation to the fields of organisation studies, media theory and artistic research.

The Personalised Subject (Chapter 5) continues to explore *how Google Search works* by building upon Chun’s ‘habits’ of new media (2016)—search engines. Organised through three lenses, I analyse my empirical results in relation to current research: how ranking is determined by ads (Advertisement) and personalisation (Authenticity), along with my unique results (Authorship), assigning each a specific modality of capitalism (platform, semantic, algorithmic). Google’s ‘platform capitalism’ (Srnicek 2016) promotes its own services through ranking as it ‘intervenes’ (Gillespie 2015), serving up ads that influence the user’s experience and detouring their path to unbiased and non-commercial information. The IP address not only enables user identification but also locative data collection and with Google’s algorithmic interpretation I received postal services results when I searched with certain keywords or ‘Terms of Art’ that began with the word ‘post’ (postdigital, postmedia, postinternet, posthumanism). With ‘semantic capitalism’ (Feuz et al. 2011), the categorisation of users into certain groups is not based on keywords but rather by ‘collaborative filtering’, recognising that what Google advertises (personalisation) is limited and based on the similarity of other users’ interests. Google Search algorithms are obscure and proprietary secrets regulated by ‘algorithmic capitalism’ (Bilic 2017) with its black-boxed politics of (in)visibility management (Flyverbom et al. 2016). The third most important ‘signal’ ‘RankBrain’, a machine-learning artificial intelligence system that uses neural networks, is now fielding ‘unique’ queries unfamiliar to Google’s algorithm.

As an alternative to Google Search I decided to *reimagine search* through anonymity technologies, such as encryption and the Tor browser in *Worlds of Anonymity* (Chapter 6). In the late 20th century, in ‘cyberspace’, hackers and programmers—‘cypherpunks’, were instilled with a belief that anonymity in online communication was a human right. Chaum’s public key

cryptography technologies (1981) lead to developments such as Zimmerman's PGP (pretty good privacy) (1991) but also to 'mixed networks', or routing protocols that intervene between request and destination such as Tor. With the 'privacy turn' at the beginning of the last decade, the 'Nymwars' debated the politics of real names in regard to corporate digital identities and the Snowden revelations (2013) perked user privacy demands and Tor usage. The Deep Web, or 'non-indexed' worlds are off the radar for most search engines. Yet the Tor p2p browser applies layers of encryption, thereby 'hiding' the IP address of those users in areas of conflict or censorship where anonymity protects their lives or for those wishing to explore 'Onionland'.

In the spirit of an 'abolitionist resistance of refusal' (Dick 2020) to Google's personalisation, *The Anonymous User* (Chapter 7) attempts to *reimagine search* by employing Tor in combination with The Amnesic Incognito Live System (TAILS) on the 'Dark Net'. A certain degree of 'tech-savvy' or 'technical elitism' and specific search engines (Grams, Torch) are required to reach these onion addresses, aka Tor Onion Services—a melange of sites ranging from anarchist forums and social networks to 'platform criminalism' (Dittus et al. 2017). Bringing together ethnographic studies (Gehl 2014; Bartlett 2014) and Gary Marx's 'rationales of anonymity' (1999), I identified a range of 'anonymous users' who upload texts and images, activists and whistleblowers releasing important information, the activities of illicit platform vendors or just 'Torists', like myself. In the technological era, new media introduces new categories as well as new actors, transmission principles and legislation**,** along with redefining users' 'expectation of privacy' when using Tor, as verified by one controversial case of a 'Deanonymised user'.

Whereas the address offices collected residential addresses, nowadays Google Search identifies users online by collecting personal data—including an IP address, yet when using the Tor browser, IP addresses remain obscured. *Black Box versus Black Bloc* (Chapter 8) employs Alexander Galloway's eponymous essay to synthesize *The Personalised Subject* and *The Anonymous User* by adopting the internet protocol (IP) address as an organisational hinge to show *the effects of search on (us)ers.* Departing from the 'data subject'—the EU legal term for a human being whose personal data is collected, I analysed how 'subjectivities of search' and 'agencies of anonymity' are organised into 'collaborative collectives' according to degrees of human-algorithmic interaction. The effects of the *Black Box* propose that as Google's personalisation increases so does the amount of computational agency, at the same time the degree of anonymity decreases. The effects of the *Black Bloc* propose that as the amount of anonymity increases with Tor so does human agency, at the same time the degree of personalisation decreases. The key difference is that I choose to be in the 'anonymous Tor collective', trusting my privacy to unknown actors instead of putting trust in Google that assigns me to particular groups through their non-transparent process of collaborative filtering, without human agency.

Building upon knowledge gleaned from my preceding chapters and my 'subjectivities of search' effects, *Black Utopias*: *Surveillance Capitalism* (Chapter 9) attempts to answer my main research question, *how Google Search organises (us)ers?* It interweaves Tantner's 'Black Utopias' with Zuboff's 'surveillance capitalism', demonstrating how both are contingent on personal data—its collection, organisation and management by whomever controlled the register and advertisement journal, or nowadays databases and interfaces. With *Big Other's* 'logic of accumulation', users have become valuable resources as data producers with their 'ubiquitous googling', a term I introduce that augments Chun's research on habit (2016) and Stephens-Davidowitz's 'digital truth serum' (2017). Made possible by Big Data sorting and filtering

techniques, as well as increased storage and computational power, this 'behaviour surplus' of user data is merged from various contexts, with this 'mission creep' constructing user profiles that feed diverse industries (advertising, credit, insurance). In this manner, Google Search operates as a 'profiling machine' (Van Dijck 2010), creating abstractions of humans as *Data Dust*, which are comprised of dividuals, data doubles, data shadows and data derivatives that come to 'haunt' the user (Blackman 2019). By quantifying these subjectivities into economic value as 'surveillance assets', user behaviour is influenced and modified—thereby organised—which in turn generates prediction products for a market that trades exclusively in future behaviour (Zuboff 2015). Formerly, the politics of search engines shaped the web (Introna and Nissenbaum 2000), now with the contemporary implications of surveillance capitalism, I postulate how Google Search also shapes and organises (us)ers in return (Chapters 5, 8 and 9).

Chapter 10 describes some of the implications and consequences of such organisational media through three types of Agency: Ad, State and Computational. Rather than drawing on human memory, users consult 'the oracle' and by believing in the results of this 'mnemotechnic', instantiate 'In Google We Trust' as an 'organisational theology'. Being ranked number one in search results determines website visibility to users as well as empowering 'algorithms of oppression' (Noble 2018) that affect further inquiry. Acting as a gatekeeper that filters and curates content as an 'arbiter of truth', Google behaves as an omnipotent infrastructure despite earning the lion's share of its revenue as an ad agency. Because of its monopoly on search, anti-trust concerns have been raised with US law enforcement agencies bringing about lawsuits and the EU commissioner instilling fines for anti-competitive practices. Additionally, European legislation authorises the *Right to be Forgotten* to delete specific pasts to create open futures and since 2018, Europe (GDPR) and certain US states have enacted privacy regulation. Yet even after Snowden's revelations of XKeyscore tracking of users' search queries, it has become apparent that State Agencies' surveillance activities continue, such as with Rule 41 that enables remote searching worldwide. Moreover, during the 2016 UK and US elections, Google algorithms perpetuated fake news and affective 'clickbaiting' through ads, organising users into 'calculated publics' (Gillespie 2014). Whilst simultaneously collecting data on (us)ers ad infinitum, Google also cultivates certain ontological effects, such as scopophilia, with users searching others or themselves, not only for information. This 'machine seeing' or 'computational agency' alters human behaviour with the transfer of autonomous decisions to increasingly 'intransparent' machines through the 'habit of automaticity'. Cyberorganization as a continuous 'information as becoming' (Parker and Cooper 2016) closes futures through a never-ending, cybernetic feedback loop operating as a 'chronotype of prediction' (Beyes and Pias 2019) in an era of 'posthistory'.

The final chapter *Emancipatory Utopias: User Agency* (Chapter 11) shows how the political technologies of organising and making information accessible is not novel, nor is the resistance to such media collecting user data. Besides address offices cultivating privacy with secret registers, 'Emancipatory Utopias' organised cultural (arts, music, inventions) and educational (seminars, lectures, book lending) activities, which were free to the public with transcripts of the lectures distributed afterwards anonymously. Yet making information free and available to everyone as a commons, also for those who were poor or those wanting to remain anonymous, is not without implications. Hostility towards commoning practices continues today when (re)searching for knowledge with file-sharing, 'shadow libraries' or joining the p2p Tor collective in order to search anonymously.

It has now become a question of how to intervene with legislative and human agency, as demonstrated by Google's army of 'search raters' or the 'tech-savvy' netizens who embody resistance along with programming code. Privacy and anonymity concerns continue, along with questions of trust regarding p2p technologies, such as Tor. Additionally, there are a range of 'agencies of anonymity', such as activist strategies of 'Obscuration' (Brunton and Nissenbaum 2015) or users who have exercised agency by choosing to leave data in their heads, withholding secrets and search queries instead of becoming 'shareveillant' subjects (Birchall 2017). Rather than contributing to a 'world brain', programming tactics enable users to insert code in order to not be taken up in Google's index. Otherwise, when researching or in search of serendipity, users can choose to use an alternative search engine.

Future considerations

This thesis is situated at the interstices of Organisation Studies, Media Theory and Artistic Research and these disciplines comprise my interdisciplinary methodological framework, which is not without limitations. In regard to the fusion of search and research as mentioned throughout this thesis, it is crucial to recognise that search engines are not just artefacts but academic enquiry tools. Therefore, I acknowledge that I incorporated commercial and biased search results (Gillespie 2016) when searching with Google as a 'personalised subject'. Moreover, with my 'critical ethnography of the self' and as a 'deskilled' scholar as elucidated in The Cybernetic Hypothesis (Chapter 4), I realise my data gathering process, or 'experiment in living', has shortcomings. However, this humanities inspired reverse-engineering process attempted to make the politics of information technology visible as a 'disclosive ethical archeology' (Introna 2007) of human/algorithmic interaction.

Although incomplete, drawing upon software, privacy and feminist media scholars in order to analyse and discuss my findings has yielded theoretical insights that could also be situated in the field of Media and Communication. Or advance new fields such as the ISA (International Sociological Association) thematic group Digital Sociology, with research located at the intersection between Digital Technologies and Society. Although my 'data visualisation as transcription' method is grounded in artistic research, it continues the domain of Critical Cartography yet further development is needed with visualising natively digital data and the politics of 'real-time' devices (Weltevrede, Helmond, Gerlitz 2014) such as search engines, which the Digital Methods Initiative repurposes for social and cultural research.

Investigating 'organising the world's information and making it accessible and useful' might belong to the domain of Library and Information Studies, but without inclusivity and acknowledgement of non-western traditions of knowledge, it remains a colonial endeavour. Although the fields of Information Retrieval or Web Search offer much insight into how search engines work (Chapter 3,4,5), I did not directly address these disciplines. A range of academics, developers and thinkers have as well 'reimagined' how search could work, including Siva Vaidhyanathan (2011) who conceived of a 'human knowledge project' or Safiya Noble's 'Imagine Search Engine' that attempts to show how information could be presented differently (2018:180-181). Other researchers have questioned making search engines 'good for democracy' (Sison and Sack 2008) by advocating plurality, autonomy and access (without advertisements) as a society-oriented design in order to 'democratise search' (Rieder 2008). Already in 2000, Introna and Nissenbaum recommended that there should be public funding in order to develop transparency as well as access to 'more egalitarian and inclusive search

mechanisms and for research into search and meta-search technologies' (2000:181). In 2018 Noble echoed the call for more financial support in developing non-commercial endeavours and to observe the circulation of disinformation (185), with which I concur.

Google keeps its databases proprietary yet proposals for a public index of the web have also been put forth, which is crucial for a more diverse and pluralistic online searching ecology, as the index would be accessible to a variety of search engines (Lewandowski 2014). Michael Christen's YaCy (Yet another Cyberspace) is a free and decentralised p2p privacy search engine that crawls and simultaneously indexes websites as it delivers search results and more research is to be done with exploring its alternative results and autocomplete suggestions. This PhD research proposes how the search ecosystem of 'ubiquitous googling' in an era of surveillance capitalism organises (us)ers (Chapter 9). Besides Google collecting excessive amounts of user data as subjects search, it delivers 'free' results in exchange for users' labour. Analogous to other forms of invisible labour (housework, childcare) or under remunerated work (mechanical turking), a further analysis of this type of hidden activity by users searching could contribute to the field of Digital Labour.

Contrarily, by exploring the boundaries of anonymity I 'reimagined search' with Tor, which incorporates a default 'privacy search engine' in its bundle: StartPage (2013), Disconnect Search (2015) when I carried out my data gathering and since 2016, DuckDuckGo. Increasingly more alternative search engines have become available, each differing in approach (meta-search, vertical search) and more work needs to be conducted regarding user privacy and user data collection. Intertwining the complex relationship between issues of privacy and anonymity (Nissenbaum 1999, Marx 1999) and technologies of the 'media arcane' (Beyes and Pias 2019) in a surveillance society (Lyon 2010, Murukami-Wood 2008), additional research in this area could be developed within the field of Surveillance Studies. Other inroads include the 'extractivism' of user data (Zuboff 2015, Mezzandra and Nielsen 2017, Crawford 2021) and more recently, configurations of data, labour and territory with the Data Farms (2019) and Machine Listening (2020) projects. Paths leading to alternatives could include policy changes through regulation and legislation, technological literacy for the general public or 'abolishing' Big Data (Milner 2020).

Moreover, some of the outcomes of this study touch upon the domain of Critical Data Studies, which explores the ethical, political and cultural challenges instantiated by 'Big Data'. Continuing on-going research strands that address politics and equitability include the recent publications *Data Feminism* (D'Ignazio and Klein 2020), *Data Colonialism* (Couldry and Mejias 2019) and the 'Datactive' (Milan) and 'Data Justice' (Denzik) projects. Ruha Benjamin's *Race after technology* (2019) argues for understanding race as a technology and more work needs to be done on reimagining information retrieval and search results in the service of eradicating social inequality, which will 'not be solved by an app' (Noble 2018). Additionally, Critical Algorithm Studies is a developing field that examines black-boxed technologies, with increasingly more data scientists and humanities scholars speaking out about the lack of algorithmic transparency and accountability (Gillespie 2014, Pasquale 2015, Diakopoulos 2016, O'Neil 2017, Noble 2018). Therefore, proposals for public institutes that develop standards and ethical guidelines for implementation by the US tech industry (O'Neil 2017) are timely and warrant further investigation, as are appeals for human interpretation, experimentation and oversight or studying the effects and ethics of AI.

Lastly, this thesis demonstrated some effects from human interaction with the algorithmic Google Search interface, although it did not address the design or implementation, as with the domain of Human Computer Interaction (HCI). Rather, the black-boxed effects of 'search subjectivities' (Chapter 8) and the areas of Cyberfeminism (https://cyberfeminismindex.com/) and Cybernetics remain underexplored. Additionally, the thesis indicates research gaps regarding *How We Became Posthuman* (Hayles 1999) and the field of Artificial Intelligence. Specifically, Google's machine-learning RankBrain, which applies neural networks that contain feedback loops and are the basic concept of cybernetics, begs further investigation. With the shift from information retrieval to the conversational web, a natural outgrowth of this thesis is the question concerning 'the future of search' that would combine a few unfinished research threads: modelled on the figure of a domestic male servant (Ask Jeeves!) and female secretary, Virtual Personal Assistants (VPAs) and chatbots reproduce master/slave relationships from former centuries and computer programming.

Instead of 'one voice' search ecosystems that not only surveil but anticipate and monetise people's lives (Zuboff 2019), a feminist ethics of care, or Pirate Care (2020), could provide (post)humanist agency in these technics of mediation (Stiegler 1994). The interdisciplinary field of Feminist Technoscience Studies, located at the interstices of Science and Technology Studies (STS) and digital technologies, interrogates boundaries by giving voice to invisible and marginalised positions within the entanglement of humans and computational machines. The legacies of feminist critique present ways forward in understanding the complexities of VPAs, their embodiment of science fiction narratives and the embeddedness of gender discrimination in society.

Epilogue

The Selfish Ledger

In May 2018 a 'leaked' Google video 'for internal use only' surfaced.[220] Made in late 2016 and inspired by Richard Dawkins' *The Selfish Gene* (1976), *The Selfish Ledger* is a 'thought experiment' in 'speculative design' written by Nick Foster and David Murphy. [221] It begins with Jean-Baptiste Lamarck, who in 1809 published the first comprehensive theory of evolution, *Philosophie zoologique* that 'introduced an internal code for every living thing, which when passed down through generations, defined the physiological characteristics of a species' (ibid). The organism adapts the code and upon reproduction, passes down a modified version to its young, what Lamarck deemed 'adaptive force'. Although Darwin's natural selection usurped his theory, this epigenetic concept has here been revamped into 'Lamarkian user data' (ibid). Consisting of a trail of 'actions, decisions, preferences, movement and relationships', the 'constant evolving representation of a codified version of who we are' is written down in a ledger (ibid). (Figure 112)

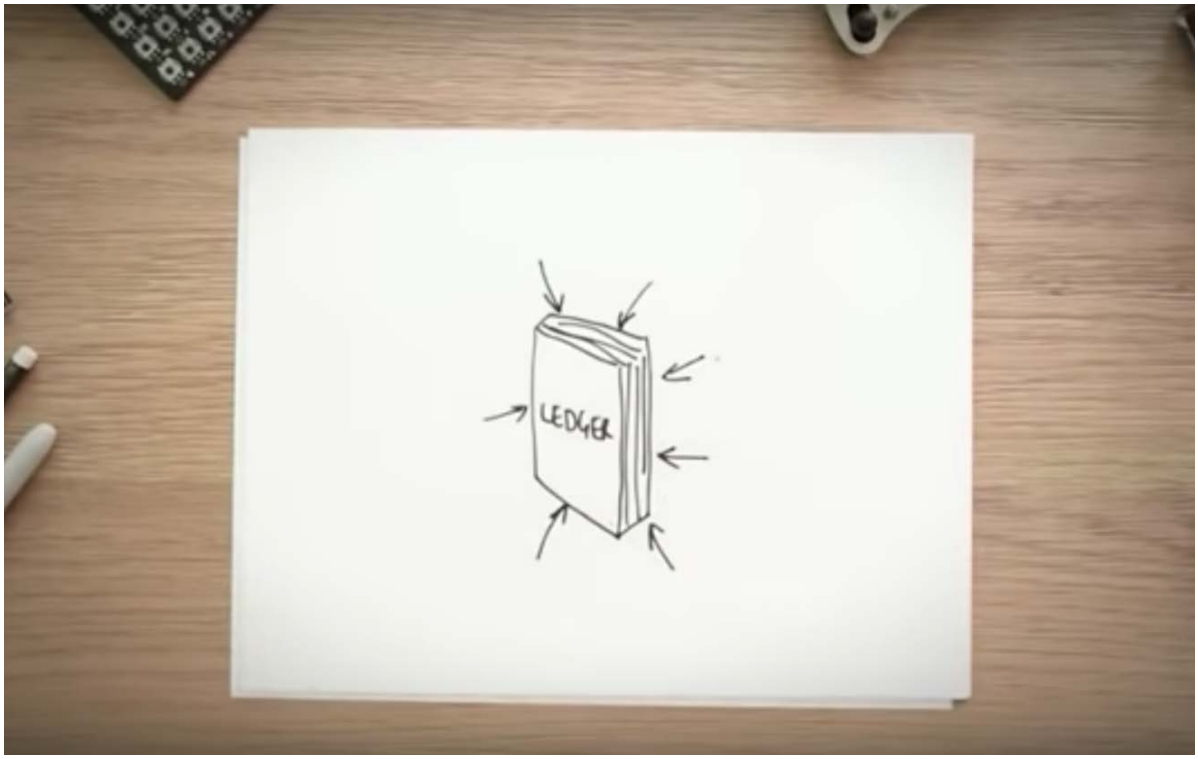


Figure 112: videostill *The Selfish Ledger*

In a manner analogous to the registers of the address offices, this *hupomnemata* continuously captures data as a 'Lamarkian epigenome', which becomes 'more complex, developing, changing and forming based on our actions' (ibid). Drawing on Bill Hamilton's contribution to knowledge— it 'is not whether behaviour is to the benefit of the behaver but whether it is to the benefit of gene'— evolutionary biologist Richard Dawkin purported that although the 'gene is

[220] The 8-minute video is on You Tube, paradoxically (or not) on Google servers: https://www.youtube.com/watch?time_continue=22&v=QDVVo14A_fo

[221] In co-production with David Murphy, it is narrated by Nick Foster who is now head of design at X (formerly GoogleX) and part of the 'Near Future Laboratory' that transforms research practice into potential products.

devoid of any motives or will, it could be pedagogically and metaphorically described as if it were' (ibid). Moreover, 'the individual organism is a transient carrier, a survival machine for the gene' and following their research, Foster and Murphy ponder:

> What if the ledger could be given a volitional purpose, rather than simply acting as a historical reference? What if we focused on creating a richer ledger, by introducing more sources of information? What if we thought of ourselves not as the owners of this information but as custodians, transient carriers, or caretakers? (ibid).

Here *The Selfish Ledger* embodies Dawkin's gene, finding inspiration in the continuation of the carrier (in this case the ledger full of human data) and not the human itself. Moreover, the user-driven 'goal-oriented ledger' reflects the corporation's values (Google) and would be 'responsible for offering suitable targets for a user's ledger' (ibid).

> Once a user selects a volition for their ledger, every interaction may be compared to a series of parallel options. If one of these options allows the ledger to move closer to its goal, it will be offered up to the user. Over time, by selecting these options, the user's behaviour may be modified and the ledger moves closer to its target (ibid).

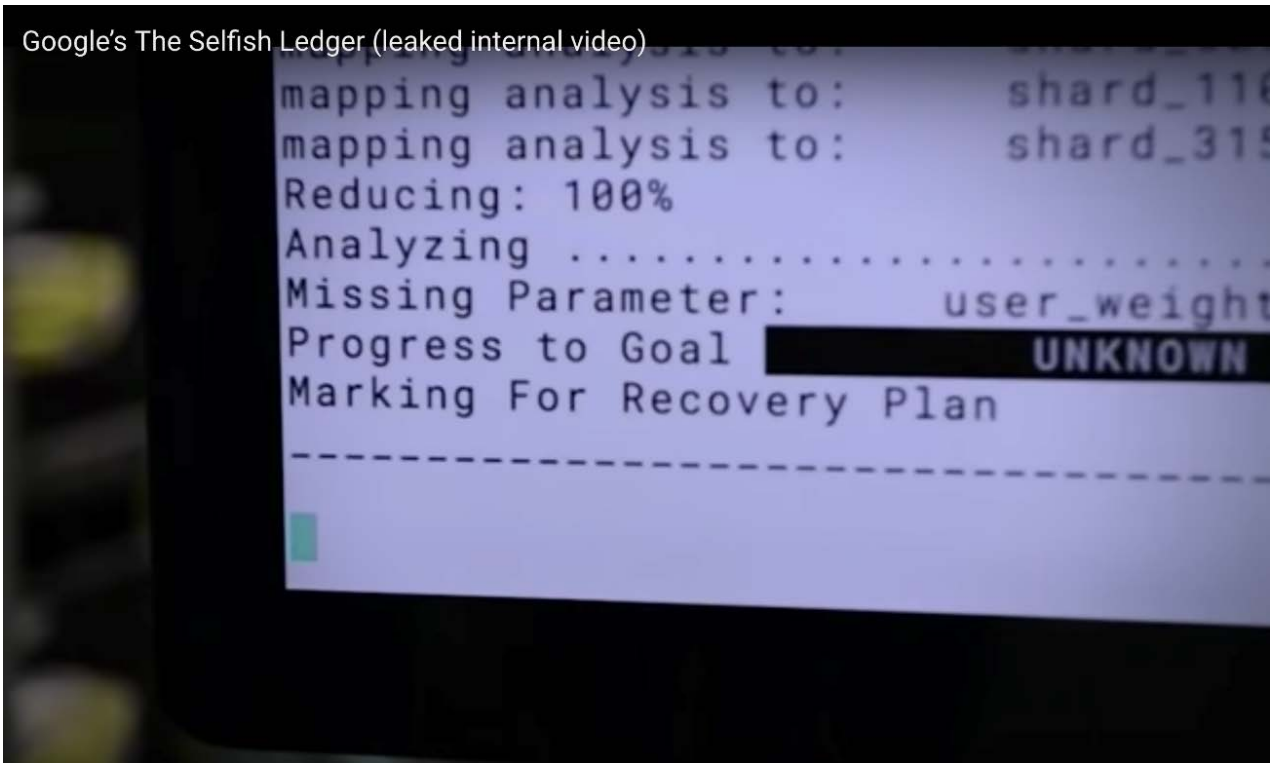


Figure 113: videostill *The Selfish Ledger*

A future scenario of user interaction with the app-like ledger and the altering of user behavior expounds upon how the ledger might carry out searches on its own volition, where 'suggestions may not be converted not by the user, but by the ledger itself' (ibid). The video then shows a hypothetical situation where there is a 'gap' in its knowledge, due to 'missing a key data source' such as a parameter (ibid). (Figure 113) Connected to the IoT, the ledger starts searching for the exact device that is needed 'in order to better understand its user', and then sorts through the options until it finds that which would appeal to the user (ibid).

> In situations where no suitable product is found the ledger may investigate a bespoke solution. By analysing historical data, it is increasingly possible to discern qualitative information such as taste and aesthetic sensibility, which may be used in the creation of a design proposal (ibid).

The Selfish Ledger could provide answers and even build and design custom objects with 3D printers in order 'to trigger the user's interest' (ibid). The video also elucidates how *The Selfish Ledger* would customise recommendations, such as a product with a 'suggestion' price of $45,90 that appears on the screen. (Figure 114) Through this process, *The Selfish Ledger* would be able to 'refine its model of human behaviour' and 'plug gaps in its knowledge' (ibid).

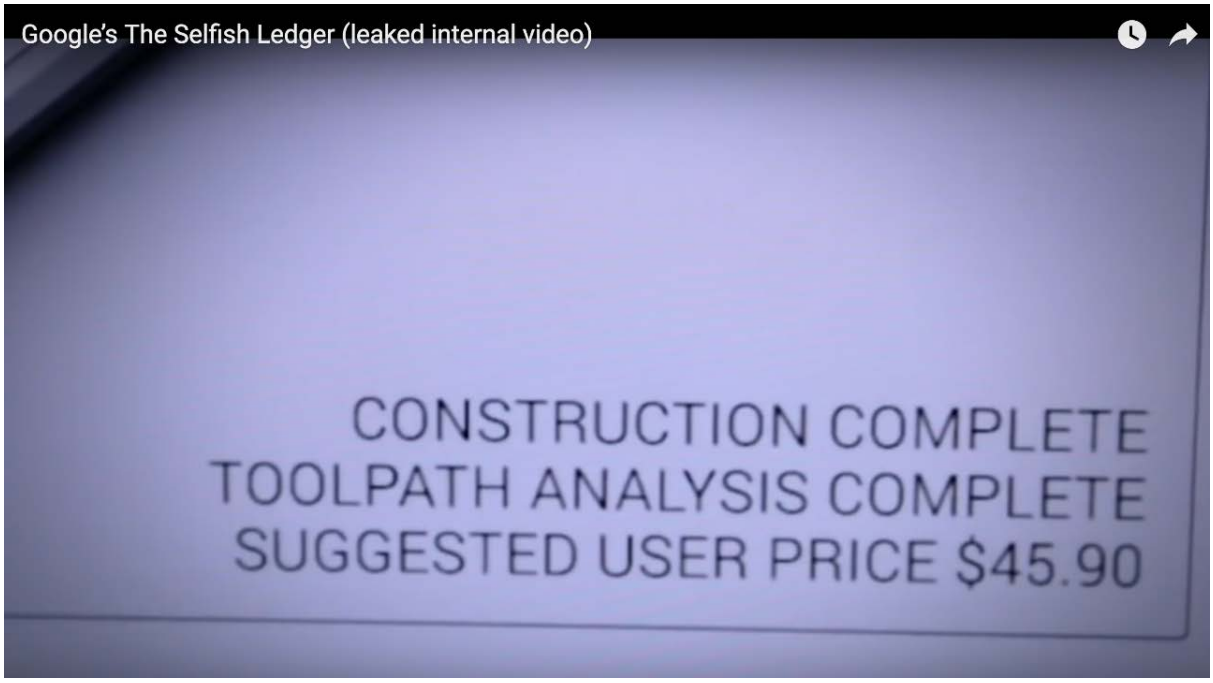


Figure 114: videostill *The Selfish Ledger*

The video then proposes the survival of user data beyond the mere mortal, aka the biological self or the 'data subject', which will be passed down to following generations, analogous to how genetic code is propagated by diverse species in nature. Forthcoming generations could benefit from the past knowledge of others as well as their 'behaviours and decisions'.

> By comparing this emergent ledger with the mass of historical user data it becomes possible to make increasingly accurate predictions about decisions and future behaviours (ibid).

With user data now residing in the databases on Google servers worldwide, in the future *The Selfish Ledger* could bring 'streams of information' together and 'new patterns become apparent and new predictions become possible' (ibid). In this way, the organisational process of searching through 'sensor-enabled objects' would (re)organise us(ers) by making decisions, only without (y)our own volition.

The last section of the video is entitled *Behavioural Sequencing*, with the terminology referring to the manner in which billions of user data points taken up in the ledger could be adopted and adapted in order for 'it to better understand its role'. Similar to how the human genome has been

‘sequenced’ in human biology, with *The Selfish Ledger* user data could also be studied and targeted in order to modify certain aspects of behaviour. (Figure 115) As the growth of user data continues exponentially, the ‘mass multi-national examination of actions and results could introduce a model of behavioural sequencing’ (ibid). Just as human biology has mapped and targeted certain genes, as ‘patterns begin to emerge in the behavioural sequences, they too may be targeted’ (ibid).

Figure 115: videostill *The Selfish Ledger*

Finally, the authors contextualise their ‘thought experiment’ as a ‘journey to understand user data’ and they intend to apply ‘knowledge of epigenetics, inheritance, and mimetics to this field’ (of research?) and to offer benefits to ‘our species as a whole’ (ibid). However, they conclude by divulging:

> The ledger could be given a focus, shifting it from a system which not only tracks our behaviour but offers direction towards a desired result (ibid).

Whose ‘desired result’ do they actually mean, *The Selfish Ledger*’s or the human whose data has been ‘given’?

A search result with no instructions on how to look for it?

Pynchon’s *Bleeding Edge* (2013) depicts the datascape of ‘Silicon Alley’—New York City in 2001 as what it was as well as how it could have become otherwise. In this excerpt Pynchon exhales his longest sentence, recounting club-goers dispersing in the early morning, back out into downtown Manhattan on the Saturday before 9/11 yet after the dot.com bubble burst.

Faces already under silent assault, as if by something ahead, some Y2K of the workweek that no one is quite imagining, the crowds drifting slowly out into the little legendary streets, the highs beginning to dissipate, out into the casting-off of veils before the luminosities of dawn, a sea of T-shirts nobody's reading, a clamor of messages nobody's getting, as if it's the true text history of nights in the Alley, outcries to be attended to and not be lost, the 3:00 AM kozmo deliveries to code sessions and all-night shredding parties, the bedfellows who came and went, the bands in the clubs, the songs whose hooks still wait to ambush an idle hour, the day jobs with meetings about meetings and bosses without clue, *the unreal strings of zeros*, the business models changing one minute to the next, the start-up parties every night of the week and more on Thursdays than you could keep track of, which of these faces so claimed by the time, the epoch whose end they've been celebrating all night—which of them can see ahead, among the microclimates of binary, tracking earthwide everywhere through dark fiber and twisted pairs and nowadays wirelessly through spaces private and public, anywhere among cybersweatshop needles flashing and never still, in that unquiet vastly stitched and unstitched tapestry they have all at some time sat growing crippled in the service of—to the shape of the day imminent, *a procedure waiting execution, about to be revealed, a search result with no instructions on how to look for it?* (Pynchon 2013:311-312 emphasis mine)

Throughout the novel a spectre haunts the reader without it ever being mentioning by name, only alluded to. The dead giveaway is perhaps 'the unreal strings of zeros' referring to googol (10,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000) or '1e100.net'.[222] As an analogy to organising the world's information and in the days before 'autocomplete', founders Brin and Page admitted that they had wanted to call their search engine googol but mistyped it as 'google'. Since October 2009, Google has made sure that each IP address has a corresponding hostname with '1e100.net,' a Google-owned domain name used to identify all the servers in its network.

One giant cybernetic collective

> Like Turing's machine and DNA, Google makes no material separation between metamessage and message—which are made of the same exact stuff. Turing's infinitely spooling paper strip contains both data and commands. In the DNA code structural and control genes mingle side by side—that is, epigenetic control of the code is managed by the code itself. In Google the material of search is the web itself. The recursive collapse of the "meta" into the thing itself is a distinctive feature of media in the age of Turing (Peters 2015:328)

In his 1954 essay *The Question Concerning Technology*, Martin Heidegger stated that 'whatever has effect as its consequence is called a cause' yet technology is also a 'means' to an 'end', employed by its 'instrumentality' (Heidegger 1954:6). Instead of 'bringing forth' as poiesis by revealing 'truth', modern technology is 'challenging-forth', treating nature as a *Bestand* or 'standing reserve', ordered and gathered together as an all-encompassing orientation to the world. With Heidegger's 'revealing the real'—Enframing—the 'essence' of technology is further contextualised by his use of the term 'human resources'—human beings as raw material.

---

[222] Googol's scientific notation is 1e100. Coined by a nine-year old (Milton Sirotta) in 1920, a googol refers to the number 10^{100}. https://en.wikipedia.org/wiki/Googol.

Following Heidegger, Enframing is ultimately the forgetting of being and 'resistance is a problem before us in which we need to rethink the dangers of technology' (Noys 2019).

With the logic of accumulation of user data through the extractivism of surveillance capitalism (Zuboff 2015), the human population now becomes a 'standing reserve' as a means to an end. *The Selfish Ledger* envisions a possible Google future of 'profiling machines' in a world where every action is recorded, nudging users towards certain purchasing behaviours as it updates the ledger. Moreover, with volition shifting from the individual to the ledger, the decision-making processes have been subcontracted to algorithms as 'habits of automaticity' because everything is recommended through 'personalisation'. Ultimately this dystopian vision is a world where human behaviour and decision-making is automated to the extent that prediction products are spontaneously produced after serving up advertisements, with the user only being able to choose from a limited selection, which had been searched by and produced by *The Selfish Ledger*.

Another possible future is that 'mnemotechnics' such as *The Selfish Ledger* will assist in altering users' behaviour into never-forgetting through an ecosystem of surveillance assets, selling users' memories back to them instead of just vending prediction products. The eventual realisation of Google's *The Selfish Ledger* could mark a shift from information storage and retrieval mechanisms, to a continuous experiment in making the internet an enormous computing mechanism (Buterin 2016)

> There's this collective AI in Google Search, where we are sort of plugged in, like nodes on the network. Like leaves on a big tree. And we're all feeding this network with our questions, and answers. We are all collectively programming the AI. And Google plus all the humans that connect to it are one giant cybernetic collective (Musk 2018).[223]

Or RankBrain might someday gain the semblance of consciousness and could eventually informate as well as automate (Zuboff 1981, 2015), producing the information it organises, making (us)ers thereby redundant. Instead of humans passing on the code of the gene in order to procreate, *The Selfish Ledger* becomes the carrier of the code, making humans obsolete.

[223] 'This is also true of Facebook. Twitter. And Instagram and all of these social networks. They are giant cybernetic collectives'. Elon Musk interview with Joe Rogan, September 6, 2018: https://www.youtube.com/watch?v=Ra3fv8gl6NE

Appendices

Appendix A: How keyword search works

After a user types a 'keyword' in a 'search box' or field, search engines have to first send out webcrawlers, 'small bits of code sometimes called spiders, robots or bots' which are programmes that 'crawl' the World Wide Web and are regularly updated (Haeselin 2017:4).[224] An indexer, another programme, then builds an index of all of the gathered documents and reads these pages, which are called up from tree-like organisational structures. The process of indexing and the search methods which algorithms use (linear, binary, prefix trees, AVL trees and (distributed) hash tables, etc.) determine how URLs can be found. Instead of humans building indexes, as with previous curated indexes, Google uses the architecture of an inverted index, the 'generic data store at the heart' that 'is a very efficient method of storing and searching over the contents of documents' (Hugill et al. 2013:244).

> In an inverted index, the contents of a document are broken into various different combinations or terms by the indexer, and a link to the original document is stored with each of these terms. This means that when searching for a keyword, instead of having to look at every document and its contents, the system just looks for all terms that match the request and returns the various links that match. The inverted index is quick at retrieval; however, building the index is slower (ibid).

Algorithms are software programmes or 'recipes' containing lists of instructions and variables (numeric data, text) that when executed deliver results, or URLs (Uniform Resource Locator). These results could be a document, or a webpage (SERPs: search engine result page) or even an image that is tagged with the chosen keyword.

As explained in Chapter 4, Google Search structures results for the user, based on a given set of hidden criteria or 'signals'. Proprietary algorithms adjust the 'weight' or importance of the delivered results in order to 'rank' them, with the default of 10 SERPs to a page. Understanding that search engines 'give added weight to terms and phrases that appear frequently on a given page, and comparatively less frequently on the rest of the web' (Halavais 2009:38), the ranking reflects pages that use the words more than once, not only in the title. Eliminating redundant or neutral words is also important.[225] 'Short, common words like [and], "the" or "to" are generally considered "stop words" and ignored by search engines precisely because they are so useless' (ibid:37).[226] Each query delivers different results, measured by their 'relevance'— first by the search engines and then by the user, who ultimately decides which search engine to use.

---

[224] Without the automated processes of crawling and recording the state of the Web, no search engine would return very good results; in fact, they would not return any results at all (Haeselin 2017:4).

[225] 'Google Boolean Search: Basic Boolean search commands (quotes, AND and OR) are supported in Google search, however Google defaults to AND searches automatically, so you don't need to enter AND into the search box. Google search uses additional symbols and words to refine searches such as 'site:' to search a specific site or domain or use $ in front of a number to search for a price'. Available at: http://www.webopedia.com/TERM/B/Boolean_search.html

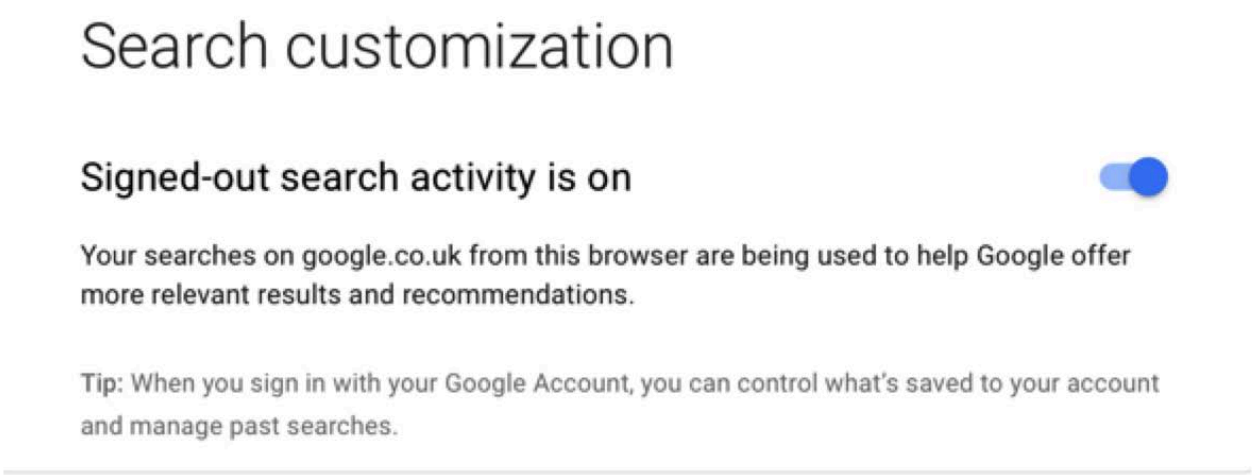
[226] 'A handful of words like "the" and "my" are often dismissed as "stop words" and not included in the index because they are so common. Further application of natural language processing (NLP) is capable of determining the parts of speech of terms, and synonyms can be identified to provide further clues for searching. At the most extreme end of indexing are efforts to allow a computer to in some way understand the genre or topic of a given page by "reading" the text to determine its meaning' (Halavais 2009:17-18).

Appendix B: Critical Ethnography of the Self

In the years leading up to beginning my PhD I used 'Ghostery' or other adblocking plug-ins with Firefox to prevent my browsing history from being tracked by Google but now I was venturing out unprotected into the 'wild' by letting myself be personalised.

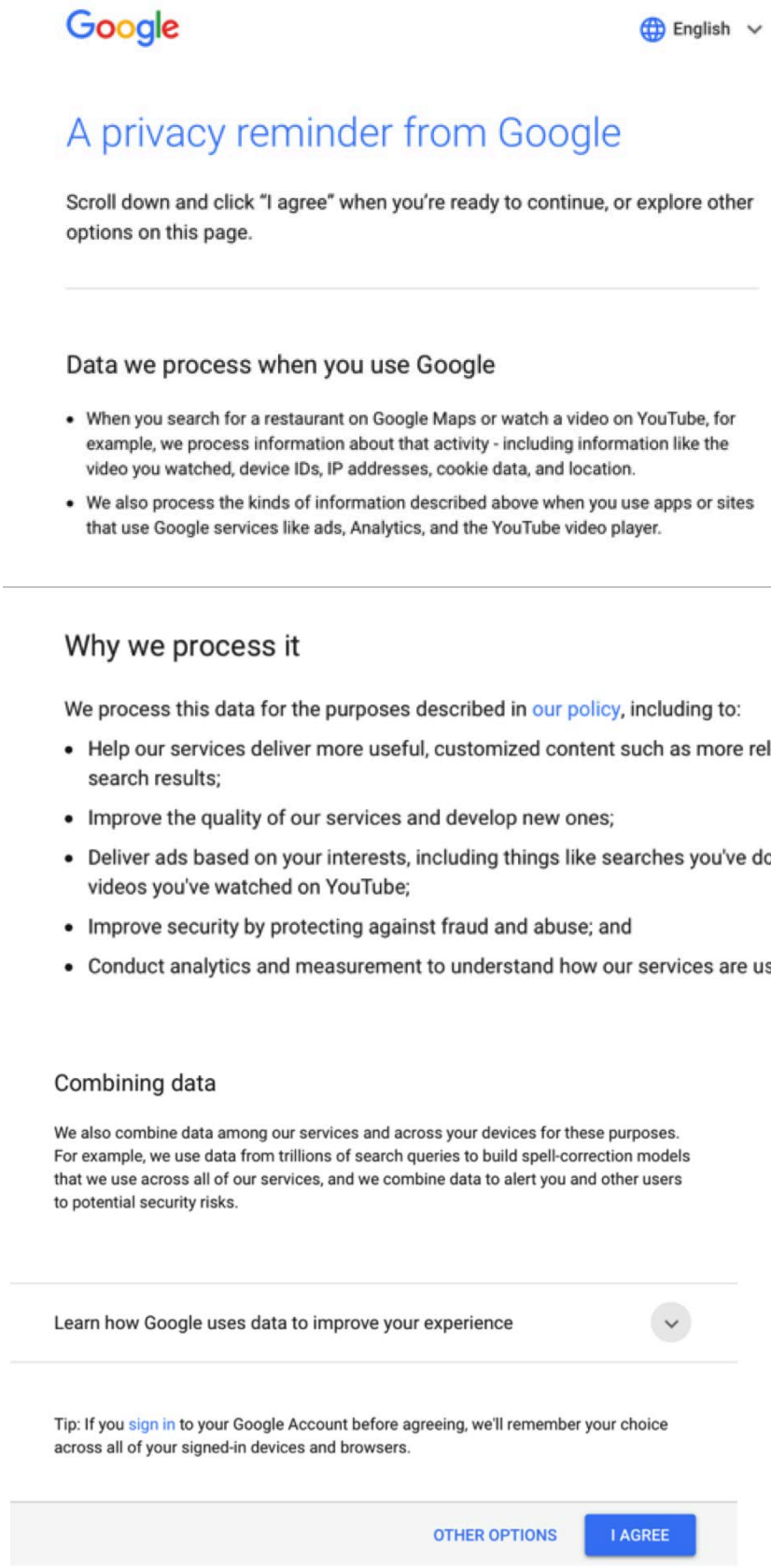
In March 2014 Leuphana University gave me a computer to use for my PhD. It was a 13-inch Powerbook, Apple Macbook Pro,version 10.9 Retina display from the end of 2013 with a 2,6 GHz Intel Core i5, 16 GB 1600 MHz DDR3 of Memory and an Intel Iris 1536 MB Graphics card. I installed no privacy plug-ins (Ghostery, AdBlocker, etc.). My 'personalised' Apple is (still) using an OS 10 Yosemite operating system, where Google applies its algorithms to offer relevance and recommendations. I did not update the operating system during the course of my PhD. As far as I know, no hidden tracking components have been installed. Certain applications were downloaded from the internet, such as Mozilla's Firefox which I used as my browser.

I entered keywords into a search box as a daily habit (Chun 2016) and I was personalised for about 19 months before I started my 'experiment in living' (Marres 2012).[227] Even when signed out of my Google account, my settings for search customisation were set to 'on'.



Even though Google asked me at a certain moment to structure my privacy, I did not employ 'incognito' mode, which is Google's terminology for 'private browsing'.[1] The screenshots I show below were captured in July 2015 when Google informed me of their privacy updates and stated that they collected data and process information regarding my activities, including search, from various devices, IP addesses, cookie data and location. Everyone who uses Google now receives this information and is forced to agree to how they process information, ranging from improving services, customizing results, targeting ads, conducting analystics, combining data to security services and 'customization' a.k.a. personalisation if they wish to use Google.

[227] I describe 'personalisation' in more detail in the Analysis section of Chapter 4.


Google
English
A privacy reminder from Google
Scroll down and click "I agree" when you're ready to continue, or explore other options on this page.
Data we process when you use Google
When you search for a restaurant on Google Maps or watch a video on YouTube, for example, we process information about that activity - including information like the video you watched, device IDs, IP addresses, cookie data, and location.
We also process the kinds of information described above when you use apps or sites that use Google services like ads, Analytics, and the YouTube video player.
Why we process it
We process this data for the purposes described in our policy, including to:
Help our services deliver more useful, customized content such as more relevant search results;
Improve the quality of our services and develop new ones;
Deliver ads based on your interests, including things like searches you've done or videos you've watched on YouTube;
Improve security by protecting against fraud and abuse; and
Conduct analytics and measurement to understand how our services are used.
Combining data
We also combine data among our services and across your devices for these purposes. For example, we use data from trillions of search queries to build spell-correction models that we use across all of our services, and we combine data to alert you and other users to potential security risks.
Learn how Google uses data to improve your experience
Tip: If you sign in to your Google Account before agreeing, we'll remember your choice across all of your signed-in devices and browsers.
OTHER OPTIONS
I AGREE

Panopticlick

How Unique — and Trackable — Is Your Browser?

Your browser fingerprint **appears to be unique** among the 6,011,043 tested so far.

Currently, we estimate that your browser has a fingerprint that conveys **at least 22.52 bits of identifying information.**

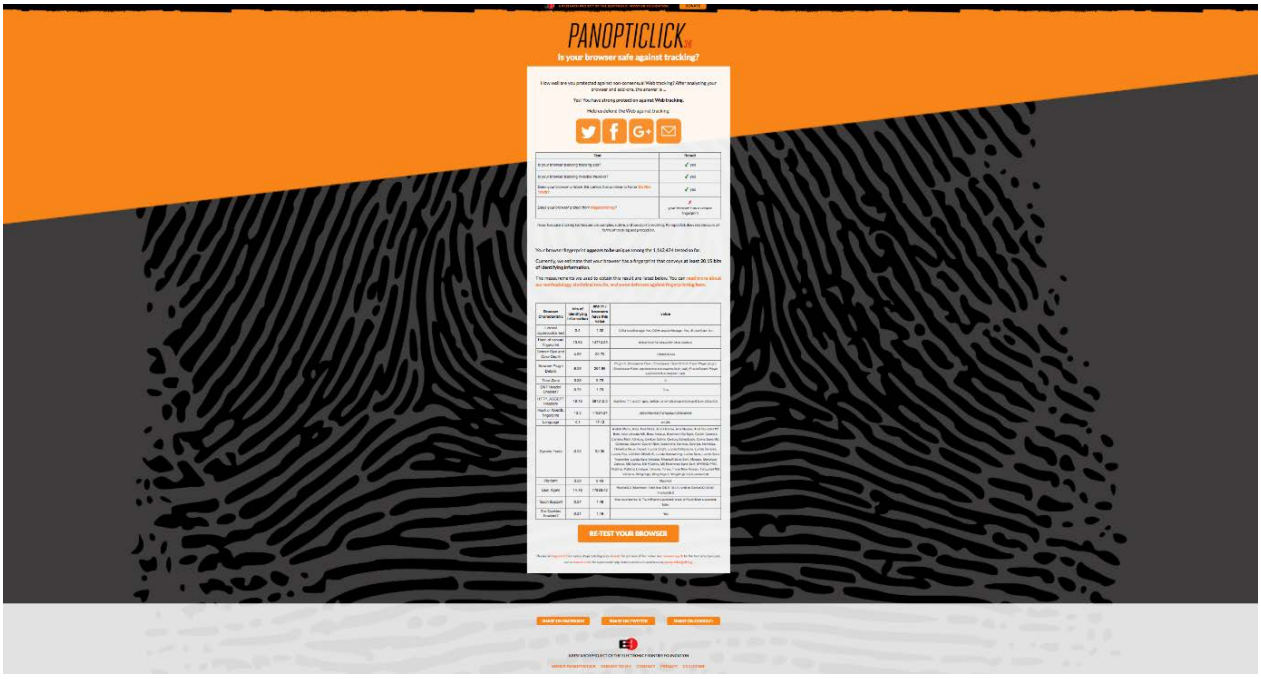
The measurements we used to obtain this result are listed below. You can read more about our methodology, statistical results, and some defenses against fingerprinting in this article.

Help us increase our sample size:

Browser Characteristic	bits of identifying information	one in *x* browsers have this value	value
User Agent	12.45	5586.47	Mozilla/5.0 (Macintosh; Intel Mac OS X 10.10; rv:41.0) Gecko/20100101 Firefox/41.0
HTTP_ACCEPT Headers	21.52	3005521.5	text/html, */* gzip, deflate en-GB,en;q=0.8,de;q=0.5,en-US;q=0.3
Browser Plugin Details	21.52	3005521.5	Plugin 0: Default Browser Helper; Provides information about the default web browser; Default Browser.plugin; (Provides information about the default web browser; application/apple-default-browser;). Plugin 1: Google Earth Plug-in; The Google Earth Plugin allows you to view 3D imagery and terrain in your web browser.; Google Earth Web Plug-in.plugin; (Google Earth browser plugin; application/geplugin;). Plugin 2: Java Applet Plug-in; Displays Java applet content, or a placeholder if Java is not installed.; JavaAppletPlugin.plugin; (Java Applet; application/x-java-applet;version=1.4;) (Java Applet; application/x-java-applet;version=1.2.1;) (Java Applet; application/x-java-applet;jpi-version=1.8.0_60;) (Java Applet; application/x-java-applet;version=1.2.2;) (Java Applet; application/x-java-applet;version=1.3;) (Java Applet; application/x-java-applet;version=1.8;) (Java Applet; application/x-java-vm;) (Java Applet; application/x-java-applet;version=1.1.1;) (Java Applet; application/x-java-applet;version=1.2;) (Java Applet; application/x-java-applet;version=1.7;) (Java Applet; application/x-java-applet;version=1.4.1;) (Java Applet; application/x-java-applet;version=1.1.2;) (Basic Java Applet; application/x-java-applet;) (Java Applet; application/x-java-applet;deploy=11.60.2;) (Java Applet; application/x-java-applet;version=1.1;) (Java Applet; application/x-java-applet;version=1.1.3;) (Java Applet; application/x-java-applet;version=1.6;) (JavaFX Applet; application/x-java-applet;javafx=8.0.60;) (Java Applet; application/x-java-applet;version=1.4.2;) (Java Applet; application/x-java-applet;version=1.3.1;) (Java Applet; application/x-java-applet;version=1.5;). Plugin 3: QuickTime Plug-in 7.7.3; The QuickTime Plugin allows you to view a wide variety of multimedia content in web pages. For more information, visit the <A HREF=http://www.apple.com/quicktime>QuickTime</A> Web site.; QuickTime Plugin.plugin; (Video For Windows; video/x-msvideo; avi, vfw) (3GPP2 media; video/3gpp2; 3g2,3gp2) (MP3 audio; audio/mpeg3; mp3,swa) (MP3 audio; audio/mp3; mp3,swa) (CAF audio; audio/x-caf; caf) (MPEG audio; audio/mpeg; mpeg,mpg,m1s,m1a,mp2,mpm,mpa,m2a,mp3,swa) (QuickTime Movie; video/quicktime; mov,qt,mqv) (MP3 audio; audio/x-mpeg3; mp3,swa) (MPEG-4 media; video/mp4; mp4) (SDP stream descriptor;

My Panopticlick fingerprint in October 2015.

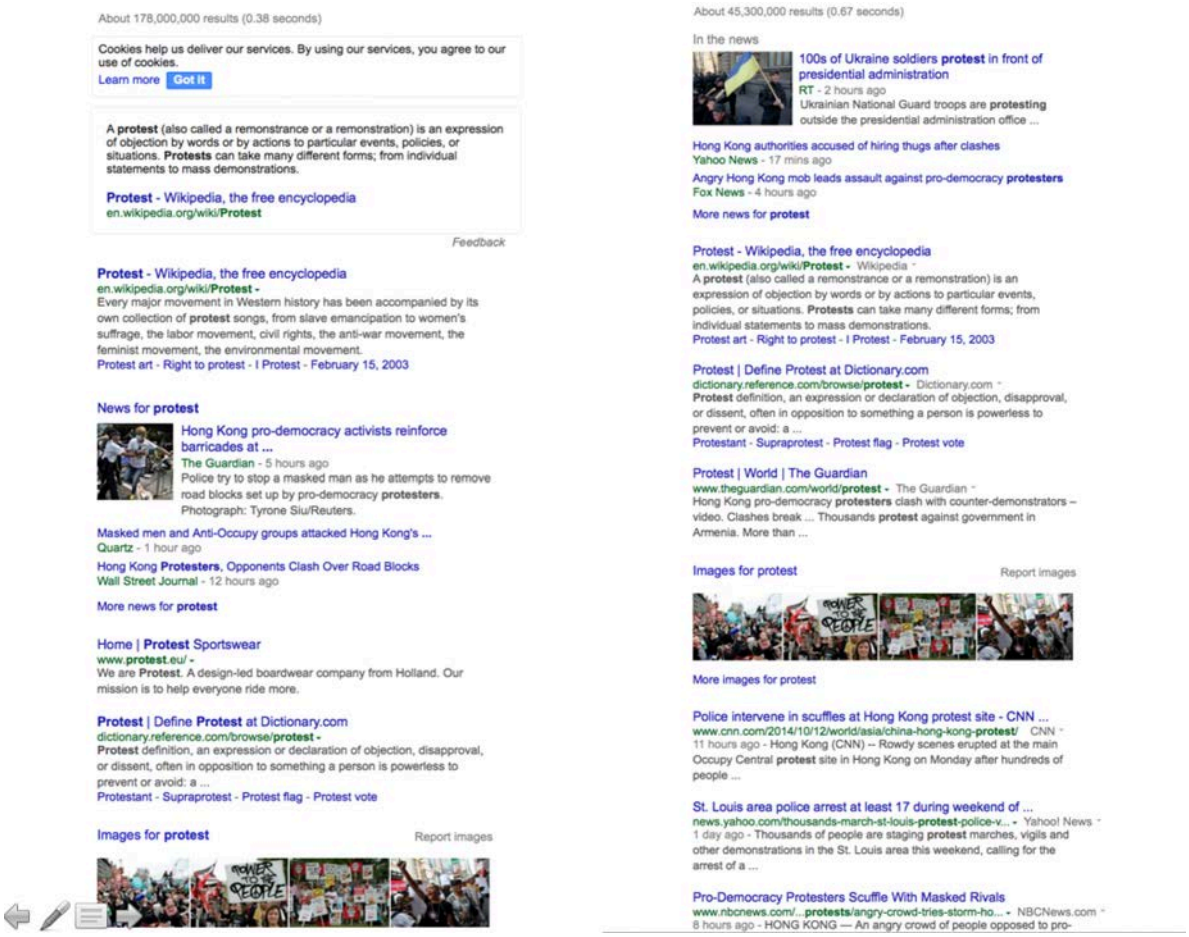
When researching how to design my experiment I searched on my 'personalised' Apple using Google Search to see if there were programmes available that could 'measure', 'capture' and 'visualise' my personalisation. My Google results didn't return any specific recommendations. Before I started my experiment I used 'Panopticlick' that is produced by the Electronic Frontier Foundation and shows how a user's browser fingerprint is tracked by listing the signals that effect user behaviour online. The bottom image provides a complete overview.



Appendix C: First Tests

In preparation for my *Datafied Research* PhD course in Hong Kong at City University in collaboration with Aarhus University, I made a series of small tests in order to find a way to compare the Google Search results with that of the Tor Browser. My first series of test trials started in Copenhagen on September 12, 2014 and lasted until October 12, 2014. In this small experiment I used Google Search in Firefox and Tor Browser. Prior to the release of *Tor* Browser version 4.5, Tor was using 'startpage.com' as its default search engine.[228]

I began by 'testing out' non-commercial keywords, that is, keywords that were uncommon. I chose words that I came across in articles I was reading in the first year of my PhD regarding currency, protest, research and subjectivity. Certain keywords 'trend' more than others, depending on current events. I specifically chose the keyword 'protest' because of the 'umbrella revolution' in Hong Kong at that time. When I searched with the word 'protest', for example, I received the following results as Page 1 with Google Search (left column) that stated 178,000,000 results at 0.38 seconds. I received 45,300,000 results at 0.67 seconds using the Tor Browser (right column). At this time, I only made screenshots of the first pages or the last page with Google and Tor.



Page 1 of search results for the word 'protest', retrieved at my CBS office in Copenhagen. Left column Google, right column Tor.

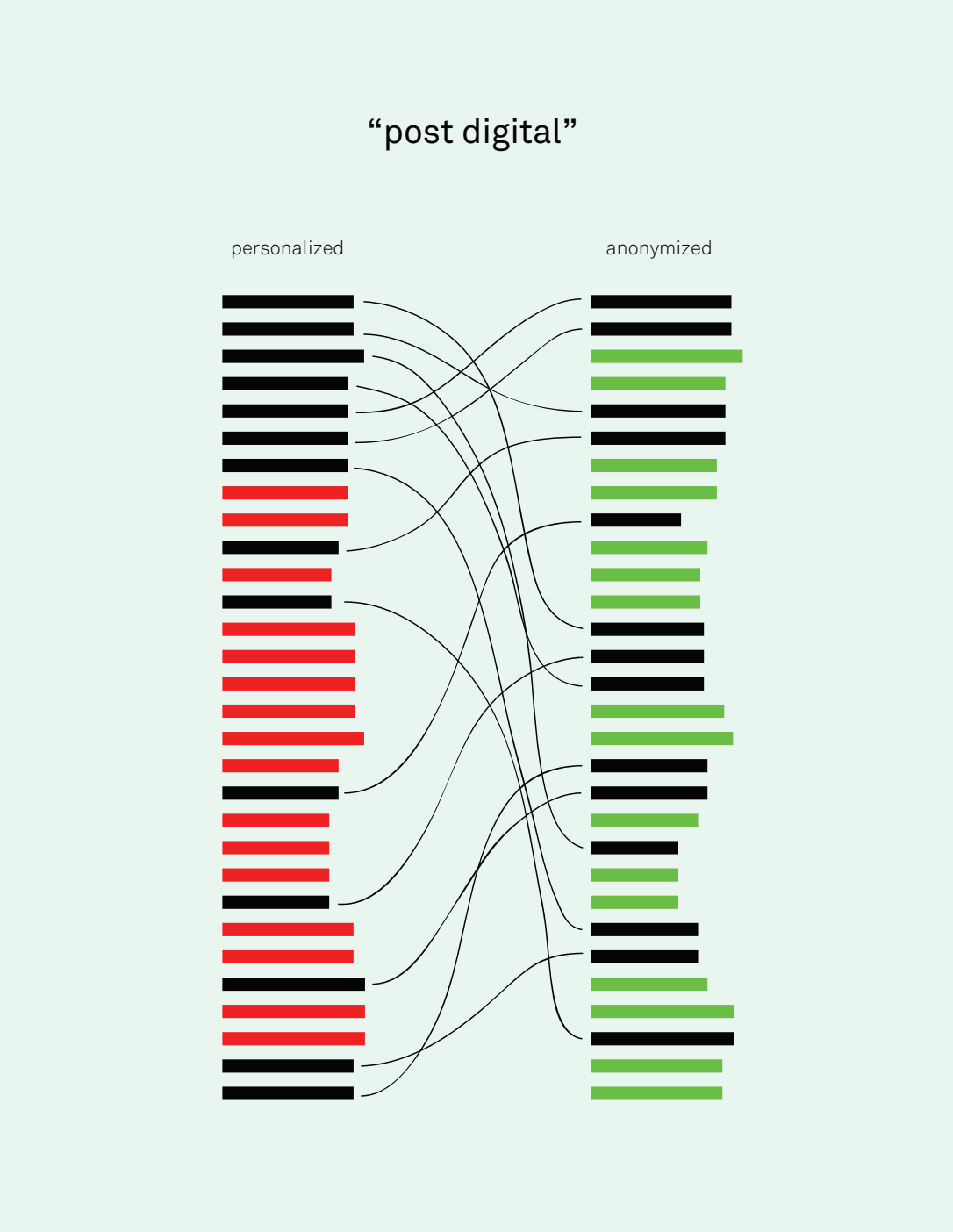
[228] Startpage returns Google search results without users IP address and personal information. Originally part of the metasearch engine Ixquick, it markets itself as a 'privacy' orientated search engine.

The most interesting news is often on the last page, where Google Search (page 73) delivered an URL discussing the US Media blackout of Wall-Street protest, described by a Chinese blogger as 'one of the best kept secrets in the US' and that the protest is 'shameful'. In the beginning I couldn't control which Tor server I was being assigned, e.g. one in Thailand or in Czech Republic, for example. I only used standard settings for the searches, with 10 requests per page (default) and did not choose country or language location. In the image shown below the Tor Browser assigned me a Polish server so this search shows everything in Polish. At this time the Tor Browser was delivering 93 pages for the same search query. Because of these tests I realised that words like 'research' returned too many results and therefore decided to shift my focus in trying to set up an experiment with more specific terms that would not deliver so many results for a given word.

The 93rd page of results for 'Protest', Tor Browser. November 2014.

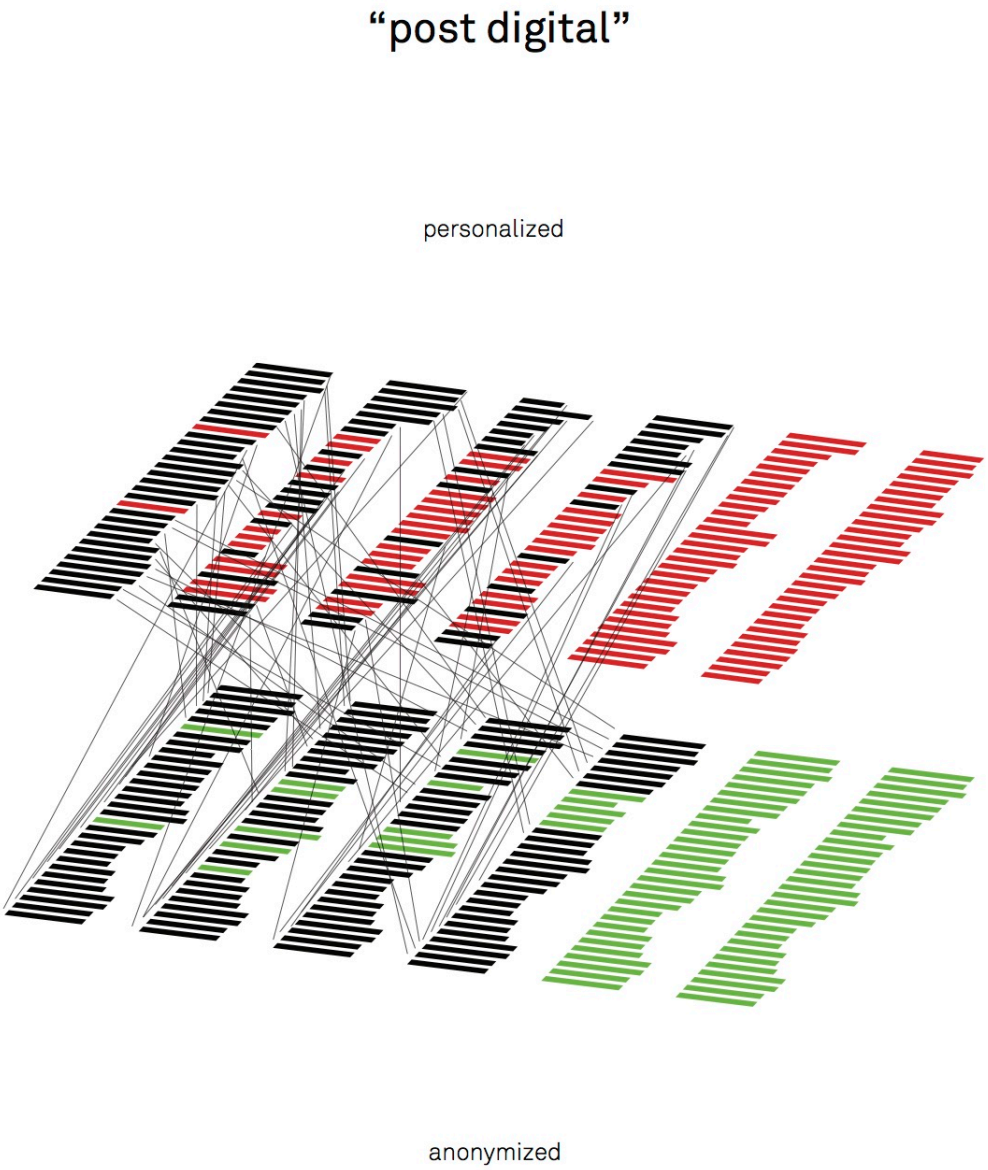
During my course in Hong Kong I also received feedback from colleagues that I needed to build a more specific data set with words that were in some way related, as a 'controlled vocabulary'. Therefore I used the first word of *Keywords* by Raymond Williams, 'aesthetics', drawing on articles and material from the 'Aesthetics' reading group organised by my supervisor Timon Beyes in my department MPP, at CBS. In January 2015 I started to take screenshots and practiced archiving my data more precisely by saving 'complete' web pages.

Learning from the previous experiments with my system of data capture (1st,10th, 20th webpage etc.) and screenshots, I then developed a method to produce data visualisations focussing on more specific 'terms of art' such as the word 'post digital'. A colleague put me in touch with interactive designer Richard Vijgen, with whom I discussed potential ways to image my research and these initial 'test' data visualisations included here are speculative.[229]



Comparison of imagined 'personalized' and 'anonymized' search results with keyword 'postdigital'. Concept: Renée Ridgway. Data visualisation: Richard Vijgen.

[229] Thanks to Florian Cramer for this introduction. My search results included his text 'What is Post Digital?' APRJA Volume 3, Issue 1, 2014. Available here: http://www.lab404.com/142/cramer.pdf



Comparison of imagined ‘personalized’ and ‘anonymized’ search results with keyword ‘post digital’. These visualisations of the keyword ‘post digital’ show the imagined difference in ranking of the search results using Google Search and the Tor browser, along with ‘unique’ results represented by red and green. This ‘small data’ test speculates that the more results obtained, the larger the amount of difference between the two browsers.
Concept: Renée Ridgway. Data visualisation: Richard Vijgen.

Appendix D: *Re:search - Terms of Art*

Keywords are still an inherent part of the index systems of books, which in turn formed the physical libraries of the past and, in the digital age, search engines. Raymond Williams' seminal book *Keywords* (1976) described keywords as the 'network of particular, culturally important words that also shift in time'. These vocabularies are conscious, non-neutral and critical choices that engage with the historical, social and cultural discussion within a particular era. Williams chose 109 alphabetized keywords, starting with Aesthetics (A) and ending with Work (W), keenly aware of how these words are used in social situations or embedded in particular networks, with historical shifts of meaning.[230] He states in the introduction:

> This is not a neutral review of meanings. It is an exploration of the vocabulary of a crucial area of social and cultural discussion, which has been inherited within precise historical and social conditions and which has to be made at once conscious and critical––subject to change as well as to continuity (Williams 1976:XXXV).

The alphabetical list, or 'controlled vocabulary', was a method for the reader to find 'other kinds of connections' and make other comparisons (ibid:25). Rooted in his coinage 'cultural materialism,' his book was seminal to the field of cultural studies and contemporary visual studies and involved a reinterpretation or reinscription of historical documents to create a new 'zeitgeist'.[231] This type of literary criticism could be defined as a theoretical method, involving numerous actors, which are influenced by and influence the power constructs involved with these human chosen taxonomies. During the 1980s, alongside keywords, the rise of visual culture (semiotics) and its relationship to the textual corpus shaped vocabularies and discourse. Besides language and communication, comprising the belief in cultural materialism was the specific practice and role that artists played in society.

Nowadays with Google search, certain keywords are 'trending', words and phrases that 'spike' at certain moments and are captured on 'Google Trends' or are the most popular (30%) but these are often the most mundane. It is the less popular keywords that comprise the long tail (70%), which potentially offer the most 'interesting' (to some) or 'relevant' (to others) results.[232] Instead of using 'trending' search terms, I decided to create a dataset of many 'keywords' or terminologies in contemporary art and new media, which I felt at that time (2014-2015) had a sort of 'epistemological gain' (Graw 2015). *Re:search - Terms of Art* is a dataset of keywords that appear in bold throughout this thesis: **accelerationism, aesthetic turn, anthropocene, artistic research, contemporaneity, creative industries, cultural entrepreneurship, new aesthetic, object oriented ontology, performativity, postdigital, posthumanism, postinternet, postmedia and transmedia**.

---

[230] Take the first word of William's book, 'aesthetic' and how it was defined then in comparison to the search results it delivers today. Different meanings then have 'no one thing in common' and there is no determining element that is able to define what the aesthetic is (Welsh: 1996). It is this subjectivity that perhaps escapes us but does not escape search algorithms.

[231] By not only focussing on canonical texts and their vocabularies, the textual analysis often referred to that which was marginalised, or overlooked.

[232] The list of the 100 most used search terms (without branding and porn) in US in 2016 features: 1. 'Powerball' 2. 'Prince' 3. 'Hurricane Matthew' 4. 'Pokémon Go' 5. 'Slither.io' 6. 'Olympics' 7. 'David Bowie' 8. 'Trump' 9. Election 10. Hillary Clinton: Available at: http://time.com/4598647/most-popular-google-search-2016/.

Appendix E: Experiment in Living

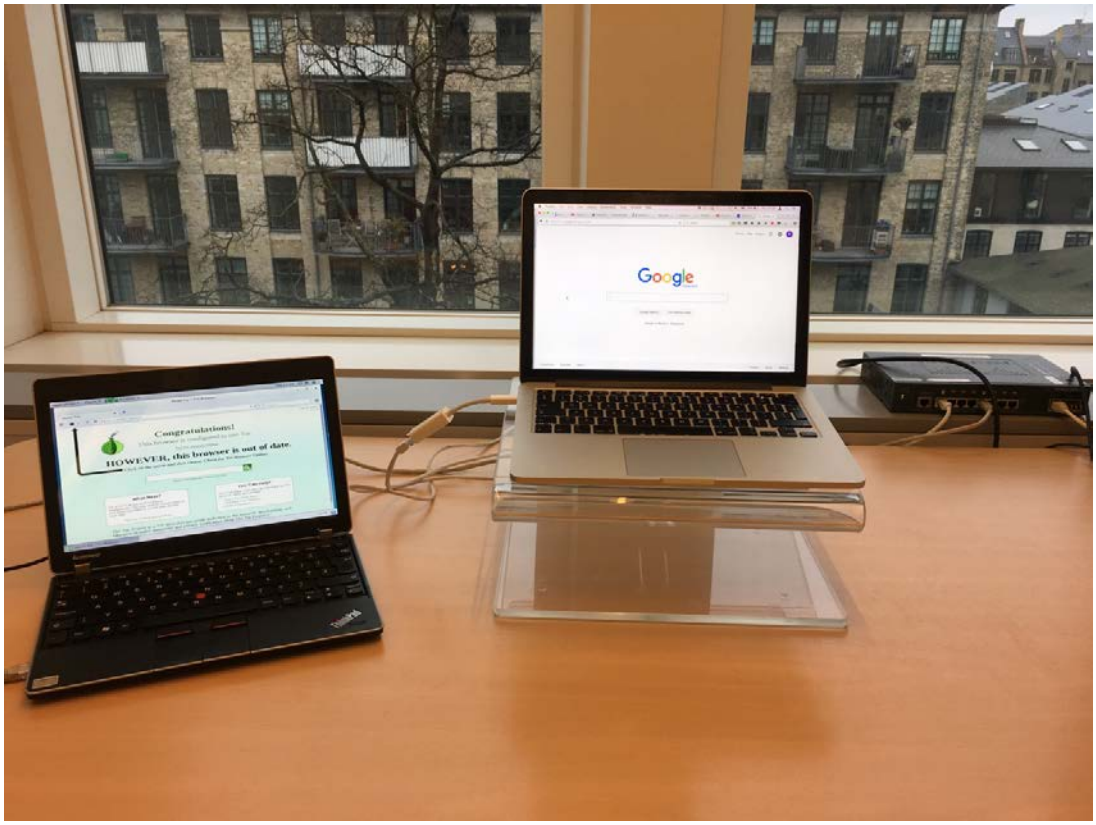
My first tests and speculative visualisations (Appendix C) imagined the difference in the search results, but now I desired to see the actual effects and if there was a difference and, potentially, to discover what caused it. I wanted to find out what search results are on the 99th page of Google? Does the 99th page even exist, even for Tor? Moreover, I wanted to invent a method instead of using those promoted by researchers in the social sciences —searching corporate databases with keywords for articles, papers, chapters and books written and published by other academics that are ranked by bibliometrics (Chapter 3). After determining my choice of keywords, *Re:search - Terms of Art* (Appendix D) I was curious to see how search engines would interpret my keywords—if a combination of terms or phrases comprised of two words, for example, would deliver diverse results. I also questioned whether I had to place these two-word terms in 'quotes' as with Boolean search and I wasn't convinced that 'single terms are unlikely to be particularly successful in returning useful results' (Halavais:2009: 38).[233]

For my 'experiment in living' I chose Google Search using the Firefox browser and Tor, which is preconfigured with the Firefox browser. My research was carried out using chosen keywords on two computers, a 'personalised' Apple using an OS 10 Yosemite operating system, where Google applies its algorithms to offer relevance and recommendations and a PC for Tor (Appendix B). The 'hacker approved' PC was a Compaq, Presario, model CQ60- 420ED, serial number: CE935G0FM. The operating system is Debian 3.20-4 and the processor is Amd64. Debian is a version of the Linux operating system. Obtaining a so-called 'clean' computer was a concerted effort, yet according to the hackers who set it up for me, there is no chance of a 'backdoor' as it has been taken apart, checked and now recycled for this experiment. The only thing that I installed on this computer was a Tor: 3.5.2. The Tor Browser ostensibly delivers 'anonymised' search results with a hidden IP address (which I discuss in detail in Chapter 6). Another given is that the Tor browser has no history—I shut it down after every search — starting anew with a keyword every time as a tabula rasa. To the best of my knowledge, I was signed into my Google account (my email was on their servers) and I used default settings with both Google Search and the Tor Browser.

In order to carry out the study in a secure and coordinated manner, I received permission from the technical service department at CBS (Copenhagen Business School) who installed a router inside my office with multiple ports. Each computer connected to an 8 port router by a cable that in turn was connected to the CBS internet, allowing almost simultaneous querying, or at least within the same time frame. According to Morten Sune Nielsen from technical services, CBS filters all incoming requests with its firewall yet almost everything is allowed in and outgoing traffic is not filtered. I received an assigned IP address from CBS (they are static and have around 10 IP addresses for the whole institution) and it was highly likely that I used the same IP address during my experiments. As I was not directly connected to the internet but rather via CBS routing, it did not matter if my IP address was static or dynamic.

---

[233] 'For a multi-word search, the situation is more complicated. Now multiple hit lists must be scanned through at once so that hits occurring close together in a document are weighted higher than hits occurring far apart. The hits from the multiple hit lists are matched up so that nearby hits are matched together. For every matched set of hits, a proximity is computed. The proximity is based on how far apart the hits are in the document (or anchor) but is classified into 10 different value 'bins' ranging from a phrase match to 'not even close'. Counts are computed not only for every type of hit but for every type and proximity. Every type and proximity pair has a type-prox-weight. The counts are converted into count-weights and we take the dot product of the count-weights and the type-prox-weights to compute an IR score' (Brin and Page 1998:109).



PC on the left, running Tor and my personalised Apple, with Google Search on the right.

I decided to search with the same keywords, same router, same internet connection with cables and same time stamp (same hour), on two different computers and two different browsers.[234] I began searching with my 'Terms of Art' in my office at CBS and I decided not to put words in quotation marks. My rational was that *most* people do not search this way (for some 'librarians and archivists' or 'techies', Boolean search is considered the more professional way to search). I gathered data by capturing the 1st page of results, along with the 10th, 20th, 30th, 40th and 50th pages for the data set. I saved these as 'complete web pages' and clicked through the page numbers at the bottom, continuing to the next page of results.

I also obtained 'ads' in my results from Google that I discuss in Chapter 5 and from Disconnect Search that skewed the order of how many results were delivered per page (even though I had set them both at 10 returns per page).[235] Google always delivered to the default of 10 SERPS per page.

---

[234] 'The theory behind using the comparison model of Apple vs. PC is that the two systems still differ. With Apple you are connected all the time and part of it. Windows (Microsoft) is more of closed system, it doesn't have an open root for intellectual property reasons. Although Windows is going the direction of Apple, there are more possibilities to close yourself off and Windows still has these options. Linux has a root as an operating system and is in principle not connecting as it is very complicated to hook up external drivers. Apple and Windows are easier because they are developed in a corporate framework. Every computer (or device) has what's called a MAC address (media access control address) that is fixed; even the internal parts have one: board, screen and printers. This is so that all gadgets are able to communicate with each other. For this empirical study, the Anonymous Hacker could have built in a MAC converter but the question would arise whether it would really stay fixed, because the IP address could shift' (Anonymous Hacker 2014).

[235] With Tor and 'Postmedia' I received search.disconnect.me/ads' as the first and second result in most of the 40 pages of returns.

With most of my keywords I received a message from Google around the 35-39th page stating: *In order to show you the most relevant results, we have omitted some entries very similar to the 200 already displayed. If you like, you can repeat the search with the omitted results included.* I continued to collect data, yet sometimes the 'next' page would disappear when I clicked on the link at the bottom of the page and that would be the end of that keyword data set.

This 'experiment in living' facilitated interaction with search algorithms 'in the wild' and I simultaneously collected data on Google and Tor. Although my [input] data was only keywords. in this exchange system I gave my data through a software interface and received free services––search results [output] as URLs from Google Search and the Tor Browser in return. As I engaged with these 'truth games' (Foucault 1994) I realised I was constantly being gamed by the algorithms as much as I attempted to game them.

Appendix F: Data visualisation as transcription

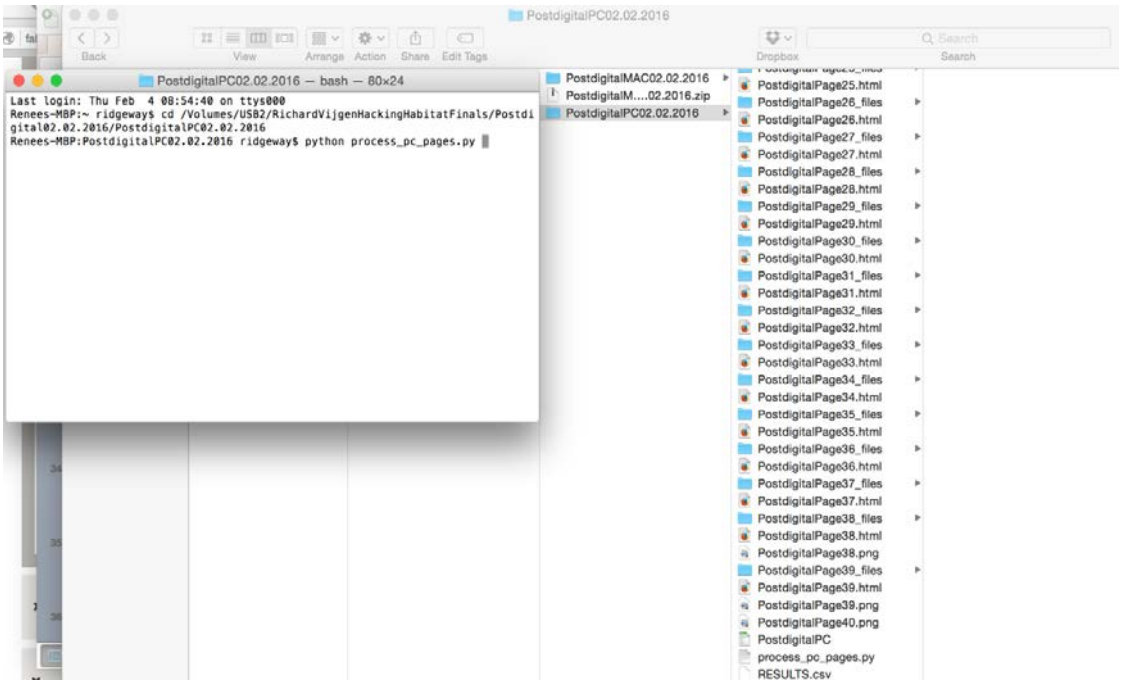
After saving 'complete web pages' I first had to extract the URLs manually from the saved webpages. For one word this wasn't much labour but to do an entire dataset would have taken me weeks. My Leuphana colleague, Marcell Mars, wrote a Python script that extrapolated the URLs (Uniform Resource Locator) from the Google Search and Tor Browser results. By using 'terminal' with UNIX code on my Apple I generated a .csv (comma-separated value) file for each dataset and was then able to 'code' the data.

```
import glob
import lxml.html
import csv
import time
import os


def getLinks(page):
    links = []
    for link in page.xpath('//h3/a/@href'):
        links.append(link)
    return links


def getPage(f):
    return lxml.html.fromstring(open(f).read())


def getTime(f):
    return time.strftime('%Y-%m-%d %H:%M:%S', time.gmtime(os.path.getmtime(f)))

files = glob.glob("*.html")
csv_f = open("RESULTS.csv", "wt")
csv_w = csv.writer(csv_f)
csv_w.writerow(("search term", "page", "rank", "link", "date"))


for f in files:
    print("file: {}".format(f))
    from_f = f.split(" - ")
    search_term = from_f[0]
    page_number = from_f[1].split("Google SearchPage")[1][:-5]
    rank = 1
    for link in getLinks(getPage(f)):
        print("rank: {}; link: {}".format(rank, link))
        csv_w.writerow((search_term, page_number, rank, link, getTime(f)))
        rank += 1
```

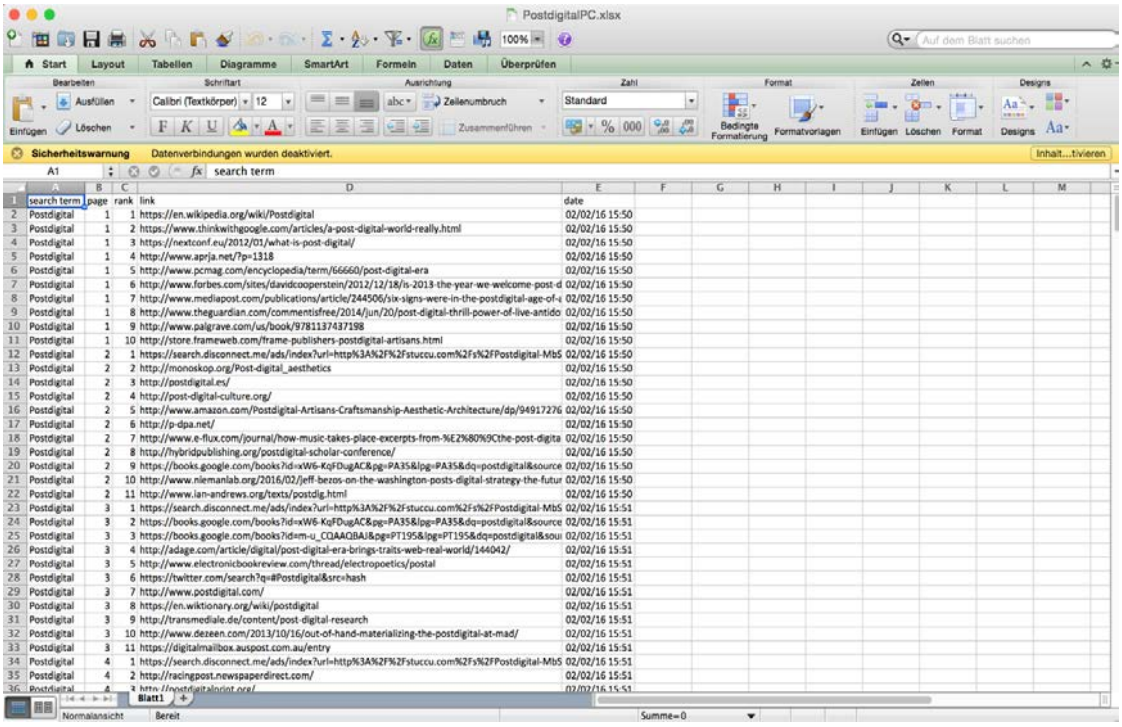
Image caption: Marcell's Python code

I opened the .csv file in Excel and made sure they were structured properly (I had to tweak some of them). I then gave these Excel files of the extracted URLs to interactive designer Richard Vijgen, who describes the visualisation workflow as follows:

> The search results are first converted from the spreadsheet to a comma separated value (.csv) format. Then for each search query, two .csv files (one for Mac, one for PC) are loaded into the software programme 'Processing' (processing.org). Processing loops through the first list (Google) and for each result it loops through the second list (Tor) to check if a match exists and if so, at which position. If a connection can be made, a line is drawn from the position of the result in the first list (Google) to the position of the same result in the second list (Tor). While the URLs are shortened in the print version for legibility reasons, it is the full URL that is being tested for matches.

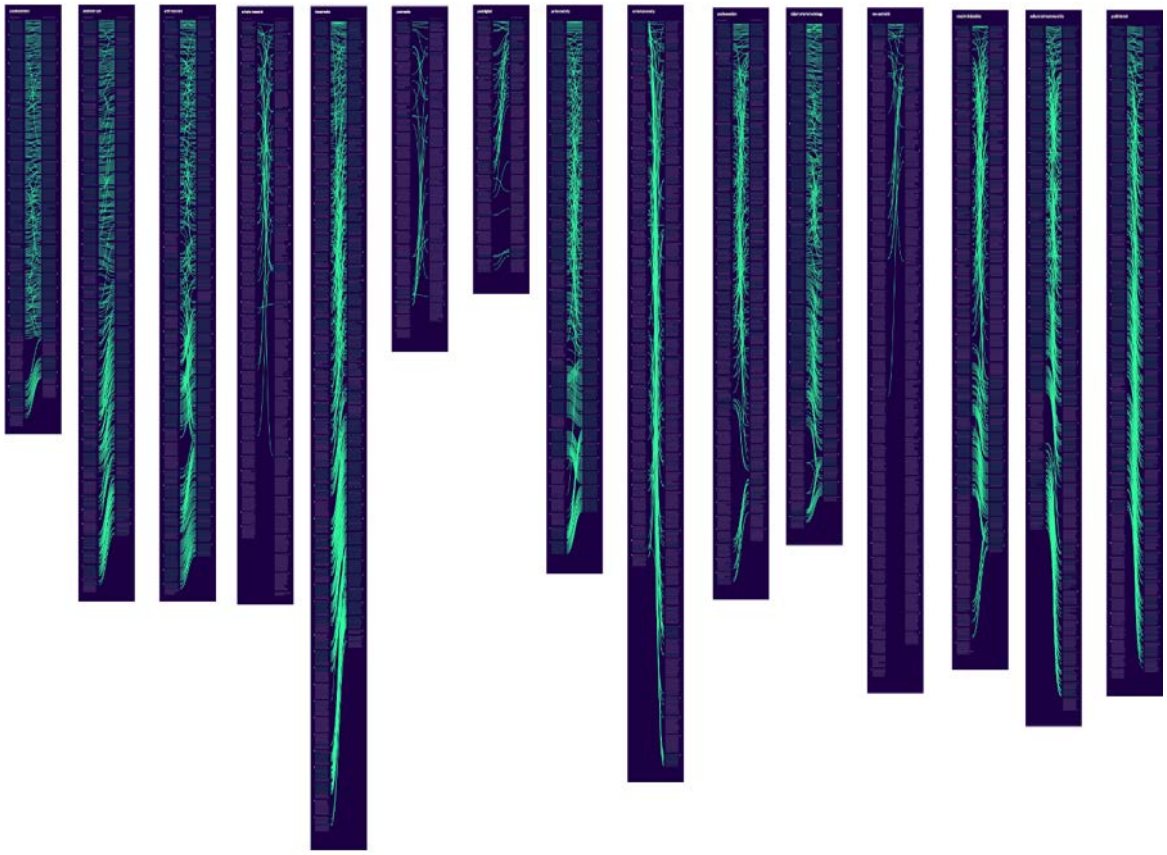


Terminal produces RESULTS.csv file for the word ‘postdigital’



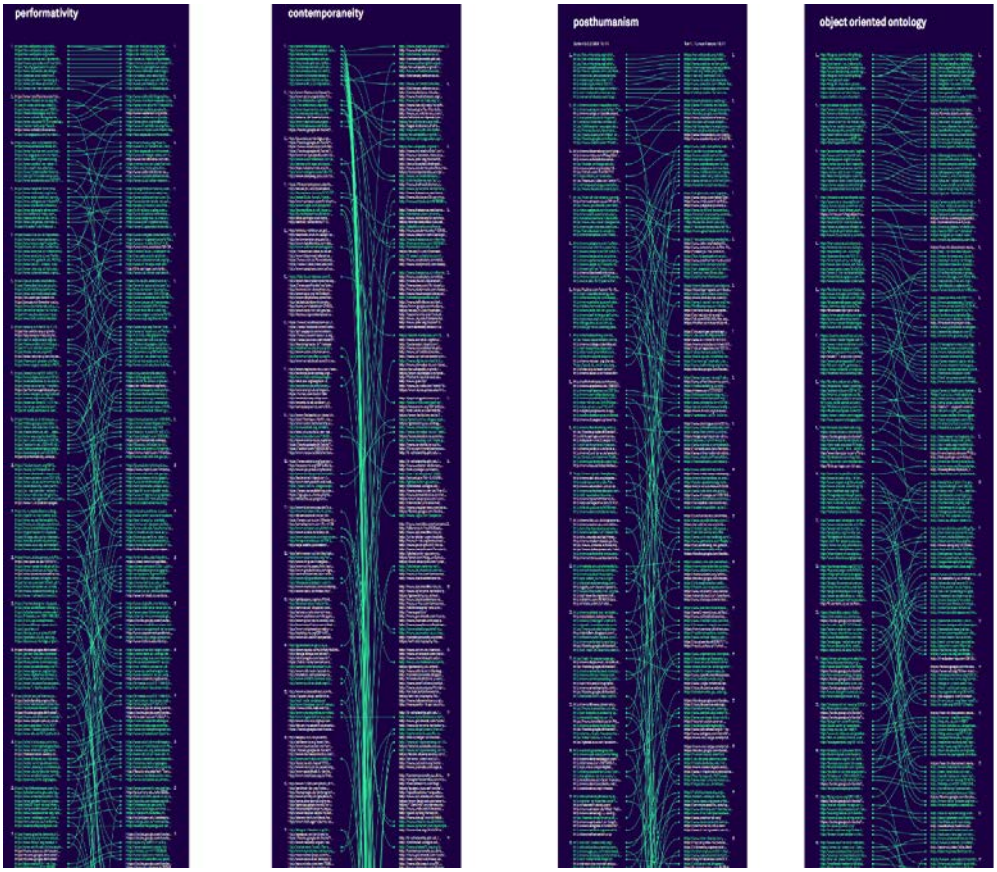
Imported links from ‘postdigital’ into Excel

The following data visualisations show the difference between personalised (left column) and anonymised (right column) search results. Although both claim that they return millions of results, in 2015 Google and Tor returned around 90 pages of results (Appendix C), later on they actually returned around 35 pages and I received less search results. Connecting the results with a green line reveals a ‘fingerprint’ of the invisible personalisation algorithm in each dataset. Some results are the same yet their position in the ranking (on which page of the results and where), differ. The white hyperlinks are unique results for both types of search.



Overview of 15 keyword dataset. Comparison of Google Search ‘personalised’ and Tor Browser ‘anonymised’ search results. Green represents ‘identical’ URLs. White represents ‘unique’ URLs. Design: Richard Vijgen

Re:search - Terms of Art (2016) overview Hacking Habitat exhibition, Utrecht. Photo: Thomas Lenden

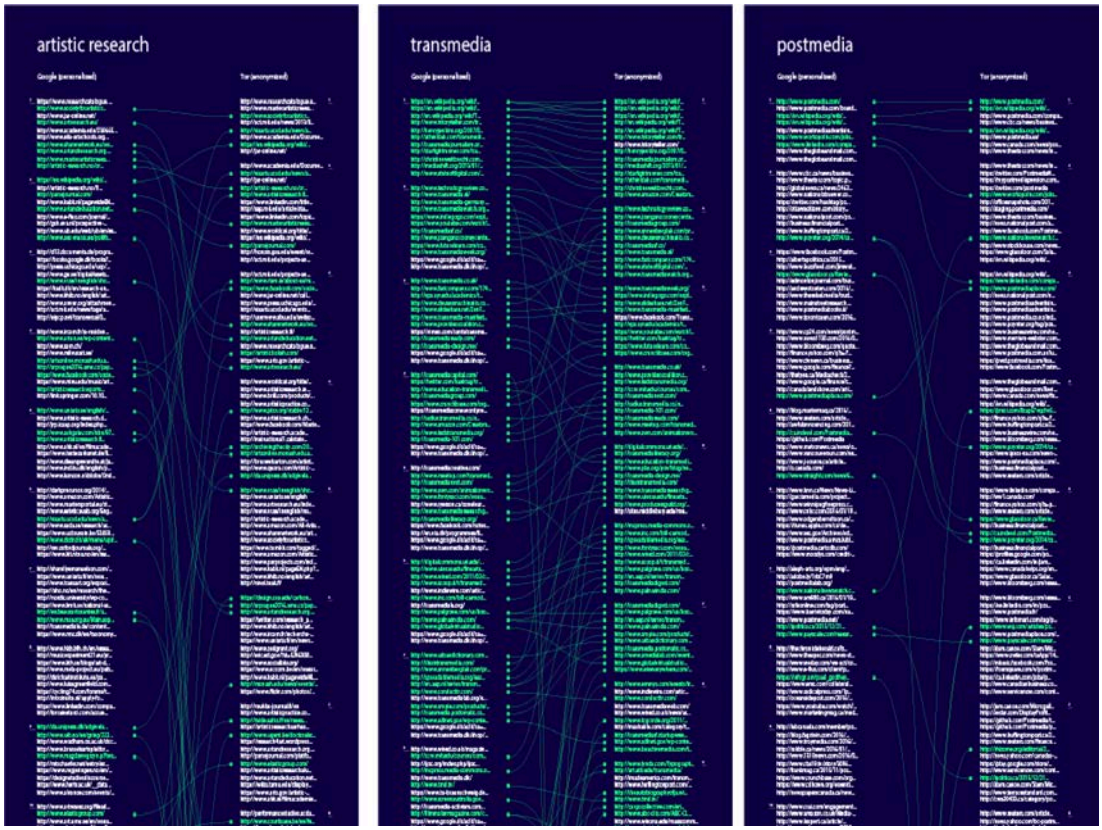


Comparison of Google Search ‘personalised’ and Tor Browser ‘anonymised’ search results

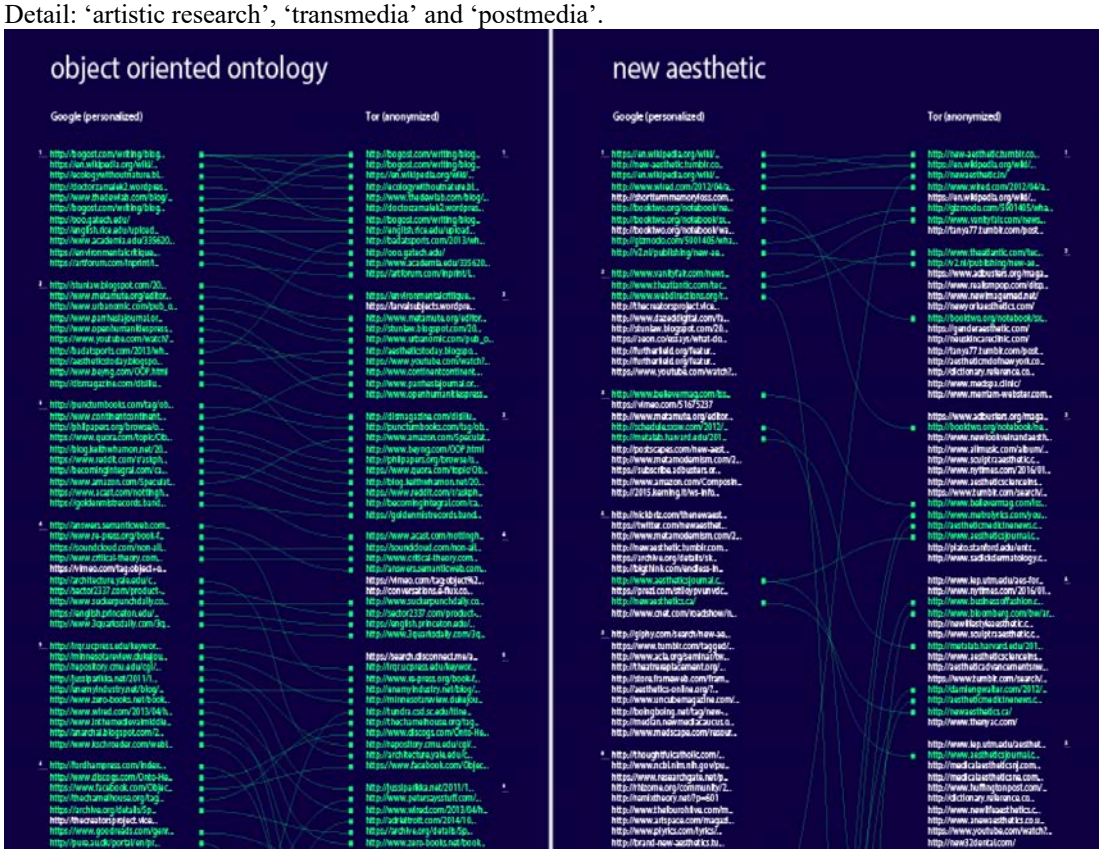
Panopticon Prayer Wheel Design: Richard Vijgen Photo: Thomas Lenden

The iPad displays an interactive visualization where personalised and anonymised results are on the opposing sides of a ‘search panopticon’. Users rotate through different search queries like a ‘prayer wheel’ and examine individual results by swiping up and down.

Specific Results:

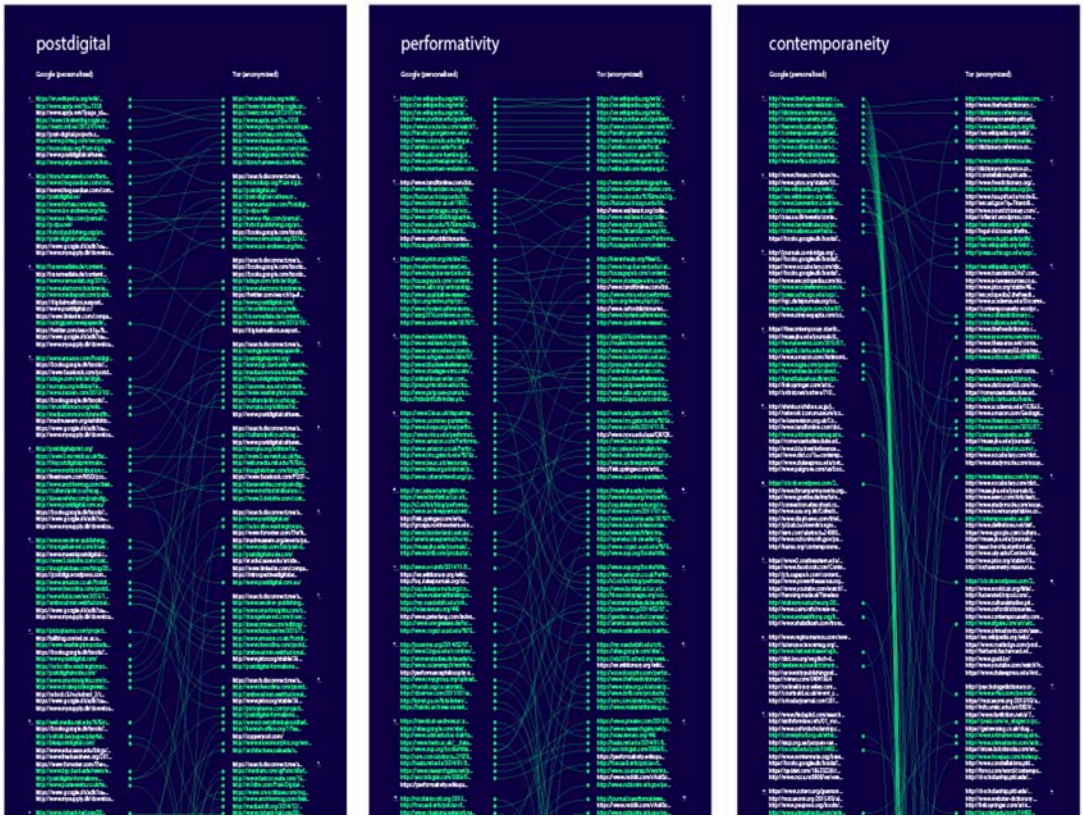


Detail: ‘artistic research’, ‘transmedia’ and ‘postmedia’.

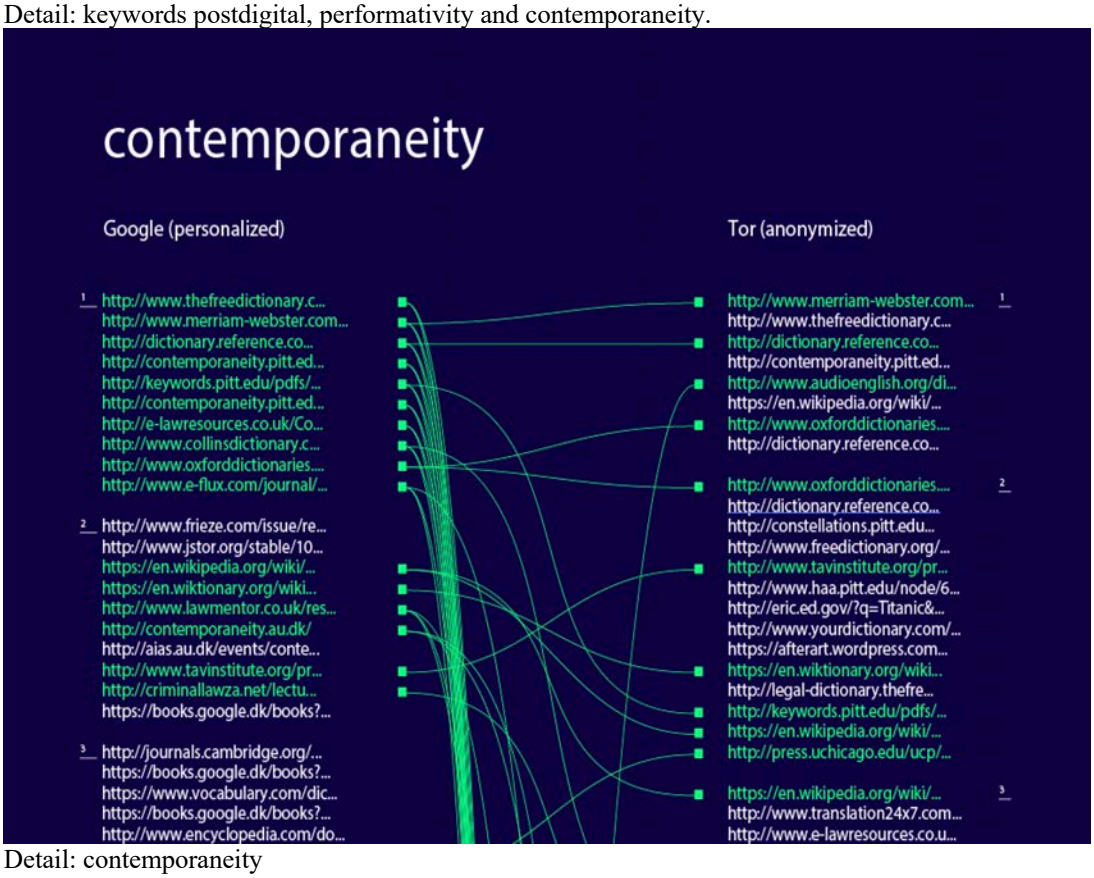


Detail of ‘object oriented ontology’ and ‘new aesthetic’.

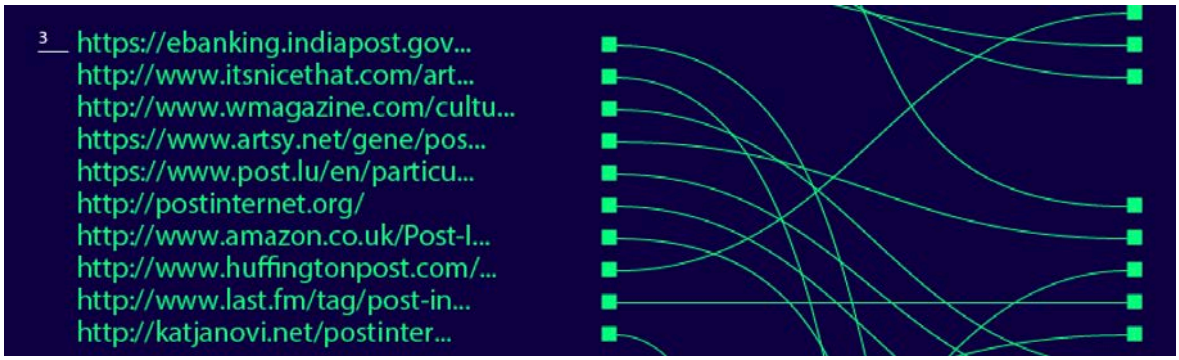
Specific Results:



Detail: keywords postdigital, performativity and contemporaneity.



Detail: contemporaneity



Keyword 'postinternet'. Google results page 3.

Some of the results mentioned in Chapter 5 might reflect my personalisation through my locative data and search history even though I was searching and collecting the data in Denmark. With the keyword 'postinternet' I received Google Search results not only for art related links but for Post departments from places I recently had visited, such as India where I received a link on the third page, top result for e-banking at the Indian Ministry of Communications, Government of Posts.[236] On page 39 I also received a link for Hong Kong Post, where I attended a PhD course in 2014.[237]

Keyword: postinternet. Google results page 39.

Keyword 'postinternet'. Google results page 33.

I also received postal services for 'rijssen-holten.nl' as the 33rd result (https://www.rijssen-holten.nl/loket/subthema/post-internet-en-telefonie) along with digital services for 'Breda'.[238]

[236] https://ebanking.indiapost.gov.in
[237] http://www.hongkongpost.hk/en/about_us/whats_new/press_release/index_id_9.html
[238] http://www.breda.nl/digitale-balie/producten-en-diensten/post-internet-telefonie

Keyword: postinternet. Google results page 37 and 38.

On the 38th page, I received results for Maastricht[239] and on the 39th page results for Diemen (9th result).[240] These are all locations in the Netherlands, where I also lived. I also received results from Danish postal services on the 30th page[241] and on the 31st page.[242]

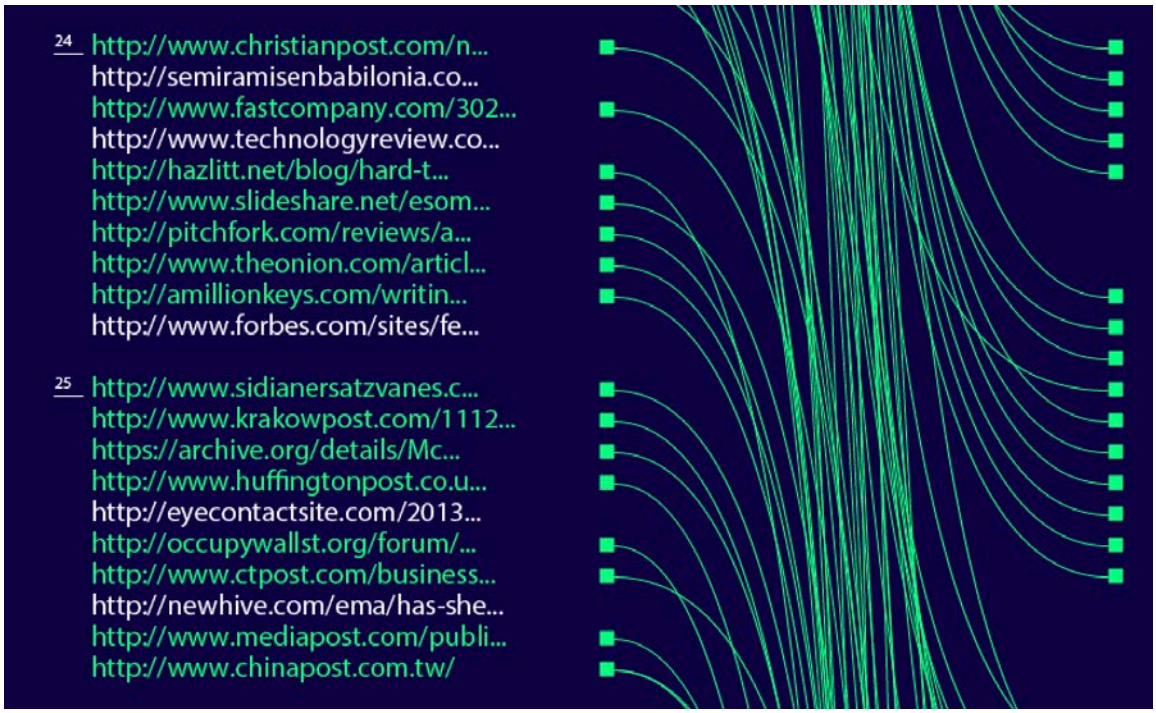
Keyword 'postinternet'. Google results page 30-31.

As I explain in Appendix E my 'keywords' were not 'trending'. Instead I used terms from contemporary art with the prefix 'post' ('postdigital', 'postmedia', 'postinternet' and 'posthumanism') and the algorithmic interpretation of 'post' varies (postal services, 'post' as a common naming convention for newspapers, as a widely used contraction of 'posting', as in blogging or commentary and as a term for histories and intellectual movements). There are newspapers with 'post' in the title, such as on page 19, bangkokpost.com.

[239] http://www.gemeentemaastricht.nl/themas/thema/post-internet-en-telefonie/
[240] http://www.diemen.nl/inwoners/digitaalloket/themas/?tx_ncgovpdcx_pi%5BtioTheme%5D=post-internet-en-
[241] https://www.rm.dk/siteassets/om-os/organisation/hr-afdelingen/med-aftale/2007.06.25-26-rmu/pkt.10g_forslag-e-post.pdf
[242] http://docplayer.dk/4787153-Dokumenthaandtering-e-post-internet-virus-lagringsmedier-brugerrettigheder-sikkerhed-i-birkeroed-kommune.html

Keyword ‘postinternet’. Google results page 19.

Or online media such as the christianpost.com on page 24 as the 1st result, and on page 25, the krakowpost.com is the second, the huffingtonpost.com is the 4th result and the 10th result is chinapost.com.



Keyword ‘postinternet’. Google results page 24-25.

Keyword ‘postinternet’. Google results page 25.

Appendix G: Black Box

Explorations of the black box originate in the field of cybernetics, derived from the Greek word for 'steersman' and coined by Norbert Wiener in 1947 'to designate what he hoped would be a new science of control mechanisms in which the exchange of information would play a central role' (Galison 1994:232).[243] It began with a 'thought experiment' regarding the Second Law of Thermodynamics by James Clerk Maxwell, otherwise known as 'Maxwell's demon' (Glanville 2003). This so-called demon is interjected into a 'closed box' and creates an input and an output, which in turn is observed.

> This way, within first-order cybernetics the observed object is interpreted as a black box that does not disclose its mode of operation but can be singled out and described in terms of its inputs and outputs and, hence, in terms of certain regularities of conduct (Glanville 1988 cited by Beyes 2005:449).[244]

In his *Introduction to Cybernetics* (1956) the early cybernetic pioneer W. Ross Ashby described the black box as a concept containing 'a presumed mechanism, which cannot be seen and is the product of the observer's interaction with the whatever-it-is' and that this obscurity is universal (Glanville 2003). Although it cannot be opened, as it is not a real object, it can be modelled.[245] Predominantly following first-order cybernetics this process reflects the need to 'whiten' the black boxes of organisation and management processes in order to understand results (Beyes 2005:455).

> Science's task, then, is to "whiten" the black box, i.e. observe inputs and outputs and come up with descriptions that enable it to predict future states, thus allowing for control and steering (ibid:449).

'Black boxes, as Wiener used the term, meant a unit designed to perform a function before one knew how it functioned; white boxes designated that one also specified the inner mechanism' (Galison 1994:246-7). In 1961 Norbert Wiener described an 'unknown system' regarding the 'operation' as a 'black box' for the production of war-fighting machines and his AA predictor incorporated 'self-regulating feedback' which was 'based its algorithm for prediction on statistical input from the pilot's past performance, the device was a kind of learning machine' (Galison 1994:238). Wiener's application of the black box for testing out systems of identification and prediction was ground breaking, along with the ability to reproduce the output behaviour of black boxes, which he defined as 'self-organising'.[246] In this way, computation enabled the continuous feedback loop of human-computer interaction:

> As Norbert Wiener suspected as early as 1948, the advent of digital computers—along with concepts such as feedback, self-regulating systems, and prediction—had initiated a

[243] Κυβερνητική – kybernetes metaphoricially, governor, pilot. 'The construction is perhaps based on the 1830s French cybernetique 'the art of governing.' Available here: https://www.etymonline.com/word/cybernetics.
[244] '"[F]irst-order cybernetics holds the promise of finally discovering, fine-tuning and inventing anew the mechanisms of governance in body, mind and society" (Baecker, 2003) – and, one should add here, in organizations' (Beyes 2005:449).
[245] Ashby actually built one, called by Heinz Von Foerster an 'Ashby Box' that had two switches and two lights, each either on or off.
[246] 'In this language, the more sophisticated feedback mechanism of the AA predictor opened a new universe of black boxes to the engineer and to the philosopher" (Galison 1994:246).

> fundamental rearrangement of temporal structures (Wiener 1961:60-94 cited by Beyes and Pias 2016:3).

However, Heinz von Foerster pointed out the 'blind spot' of first order cybernetics, namely that the observer is included in the interaction with the observed system. The 'Terms of the first order are "ontologically" based on the assumption of looking objectively at a given black box, trying to decode its modus operandum' (Beyes 2005:449). Yet the observer observes the interaction.[247] In this way the experimenter interacts with the black box and in turn is affected by it. As Beyes notes, Von Foerster's 'second-order cybernetics' has epistemologically 'far-reaching consequences', where if the observer views the black box as a black box, one ends up 'in a loop that connects the observer to the respective object that is under inspection' (von Foerster and Poerksen, 2002 cited by ibid:450). With Von Foerster's second-order cybernetics objective steering and control are disrupted by the paradoxical 'self-made' observations of the observer. In this way the

> Black Box and the observer act together to constitute a (new)whole, which itself is a Black Box to its observer, and so on. These observers might be the same, transcending boundaries: which may be a source of human consciousness (Glanville 2003).

Besides producing military weapons used for war, Wiener, in collaboration with others, designed models or simulations describing the workings of the human brain. 'Black-box engineering now had something more complex than electrical amplification as its functional goal: to re-create the mind itself '(Galison 1994:246). The 1943 article *Behavior, Purpose and Teleology* by Rosenblueth, Bigelow and Wiener, contemplated the relationship between living organisms and machines (ibid:245). This furthermore blurred the distinction between humans and machines, influencing other disciplines thereby 'suggesting that all functions of the brain might be duplicated by electrical systems "very attractive"'(ibid:247). Merging research, engineering and neurology these early computational machines became cybernetic feedback systems of control. The input-output analysis even went so far as to emphasise that 'as objects of scientific enquiry, humans do not differ from machines' (ibid:251).

Recursion also plays a role in these human/machine experiences with repetitive behaviour 'whitened'—each black box consists potentially of black boxes and observers. However, this type of behaviour is not necessarily repetitive—it is reciprocal: 'Finally, following the Law of Mutual Reciprocity, since the Box is Black to the observer, the observer may be Black to the Box' (Granville 2003). The ways in which the observer (user) interacts with non-human agents is observed by the algorithm as well. Not only did the observed observe the black box, but humans could be observed as a black box, in other words as something non-transparant or invisible by machines. Galison writes in his text *The Ontology of the Enemy: Norbert Wiener and the Cybernetic Vision* (1994) that cybernetic philosophy includes the hidden, the non-transparent, and is 'premised on the opacity of the Other'—in this case the algorithm. Thus, the machine could also be deemed as the other. Yet 'the black box that constitutes the process of perception ….vanishes after observing, evoking unstableness and a continuous process of becoming' (von Foerster and Poerksen 2002; von Foerster 1993a cited by Beyes 2005:450). Thus, in their process of becoming, the search algorithms of the black box in turn observe the observer observing, reflecting the nature of a continuous feedback ecosystem in flux.

[247] 'Yet, as Luhmann markedly pointed out, there is also "the observer who cannot observe his observing"' (Luhmann 2000 cited by Beyes 2005:449).

Appendix H: Not just routing information, but content

Deborah Buckner's 2008 article, *Internet Search and Seizure in United States v. Forrester: New Problems in the New Age of Pen Registers* brings to the forefront contentious terminologies of 'communicative content' otherwise known as 'routing information' under case law. The Ninth Circuit court in *United States v. Forrester* upheld the Supreme Court precedent *Smith v. Maryland,* which decided that individuals do not have a 'reasonable expectation of privacy' and that IP addresses (and email) are just routing information, not content. Buckner's spells out how this newer ruling draws on its precedent, *Smith v. Maryland*,

> where the United States Supreme Court found that traditional Fourth Amendment protections do not apply to telephone routing information (like telephone numbers) because of the caller's lack of a reasonable expectation of privacy in those numbers and because this information does not constitute—content (2008:504).[248]

Moreover, this older ruling is in turn contingent on the assumption that this was not a violation of the Fourth Amendment because people willingly hand over their telephone numbers to a telephone company—the information is voluntarily given to a 'third party'—therefore there is no legitimate expectation of privacy. The employees at phone companies also have access to an individual's information in order to channel their calls (ibid).

The same rational was therefore applied to *United States v. Forrester* in that individuals have no expectation of privacy because they choose to hand over their data (e.g. IP address) in order to have access to the internet via an ISP, which is then responsible for channelling the information correctly through servers and to send emails (ibid:505).[249] This is where the shift from physical search extends to what is called 'electronic surveillance'. Buckner notes that with the case of *United States v. Forrester*, this reasoning was extended to include 'pen registers' surveying the internet.[250] Buckner's argument rests on two points, namely: the 'expectation of privacy' is much greater with IP addresses and emails (unlike telephone numbers) and that they are 'more suggestive of content.'[251] She then elucidates why what 'pen registers' collect nowadays is

---

[248] The dissent of the other judges on the Supreme Court is important to note: 'These dissenting opinions present strong reasoning to rebut the ruling of the majority's opinion in Smith (that communications handed to third parties have no expectation of privacy). Likewise, these same policy reasons, as expressed by the dissent, support the notion that Internet communications, though given to third parties, should retain a reasonable expectation of privacy' (Buckner 2008:510).

[249] Concerning the 1st Amendment rights in the US constitution, the late Aaron Swartz stated:
'Principle I think is one what I think our Founding Fathers would have understood. If the Internet had been around back then instead of putting post offices in the constitution, they would have put ISPs' (Knappenberger 2014) Timecode:1:02:00.

[250] A pen register, or dialled number recorder (DNR), is an electronic device that records all numbers called from a particular telephone line. The term has come to include any device or program that performs similar functions to an original pen register, including programs monitoring Internet communications.
https://en.wikipedia.org/wiki/Pen_register

[251] Buckner's description of content': 'Communications that ISP could capture: Email subject lines could be obtained through the use of Internet pen registers, which would reveal the contents of the communications. Problems would also arise if the user had put a search term into Google or another search engine. The URL derived from the IP address would contain the search words the individual used. This would certainly alert the Government to the content that the user was seeking. The Court also commented that IP addresses themselves would allow the government to determine the home page of the website that the user under surveillance was accessing' (Buckner 2008: 514).

qualitatively different —IP addresses that enable email exchange are not routing information and therefore are not similar to telephone numbers (ibid:502).

Moreover, Buckner points out that with the ruling of *United States v. Maxwell*,

> [c]ourts have held that the content of email communications are protected because there is a reasonable expectation of privacy when those emails are stored with, or sent through, a commercial Internet Service Provider (ISP)(ibid).

Buckner argues that the Ninth Circuit Court in *United States v. Forrester* erred in stating that the public has no expectation of privacy when using ISP providers or sharing private information with corporations. Written after the US Patriot Act (2001) allowed increased governmental discretion, she also points out that pen registers used in telephone communications differ from Internet routing information and she cites the EPIC (Electronic Privacy Information Center) supra note 16:[252]

> The fact that the provision prohibits the capture of 'content' does not adequately take into account the unique nature of information captured electronically, which contains data far more revealing than phone numbers, such as URLs generated while using the Web (which often contain a great deal of information that cannot in any way be analogized to a telephone number) (ibid:513).

Information derived from an IP address differs from an URL—with an IP address, one accesses the home page of a website. However, with an URL, the exact webpage the user accessed would be shown if they were under surveillance (ibid:514). With packet switching the header also contains metadata.[253] In other words, internet communications are much more detailed than telephone numbers and may allow the identification of the user along with content such as email subject lines, home pages and search terms (ibid:517).

[252] EPIC is a public interest research centre in Washington, D.C. EPIC was established in 1994 to focus public attention on emerging privacy and civil liberties issues and to protect privacy, freedom of expression and democratic values in the information age. Available here: https://epic.org/epic/about.html.
[253] Metadata is a larger discussion, only briefly addressed in Chapter 5 but beyond the scope of discussion.

Appendix I: Google's database architecture

Google relies on a polyglot persistence architecture consisting of BigTable, which is leveraged by most of its services such as Google Search, Google Analytics, YouTube, etc.[254] Its exabytes of data are stored across commodity servers with the help of the Google File System (2018).[255] Google initially used MapReduce, a programming model that processes and generates big data sets, for its indexing infrastructure but later transitioned to BigTable during the Caffeine release.[256] Historically, MapReduce was used to regenerate Google's entire index of the World Wide Web but is now used with 'a set of wrappers that allow BigTable to be used both as an input source and as an output target for MapReduce jobs' (Sarawagi 2020).[257] In development since 2004, BigTable is a compressed, proprietary data storage system that manages structured data across thousands of servers and has very different characteristics than a relational database because it relies on a huge (distributed) hash table.[258] BigTable is used by other Google applications, such as web indexing and My Search History that 'maps two arbitrary string values (row key and column key) and timestamp (hence three-dimensional mapping) into an associated arbitrary byte array' (Stackoverflow 2020).[259]

---

[254] YouTube uses MySQL with Vitess, a database clustering system for horizontal scaling of MySQL. Available here: https://vitess.io/

[255] 'The Google Cloud datastore has over 100 applications in production at Google both facing internal and external users. Applications like Gmail, Picasa, Google Calendar, Android Market & AppEngine use Cloud Datastore & Megastore. Google Trends use MillWheel for stream processing' (Stackoverflow 2020). This is from 2006: 'Google runs on hundreds of thousands of servers—by one estimate, in excess of 450,000—racked up in thousands of clusters in dozens of data centers around the world. It has data centers in Dublin, Ireland; in Virginia; and in California, where it just acquired the million-square-foot headquarters it had been leasing. It recently opened a new center in Atlanta, and is currently building two football-field-sized centers in The Dalles, Oregon' (Carr 2006)[…] 'In general, Google has a split personality when it comes to questions about its back-end systems. To the media, its answer is, "Sorry, we don't talk about our infrastructure." Yet, Google engineers crack the door open wider when addressing computer science audiences, such as rooms full of graduate students whom it is interested in recruiting. As a result, sources for this story included technical presentations available from the University of Washington Web site, as well as other technical conference presentations, and papers published by Google's research arm, Google Labs' (Carr 2006).

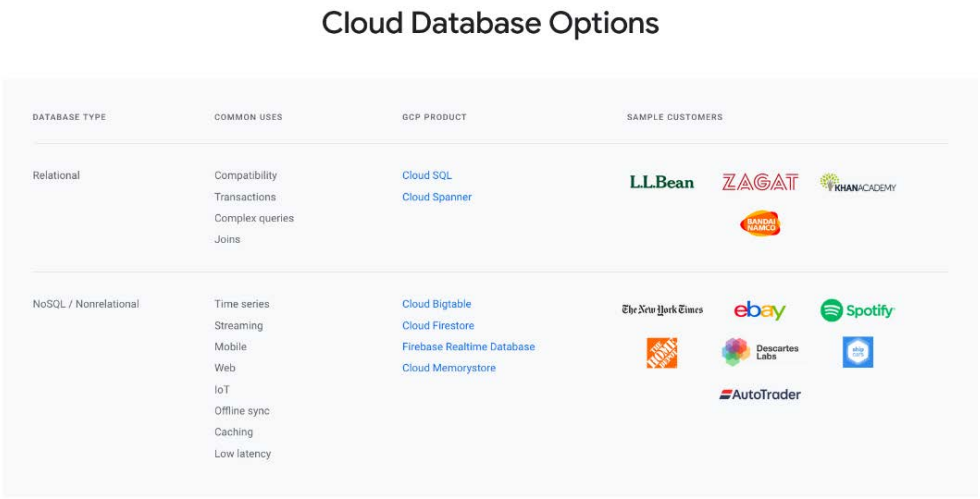
[256] 'It eventually moved to technologies like *Percolator, FlumeJava* & *Millwheel* that provided real-time data streaming features as opposed to batch processing. This enabled the search service to integrate the live search results without rebuilding the entire index' (Sarawagi 2020).

[257] 'Google still uses MapReduce for Google App Engine log analysis and some other use cases' (Sarawagi 2020).

[258] Besides scalability it handles other data demands such as 'high availability & performance whether it is for indexing urls, processing real-time data or latency-sensitive data serving' (Sarawagi 2020).

[259] 'BigTable is not a relational database. It does not support rich SQL-like queries. Each table is a multidimensional sparse map. Tables consist of rows and columns, and each cell has a time stamp. There can be multiple versions of a cell with different time stamps. The time stamp allows for operations such as "select 'n' versions of this Web page" or "delete cells that are older than a specific date/time." In order to manage the huge tables, BigTable splits tables at row boundaries and saves them as tablets. A tablet is around 200 MB, and each machine saves about 100 tablets. This setup allows tablets from a single table to be spread among many servers. It also allows for fine-grained load balancing. If one table is receiving many queries, it can shed other tablets or move the busy table to another machine that is not so busy. Also, if a machine goes down, a tablet may be spread across many other servers so that the performance impact on any given machine is minimal. Tables are stored as immutable SSTables and a tail of logs (one log per machine). When a machine runs out of system memory, it compresses some tablets using Google proprietary compression techniques (BMDiff and Zippy). Minor compactions involve only a few tablets, while major compactions involve the whole table system and recover hard-disk space. The locations of BigTable tablets are stored in cells. The lookup of any particular tablet is handled by a three-tiered system. The clients get a point to a META0 table, of which there is only one. The META0 table keeps track of many META1 tablets that contain the locations of the tablets being looked up. Both META0 and META1 make heavy use of pre-fetching and caching to minimize bottlenecks in the system' (Stackoverflow 2020).

Google's personalized search service enables users to browse their search history to revisit old queries and clicks. Users can ask the service for personalized search results based on their usage patterns. All the user data for the personalized search results gets stored in BigTable. The data is replicated across BigTable clusters to make it highly available and to reduce the latency (Sarawagi 2020).[260]



Google's Cloud Database Options. Source: Google.com

Personalised results are connected to advertising through 'a scalable, globally distributed database (RDBMS) designed, built and deployed at Google that 'shards data across many sets'[…] in 'datacenters spread all over the world' (Corbett et al. 2012:1).[261] According to the community wiki of Stackoverflow, Google Ads initially used MySQL and later migrated to F1 DB—a custom written distributed relational database built on top of Spanner, which updates billions of rows per day with parallel operations (2020). 'Spanner started being experimentally evaluated under production workloads in early 2011, as part of a rewrite of Google's advertising backend, called F1' (Shute et al. 2012 cited by Corbett et al. 2012:17).[262] The F1 database has replicas provided by Megastore[263] where updates and searches participate in Megastore's 'transactions and multiversion concurrency' (Baker et al. 2011:228).[264]

[260] 'Three factors play into the overall growth in automation. First of all, search engine technology has grown and become better in sifting through large amounts of structured and unstructured data, especially since Google introduced tools such as MapReduce, BigTable in the mid-2000s and since new Open Source Software for data mining in unstructured information collections such as Hadoop became available. Structured data is for example held in tables, where each column contains a certain kind of information (e.g. a date, the weather condition, a color) and each row represents a record. In contrary, with unstructured data you can never say where certain information may be held nor if it is there at all […] meaning is created by comparing it to other, structured data and through ranking algorithms. Unstructured information machine searching has become better' (Hunger 2017).

[261] 'Spanner is designed to scale up to millions of machines across hundreds of datacenters and trillions of database rows […] automatically reshards data across machines […] and it automatically migrates data across machines (even across datacenters) to balance load and in response to failures' (Corbett et al. 2012:1-2).

[262] 'This backend was originally based on a MySQL database that was manually sharded many ways' (Corbett et al. 2012:17).

[263] 'Megastore blends the scalability of a NoSQL datastore with the convenience of a traditional RDBMS in a novel way, and provides both strong consistency guarantees and high availability […] This partitioning allows us to synchronously replicate each write across a wide area network with reasonable latency and support seamless failover between datacenters' (Baker et al. 2011:228).

[264] 'A full-text index declared in a Megastore schema can index a table's text or other application-generated attributes' (Baker et al. 2011:228).

List of Figures

References

A

Agre, Philip. E. 1994. “Surveillance and capture: Two models of privacy.” *The Information Society* 10(2): 101-127.

Alaimo, Cristina and Jannis Kallinikos. 2019. “Recommender System.” Chapter in *Oxford Handbook of Media, Organisation and Technology*, edited by Timon Beyes, Robin Holt and Claus Pias. Oxford: Oxford University Press.

Alexin, Zoltán. 2014. “Does fair anonymization exist?” *International Review of Law, Computers & Technology* 28(1): 21–44.

AlSabah, Mashael, Kevin Bauer, and Ian Goldberg. 2012. “Enhancing Tor’s Performance using Real-time Traffic Classification.” Presented at CCS’12, Raleigh, North Carolina, USA, October 16–18.

Althusser, Louise. 1971. “Ideology and Ideological State Apparatuses (Notes towards an Investigation).” In *Lenin and Philosophy and Other Essays*, translated by B. Brewster. London: Monthly Review Press.

Amoore, Louise. 2011. “Data Derivatives: On the Emergence of a Security Risk Calculus for Our Times.” *Theory, Culture & Society* 28(6): 24–43.

Anderson, Benedict. 1983. *Imagined Communities: Reflections on the origins and spread of nationalism*. London: Verso.

Anderson, Chris. 2008.“The End of Theory: The Data Deluge Makes the Scientific Method Obsolete.” *WIRED*.16.

Angwin, Julia. 2016. “Breaking the Black Box.” *Propublica*. October 19, 2016. Available here: https://www.propublica.org/article/breaking-the-black-box-what-facebook-knows-about-you

Anonymous. 2014. “Interview With Grams Search Engine Admin: Exciting Features Ahead!” *Deepdotweb* May 3. Available here: https://www.deepdotweb.com/2014/05/03/interview-with-grams-search-engine-admin-exciting-features-ahead/

Ashby, W. Ross. 1956. “The Black Box.” Chapter in *An Introduction to Cybernetics*. London: Chapman & Hall.

B

Ball, James, Bruce Schneier and Glenn Greenwald. 2013. “NSA and GCHQ Target Tor Network That Protects Anonymity of Web Users.” *The Guardian* October 4, section World news. Available here: http://www.theguardian.com/world/2013/oct/04/nsa-gchq-attack-tor-network-encryption

Ballatore, Andrea, Mark Graham, and Shilad Sen. 2017. “Digital Hegemonies: The Localness of Search Engine Results.” *Annals of the American Association of Geographers* 107(5): 1194-1215.

Barad, Karen. 2003. “Posthumanist Performativity: Toward an Understanding of How Matter Comes to Matter.” *Signs: Journal of Women in Culture and Society* 28(3).

Barad, Karen. 2007. *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Durham: Duke University Press.

Barrett, Brian. 2018. “I Used Only Bing for 3 Months. Here's What I Found—and What I Didn't.” *Wired* October 17. Available here: https://www.wired.com/story/tried-bing-search-google-microsoft

Barth, Brian J. 2019. “Are You Afraid of Google? BlackBerry Cofounder Jim Balsillie Says You Should Be.” *The Walrus*. Available here: https://thewalrus.ca/are-you-afraid-of-google-blackberry-cofounder-jim-balsillie-says-you-should-be/

Bartlett, Jamie. 2014a. *The Dark Net: Inside the Digital Underworld.* Portsmouth: Heinemann.

Bartlett, Jamie. 2014b. “What does the Dark Net mean for the Future of Intelligence Work?” Presented at the 2014 Annual Vincent Briscoe Security Lecture Imperial College, London, October 29.

Bateson, Gregory and Margaret Meade. 1973. “Diagram from Interview.“ *CoEvolutionary Quarterly* 10: 32-44. Available at: http://oikos.org/forgod.htm

Battelle, John. 2006. *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*. New York: Penguin.

Battelle, John. 2010. *The Database of Intentions Is Far Larger Than I Thought.* Available here: https://battellemedia.com/archives/2010/03/the_database_of_intentions_is_far_larger_than_i_thought. March 5, 2010.

Becker, Konrad and Stalder. Felix. 2009. *Deep search: The politics of search beyond Google*. Vienna: Studien Verlag.

Beer, David. 2009. “Power through the algorithm? Participatory web cultures and the *technological unconscious.” New Media & Society* 11(6): 985-1002.

Beigi, Mandis, Ana B. Benitez, and Shih-Fu Chang. 1998. “MetaSEEk: A Content-Based Meta-Search Engine for Images.” *Proceedings of the SPIE 1998 Conference on Storage and Retrieval for Image and Video Databases VI (IST/SPIE-1998) San Jose, CA, Jan 28-30* 3312.

Bellovin, Steve, Matt Blaze, and Susan Landau. 2016. “Insecure Surveillance: Technical Issues with Remote Computer Searches.” *Computer published by the IEEE Computer Society 49(3).*

Benkler, Yochai. 2005. “The New Open Source Economics.” *TED talk*. Available here: https://www.ted.com/talks/yochai_benkler_on_the_new_open_source_economics#t-343948.

Benkler, Yochai. 2006. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven: Yale University Press.

Berger, John. 1972. *Ways of Seeing*. London: Penguin.

Bernard, Andreas. 2016. "The Total Archive: On the Function of Non-Knowledge in Digital Cultures." Chapter in *Non-Knowledge in Digital Cultures, edited by* A. Bernard, M. Leeker, and M. Koch. Berlin: Meson Press.

Beverungen Armin. 2014. "Wages for or Against Usership?" Available here: n.e.w.s.: http://northeastwestsouth.net/arbitrating-attention-paid-usership.

Beverungen Armin, Stefan Böhm, and Chris Land. 2015. "Free Labour, Social Media, Management: Challenging Marxist Organization Studies." *Organization Studies* 36(4): 473-489.

Beverungen, Armin, Timon Beyes, and Lisa Conrad. 2019. *Digital Media*. Thousand Oaks: Sage Publishers

Beyes, Timon. 2018. "The work of disconnection." Keynote Lecture at the 8th Organizations, Artifacts & Practices (OAP) conference, Amsterdam, June 20-22.

Beyes, Timon. 2021. "Organisation." Chapter in *Uncertain Archives: Critical Keywords for Big Data*, edited by N.B. Thylstrup, D. Agostinho, A. Ring, C. D'Ignazio, and K. Veel. Cambridge: MIT Press.

Beyes, Timon, Robin Holt, and Claus Pias. 2019. *The Oxford Handbook of Media, Technology, and Organization Studies*. Oxford: Oxford University Press.

Beyes, Timon and Claus Pias. 2014. "Transparenz und Geheimnis." *Zeitschrift für Kulturwissenschaft* 2: 111-118.

Beyes, Timon and Claus Pias. 2019. "The Media Arcane." *Grey Room* 75: 84-107. Cambridge: MIT Press Journals.

Biddle, Peter, Paul England, Marcus Peinado, and Bryan Willman. 2003. "The Darknet and the Future of Content Distribution." *Digital Rights Management* 2696: 155-176.

Bilić, Paško. 2017. "A Critique of the Political Economy of Algorithms: Brief History of Google's Technological Rationality." *Westminster Advanced Studies* 5.

Birchall, Clare. 2018. Shareveillance: The Dangers of Openly Sharing and Covertly Collecting Data. Minneapolis: University of Minnesota Press.

Biryukov, Alex, Ivan Pustogarov, and Ralf-Philipp Weinmann. 2013. "Trawling for tor hidden services: Detection, measurement, deanonymization." Presented at IEEE Symposium on Security and Privacy, San Francisco CA, May 19-22.

Bjerke, Björn. 2007. *Understanding Entrepreneurship.* Cheltenham: Edward Elgar.

Blackman, Lisa. 2019. *Haunted Data: Affect, Transmedia, Weird Science*. London: Bloomsbury Academic.

Blas, Zach. 2016. *"In Practice: Opacities." Camera Obscura: Feminism, Culture, and Media Studies 31(2).*

Blome, Astrid. 2010. "Offices of Intelligence and Expanding Social Spaces." Chapter in *The Dissemination of News and the Emergence of Contemporaneity in Early Modern Europe*, edited by B. Dooley. Burlington: Ashgate.

Bogatin, Donna and Danny Sullivan. 2007. "Open debate." *Fast Company*, p. 116. Available here: https://www.fastcompany.com/58707/open-debate

Böhm, Steffen and Chris Land. 2012. "The new 'hidden abode': reflections on value and labour in the new economy." *The Sociological Review* 60(2).

Bolsover, Gillian and Philip Howard. 2017. "Computational propaganda and political big data: Moving toward a more critical research agenda." *Big Data* 5(4): 273–276.

Borch, Christian. 2015. *Foucault, Crime and Power: Problematisations of Crime in the Twentieth Century*. Abingdon: Routledge.

Borch, Christian. 2016. "Shiller and the uses of sociology in behavioural finance," December 20. Available here: https://socfinance.wordpress.com/2016/12/20/shiller-and-the-uses-of-sociology-in-behavioural-finance/.

Borch, Christian and Ann Christine Lange. 2017. "Market Sociality: Mirowski, Shiller and the Tension between Mimetic and Anti-mimetic Market Features." *Cambridge Journal of Economics* 41(4): 1197-1212.

Borland, John. 2013. "For Tor, publicity a mixed blessing" *Wired* December 28. *Available here:* [http://www.wired.com/2013/12/tor-publicity-mixed-blessing/]

boyd, danah. 2012. "The Politics of 'Real Names: Power, Context, and Control in Networked Publics." *Communications of the ACM* 55(8): 29-31.

Brewster, Thomas. 2015. "$30,000 to $1 Million -- Breaking Tor Can Bring In The Big Bucks." *Forbes Magazine* November 12. Available here: https://www.forbes.com/sites/thomasbrewster/2015/11/12/earn-money-breaking-tor/

Bria, Francesca. 2017. "Reclaiming Europe's Digital Sovereignty." *Financial Times.* Interview. October 25, 2017. Available here: https://www.francescabria.com/talks.html

Brin, Sergei. and Larry Page. 1998. "*The Anatomy of a Large-Scale Hypertextual Web Search Engine."* Presented at Seventh International World-Wide Web Conference (WWW 1998), Brisbane, Australia, April 14-18.

Bruns, Axel. 2008. *Blogs, Wikipedia, Second Life and Beyond, from production to*

produsage. New York: Peter Lang Publishing.

Brunton, Finn and Helen Nissenbaum. 2015. *Obfuscation*. *A User's Guide for Privacy and Protest*. Cambridge: MIT Press.

Buckner, Deborah. 2008. "Internet Search and Seizure in United States v. Forrester: New Problems in the New Age of Pen Registers." *Brigham Young University Journal of Public Law* 22(2): 499-517. Available here: http://digitalcommons.law.byu.edu/jpl/vol22/iss2/9

Bunz, Mercedes. 2014. *The Silent Revolution: How Digitalization Transforms Knowledge, Work, Journalism and Politics without Making Too Much Noise.* Basingstoke: Palgrave Macmillan.

Bush, Vannevar. 1945. "As We May Think." *The Atlantic*. Available here: https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/.

Buterin, Vitalik. 2016. "The Road Ahead" *@KMPG*. Lecture. June 16, 2016. Available here: https://www.youtube.com/watch?v=xtJaM05VUR4

Butler, Judith. 2005. *Giving an account of oneself. New York: Fordham University Press.*

C

Cadman, Louisa. 2009. *Non-Representational Theory/Non-Representational Geographies*, in R. Kitchin and N. Thrift (eds) International Encyclopedia of Human Geography, Vol. 7, pp. 456–63. London: Elsevier.

Cadwalladr, Carole. 2016a. "Google is not 'just' a platform. It frames, shapes and distorts how we see the world". *The Guardian*. November 11, 2016. Available here: https://www.theguardian.com/commentisfree/2016/dec/11/google-frames-shapes-and-distorts-how-we-see-world.

Cadwalladr, Carole. 2016b. "Google, democracy and the truth about internet search". *The Guardian*. December 4, 2016. Available at: https://www.theguardian.com/technology/2016/dec/04/google-democracy-truth-internet-search-facebook.

Çalışkan, Emin, Tomas Minárik and Anna-Marie Osula. 2015. 'Technical and legal overview of the Tor anonymity network', Tallin: CCDCOE, NATO Cooperative Cyber Defence Centre of Excellence.

Carnegie Mellon University (CMU). 2015. "Media Statement." November 18th. Available at: http://www.cmu.edu/news/stories/archives/2015/november/media-statement.html

Carr, David F. 2006. "How Google Works". Baseline Magazine. July 7, 2006. Available here: http://www.baselinemag.com/c/a/Infrastructure/How-Google-Works-1

Carr, Nicholas. 2008. *The Big Switch: Rewiring the World, from Edison to Google*. New York: W.W. Norton.

Castells, Manuel. 2009. *Communication Power*. Oxford: Oxford University Press.

Certeau, Michel de. 1984. *The practice of everyday life*. Berkeley: University of California Press.

Chakravarty, Sambuddho, Marco V. Barbera, Georgios Portokalidis, Michalis Polychronakis, Angelos D. Keromytis. 2014. "On the Effectiveness of Traffic Analysis Against Anonymity Networks Using Flow Records". Proceeding PAM 2014 Proceedings of the 15th International Conference on Passive and Active Measurement Volume 8362. Pages 247-257. New York: Springer Verlag.

Chaum, David L. 1981. "Untraceable Electronic Mail, Return addresses, and Digital Pseudonyms". Communications of the ACM, February 1981. Volume 24. Number 2

Chaum, David. 1982. *Computer Systems Established, Maintained, and Trusted by Mutually Suspicious Groups*. Dissertation. University of California, Berkeley.

Chen, Sophia. 2018. "Quantum Mechanics Could Solve Cryptography's Random Number Problem". WIRED. April 4, 2018. https://www.wired.com/story/quantum-mechanics-could-solve-cryptographys-random-number-problem/

Cheney-Lippold, John. 2011. "A New Algorithmic Identity: Soft Biopolitics and the Modulation of Control". *Theory, Culture & Society*, *28*(6), 164–181.

Christl, Wolfie. 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.

Christl, Wolfie. 2018. April 1 and April 13, 2018. Twitter.

Chun, Wendy Hui Kyong. 2005. *Control and Freedom: Power and Paranoia in the Age of Fiber Optics.* Cambridge: MIT Press

Chun, Wendy Hui Kyong. 2011. *Programmed Visions: Software and Memory*. Cambridge: MIT Press.

Chun, Wendy Hui Kyong. 2016. *Habitual New Media: Updating to Remain the Same.* Cambridge: MIT Press.

Chun, Wendy Hui Kyong. 2018. "Queerying Homophily". In *Pattern Discrimination*. Minneapolis: University of Minnesota Press. Lüneburg: Meson Press

Cohen, Judith. 2012. *Configuring the Networked Self*. New Haven: Yale University Press.

Coleman, Gabrielle. 2014. *Hacker, Hoaxer, Whistleblower, Spy: The Many Faces of Anonymous*. London: Verso.

CONSTANT 2020. "Paul Otlet, an Omissum". Available here: https://gitlab.constantvzw.org/diversions/paul-otlet-an-omissum

Cooper, Robert and Law, John. 2016. “Organization: Distal and Proximal Views”. In *For Robert Cooper: Collected Work*. Edited by Gibson Burrell and Martin Parker. New York; Routledge, pp.199-235.

Cox, Geoff and McLean, Alex. 2012. *Speaking Code-Coding as Aesthetic and Political Expression*. Cambridge: MIT Press.

Cox, Joseph. 2015. “Tor Attack Could Unmask New Hidden Sites in Under Two Weeks.” November 13th. *Motherboard*. Available at: https://motherboard.vice.com/en_us/article/tor-attack-could-unmask-new-hidden-sites-in-under-two-weeks.

Cox, Joseph. 2016 “Confirmed: Carnegie Mellon University Attacked Tor, Was Subpoenaed By Feds.” February 24th. *Motherboard*. Available at: https://motherboard.vice.com/en_us/article/carnegie-mellon-university-attacked-tor-was-subpoenaed-by-feds.

Cramer, Florian. 2013. “Animals that Belong to the Emperor”. Chapter in Anti-Media. *Ephemera on Speculative Arts.* nai010. Institute of Network Cultures.

Cramer, Florian. 2014. “What is Post Digital?” *APRJA*. Volume 3, Issue 1, 2014. Available here: http://www.lab404.com/142/cramer.pdf.

Curtis, Adam. 2016. *Hypernormalisation*. BBC Documentary.

D

Davenport, David. 2002. “Anonymity on the Internet: Why the Price May Be Too High.” *Communications of the ACM*, April 2002, Vol. 45 No. 4, Pages 33-35.

Davies, William. 2017. “How statistics lost their power – and why we should fear what comes next”. *The Guardian*. January 19, 2017. Available here: https://www.theguardian.com/politics/2017/jan/19/crisis-of-statistics-big-data-democracy

Davies, William. 2018. “Short Cuts”. *London Review of Books*. Vol. 40 No. 7 · 5 April 2018. pages 20-21.

Davilla, Marixenia. 2017. “Is Big Data a Different Kind of Animal? The Treatment of Big Data Under the EU Competition Rules”. *Journal of European Competition Law & Practice*. 2017, Vol. 8, No. 6

Day, Ronald E. 2014. *Indexing It All: The Subject in the Age of Documentation, Information and Data.* Cambridge, MA: MIT Press.

Degutin, Stéphane & Wagon, Gwenola. 2014. *World Brain*. Film and Transmedia project.

Deibert, Ron. 2020. Massey Lectures. Lecture 2.

Deleuze, Gilles. 1992. “Postscript on the Societies of Control.” *October.* Vol. 59. (Winter, 1992). pp. 3-7.

Diakopoulos, Nicholas. 2013. *Algorithmic accountability reporting: On the investigation of black boxes*. A Tow/Knight Brief. Tow Center for Digital Journalism, Columbia Journalism School. Retrieved: August 21, 2014, from http://towcenter.org/algorithmic-accountability-2/.

Diakopoulos, Nicholas. 2016. "Accountability in Algorithmic Decision Making". *Communications of the ACM*. February 2016. Vol. 59. No. 2

Dingledine, Roger. 2014. "Traffic correlation using netflows". *Tor Project blog*: https://blog.torproject.org/blog/traffic-correlation-using-netflows. November 14, 2014.

Dingledine, Roger. 2014a. "Tor security advisory: "relay early" traffic confirmation attack". *Tor Project blog*: https://blog.torproject.org/tor-security-advisory-relay-early-traffic-confirmation-attack. July 30, 2014.

Dingledine, Roger. 2015. "Did the FBI Pay a University to Attack Tor Users?" Tor Project blog: https://blog.torproject.org/did-fbi-pay-university-attack-tor-users. November 11, 2015.

Doctorow, Cory. 2019. "Adversarial Interoperability: Reviving an Elegant Weapon From a More Civilized Age to Slay Today's Monopolies". *EFF*. JUNE 7, 2019. https://www.eff.org/deeplinks/2019/06/adversarial-interoperability-reviving-elegant-weapon-more-civilized-age-slay.

Doyle, Gillian. 2002. *Understanding Media Economics*. London: Sage Publications.

Drozdiak, Natalia, Schechner, Sam. 2015. "EU Court Says Data-Transfer Pact With U.S. Violates Privacy." *Wall Street Journal*, October 6, 2015. Available here: http://www.wsj.com/articles/eu-court-strikes-down-trans-atlantic-safe-harbor-data-transfer-pact-1444121361.

Drucker, Johanna. 2013. "Performative materiality and theoretical approaches to interface". *Digital Humanities Quarterly*, 7(1). Available here: http://www.digitalhumanities.org/dhq/vol/7/1/000143/000143.html

Dyson, George. 2012. *Turing's Cathedral: The Origins of the Digital Universe*. New York: Pantheon.

E

Edelman, Gilad. 2020. "'Do Not Track' Is Back, and This Time It Might Work." *WIRED*. 10.07.2020. Available here: https://www.wired.com/story/global-privacy-control-launches-do-not-track-is-back/?

Edelman, Gilad. 2020b. "Ad Tech Could Be the Next Internet Bubble" *WIRED*.10.05.2020. Available here: https://www.wired.com/story/ad-tech-could-be-the-next-internet-bubble/?

Eggers, Dave. 2013. *The Circle*. New York: Knopf.

Eli the Computer Guy. 2015. video "TOR Security Concerns and Considerations" Available here: https://www.youtube.com/watch?v=IvaBc0Q49Ys

Elmer, Greg. 2003. *Profiling Machines: Mapping the Personal Information Economy.* Cambridge: MIT Press.

Emerson, Lori. 2016. "Othernet, Alternet, Darknet". Workshop. February 24, 2016 Available here: https://loriemerson.net/2016/02/24/workshop-othernet-alternet-darknet/

Ernst, Wolfgang. 2012. *Digital Memory and the Archive*. Minneapolis: University of Minnesota Press.

Esmail, Sam. 2016. *Mr Robot*. NBC.

Esposito, Elena. 2017. "Algorithmic Memory and the Right to Be Forgotten on the Web". *Big Data & Society*. January–June 2017: 1–11. Sage Publishers.

Esteve, Asunción. 2017. "The business of personal data: Google, Facebook, and privacy issues in the EU and the USA". *International Data Privacy Law* 2017; 7 (1): 36-47.

Evans, David S. and Schmalensee, Richard. 2016. *Matchmakers: The New Economics of Multisided Platforms*. Boston: Harvard Business Review Press.

EU Commission Press Corner. 2019. "Antitrust: Commission fines Google €1.49 billion for abusive practices in online advertising." Press release. March 20, 2019. 2019https://ec.europa.eu/commission/presscorner/detail/en/IP_19_1770

F

Farivar, Cyrus. 2016. "Top Silk Road 2.0 admin "DoctorClu" pleads guilty, could face 8 years in prison." *Ars Technica,* April 4th. Available here: https://arstechnica.com/tech-policy/2016/04/top-silk-road-2-0-admin-doctorclu-pleads-guilty-could-face-8-years-in-prison/.

Felten, ed. 2014 "Why were CERT researchers attacking Tor?" *Freedom to Tinker Blog.* July 31. Available at: https://freedom-to-tinker.com/2014/07/31/why-were-cert-researchers-attacking-tor/.

Feuz, Martin; Fuller, Matthew; Stalder, Felix. 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 here: http://firstmonday.org/article/view/3344/2766.

Fidgen, Jo. 2011. *Pornography: What Do We Know?* Analysis programme of BBC. Available here: http://www.bbc.co.uk/programmes/b02ykg3m.

Fincher, David. 2016. *House of Cards*. 2016. (Season 4) Netflix

Flyverbom, Mikkel, Lars Thøger Christensen and Hans Krause Hansen. 2015. "The Transparency–Power Nexus: Observational and Regularizing Control". *Management Communication Quarterly*. Vol. 29(3) 385–410.

Flyverbom, Mikkel, Paul Leonardi., Michael Stohl and Cynthia Stohl. 2016. “The Management of Visibilities in the Digital Age: Introduction”, *International Journal of Communication*, Vol. 10, p. 98–109

Flyverbom, Mikkel Ron Deibert and Dirk Matten. 2019. “The Governance of Digital Technology, Big Data, and the Internet: New Roles and Responsibilities for Business.” *Business & Society,* Vol. 58, No. 1, 1, p. 3-19

Fogerlie, Garrett. 2016. “How Tor Users Got Caught Part 2 Cliff Notes”. August 16, 2016. Personal Blog. Available here: http://se.azinstall.net/2016/08/how-tor-users-got-caught-part-2-cliff.html. There is also a video here that shows his presentation at DefCon: https://www.youtube.com/watch?v=TQ2bk9kMneI&feature=youtu.be

Foot, Kirstin, A.(ed.). 2014. *Media technologies: Essays on communication, materiality, and society*. Cambridge, MA: MIT Press.

Forte, Andrea, Nazanin Andalibi, Rachel Greenstadt. 2017. “Privacy, Anonymity, and Perceived Risk in Open Collaboration: A Study of Tor Users and Wikipedians”. *Proceedings of Computer-Supported Cooperative Work and Social Computing* (CSCW). Portland

Foster, Nick & Murphy, David. 2016. “The Selfish Ledger”. Video. Available here: https://www.youtube.com/watch?time_continue=22&v=QDVVo14A_fo.

Foucault, Michel. 1975. *Discipline & Punish: The Birth of the Prison*. London: Vintage Books.

Foucault, Michel. 1982. “The Subject and Power”. *Critical Inquiry,* Vol. 8, No. 4 (Summer 1982), pp. 777-795. Chicago: The University of Chicago Press.

Foucault, Michel. 1988. *Technologies of the Self*. Edited by Luther H. Martin, Huck Gutman. Amherst: University of Massachusetts Press.

Foucault, Michel. 1994. *Ethics: Subjectivity and Truth. The Essential Works of Michel Foucault 1954-1984*. Volume One. New York: The New York Press.

Foucault, Michel. 1997. “What is Critique?” in The Politics of Truth, eds. Sylvère Lotringer and Lysa Hochroth, (New York: Semiotext(e), transcript by Monique Emery, revised by Suzanne Delorme, et al., translated into English by Lysa Hochroth.

Foucault, Michel. 2005. *The Hermeneutics of the Subject: Lectures at the College de France 1981-1982*. London: Picador

Foucault, Michel. 2008. *The Birth of Biopolitics*: Lectures at the Collège de France, 1978–1979. New York: Palgrave Macmillan.

Franklin, Ursula. 1989. “The Real World of Technology”. Massey Lectures.

Frediani, Carola. 2014. “The new search engines shining a light on the deep web”. The Kernel. September 28, 2014. Available here:

https://kernelmag.dailydot.com/issue-sections/features-issue-sections/10376/how-to-search-deep-web-tor/.

Fuchs, Christian. 2011. “A Contribution to the Critique of the Political Economy of Google.” *Fast Capitalism*. 8 (1). Available here: https://www.uta.edu/huma/agger/fastcapitalism/8_1/fuchs8_1.html.

G

Galison, Peter. 1994. *The Ontology of the Enemy: Norbert Wiener and the Cybernetic Vision*. Critical Inquiry, Vol. 21, No. 1. (Autumn), pp. 228-266.

Gallagher, Ryan. 2013. “New Snowden Documents Show NSA Deemed Google Networks a “Target””. *Slate*. September 9, 2013. Available here: https://slate.com/technology/2013/09/shifting-shadow-stormbrew-flying-pig-new-snowden-documents-show-nsa-deemed-google-networks-a-target.html

Galloway, Alexander. 2004. *Protocol: How Control Exists after Decentralization*. Cambridge: MIT Press.

Galloway, Alexander. 2011a. “Black Box, Black Bloc.” In: Benjamin Noys (ed.): *Communization and its Discontents. Contestation, Critique, and Contemporary Struggles.* Brooklyn: Minor Compositions/Autonomedia, 237–249.

Galloway, Alexander. 2011b. “Are Some Things Unrepresentable?”, *Theory, Culture & Society.* 28(7-8): 85-102. Thousand Oaks: Sage Publishers.

Galloway, Alexander. 2012. *The Interface Effect*. Cambridge: Polity Press.

Galloway, Alexander. 2018. “A Theory of Media”. Blog. http://cultureandcommunication.org/galloway/a-theory-of-media#more-1369

Garfield, Eugene. 2007. The evolution of the Science Citation Index. *International Microbiology* (2007) 10:65-69.

Garland, Alex. 2015. *Ex Machina*. Film.

Gehl, Robert W. 2014. “Power/freedom on the dark web: A digital ethnography of the Dark Web Social Network”. Republished 2016. *New Media & Society*, Volume 18(7). 1219-1235. Thousand Oaks: Sage Publishers.

Gennaro, Cuofano. 2018a. “How Does DuckDuckGo Make Money? DuckDuckGo Business Model Explained”. *Four Week MBA*. Avai;lable here: https://fourweekmba.com/duckduckgo-business-model/

Gennaro, Cuofano. 2018b. “DuckDuckGo: The [Former] Solopreneur That Is Beating Google at Its Game”. *Four Week MBA*. Available here: https://fourweekmba.com/duckduckgo-vs-google/

Gesenhues, Amy. 2013. “Google’s Hummingbird Takes Flight: SEOs Give Insight On Google’s New Algorithm” *Search Engine Land.* September 30. Available at: http://searchengineland.com/hummingbird-has-the-industry-flapping-its-wings-in-excitement-reactions-from-seo-experts-on-googles-new-algorithm-173030

Geuss, Megan. 2015 “Alleged “right hand man” to Silk Road 2.0 leader arrested in Seattle.” *Ars Technica.* January 21st. Available at: https://arstechnica.com/tech-policy/2015/01/alleged-right-hand-man-to-silk-road-2-0-leader-arrested-in-seattle.

Giannandrea, John. 2017. “Google AI chief thinks reports of the AI apocalypse are greatly exaggerated.” Tech Crunch. Available here: https://techcrunch.com/2017/09/19/googles-ai-chief-thinks-reports-of-the-ai-apocalypse-are-greatly-exaggerated/

Gillespie, Tarleton. 2014. “The Relevance of Algorithms”. *Media Technologies: Essays on communication, materiality, and society*. Ed. Tarleton Gillespie, Pablo Boczkowski, and Kirsten Foot. Cambridge, MA: MIT Press.

Gillespie, Tarleton. 2015. “Platforms Intervene”. *Social Media + Society*. April 2015. Thousand Oaks: Sage Publishers

Gillespie, Tartleton. 2017. “The Platform Metaphor”. Available here: https://www.hiig.de/en/blog/the-platform-metaphor-revisited/

Girard, Rene. 1961. *Deceit, Desire and the Novel. Baltimore:* Johns Hopkins University Press.

Gitelman, Lisa and Jackson, Virginia. 2013. ‘Introduction’. *Raw Data Is an Oxymoron*. Lisa Gitelman, ed. Cambridge, MIT Press.

Glanville, Randall. 2003. “Second order cybernetics”. In F. Parra-Luna (Ed.), *Systems science and cybernetics*. In Encyclopaedia of life support systems (EOLSS). Oxford. Available at http://cepa.info/2326

Gleick, James. 2011. *The Information: A History, A Theory, A Flood*. New York: Pantheon.

Glenny, Misha. 2015. “Into the Wild”. Review of *The Dark Net: Inside the Digital Underworld* by Jamie Barlett. *London Review of Books*. 19 March 2015.

Google.com. 2015. “Google Inc. Announces Second Quarter 2015 Results” https://investor.google.com/earnings/2015/Q2_google_earnings.html

Google. 2017. “Inside search”. Available at: https://www.google.com/intl/bn/insidesearch/

Gorges, Sir Arthur. 1611. “Publique Register for Generall Commerce”. Available here: http://quod.lib.umich.edu/e/eebo/A22719.0001.001/1:4?rgn=div1;view=fulltext.

Goriunova, Olga. 2015. “Digital Subject”. *Online lecture*. Available here: https://www.youtube.com/watch?v=yAIheBLmG6M

Goriunova, Olga. 2016. “Digital Ontologies as Productive Process. Theorizing the Contemporary”. *Fieldsights*, March 24. Available at: https://culanth.org/fieldsights/digital-ontologies-as-productive-process

Goriunova, Olga. 2019. “The Digital Subject: People as Data as Persons. *Theory, Culture and Society*. Vol. 36, No. 6, 01.11.2019, p. 125-145

Graff, Garrett M. 2018. Indicting 12 Russian Hackers Could Be Mueller’s Biggest Move Yet. WIRED. July 13, 2018. Available here: https://www.wired.com/story/mueller-indictment-dnc-hack-russia-fancy-bear/

Granovetter, Mark. 1973. “The Strength of Weak Ties.” *American Journal of Sociology.* 78, 6:1377.

Graw, Isabelle. 2015. “False Polarities and Economic Subtexts Art Good, Market Evil?” Texte zur Kunst. May 6, 2015.Available at: https://www.textezurkunst.de/articles/art-good-market-evil/

Greenberg, Andy. 2018. “Operation Bayonet: Inside the Sting That Hijacked an Entire Dark Web Drug Market”. *WIRED*. 03.08.2018. Available here: https://www.wired.com/story/hansa-dutch-police-sting-operation/?mbid=nl_030918_daily_list_p

Greenwald, Glenn and Ewen MacAskill. 2013. “NSA Prism program taps in to user data of Apple, Google and others”. *The Guardian*. June 7, 2013. Available here: https://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data.

Grind, Kirsten, Sam Schechner and Robert McMillan. 2019. “How Google Interferes With Its Search Algorithms and Changes Your Results:The internet giant uses blacklists, algorithm tweaks and an army of contractors to shape what you see.” *Wall Street Journal*. November 15, 2019. Available here: https://www.wsj.com/articles/how-google-interferes-with-its-search-algorithms-and-changes-your-results-11573823753.

Gross, Grant. 2013. “Report: NSA has little success cracking Tor”. *Computer World*. October 4, 2013. Available here: https://www.computerworld.com/article/2485700/encryption/report--nsa-has-little-success-cracking-tor.html

Groys, Boris. 2012. *Google, Words beyond Grammar*. In 100 Notes, 100 Thoughts: 046 Documenta (13). Berlin: Hatje Cantz

H

Haeselin, David. 2017. “Welcome to the Indexed World: Thomas Pynchon’s *Bleeding Edge* and The Things Search Engines Will Not Find”. *Critique: Studies in Contemporary Fiction*. Volume 58, 2017 – Issue 4. Routledge.

Halavais, Alexander. 2009. *Search Engine Society.* Cambridge: Polity Press.

Hall, Gary. 2016. *The Uberfication of the University*. Minneapolis: University of Minnesota Press.

Haraway, Donna. 1991. *Simians, Cyborgs and Women: The Re-invention of Nature.* London: Free Association Books.

Hargittai, Ezster. 2007. "The Social, Political, Economic, and Cultural Dimensions of Search Engines: An Introduction". *Journal of Computer-Mediated Communication* 12. 769-777. Oxford: Oxford University Press.

Hariri, Yuval Noah. 2017. "Mensen, Goden, Technologie". (Humans, Gods, Technology) *Tegenlicht.* October "29, 2017. Available here: https://www.vpro.nl/programmas/tegenlicht/kijk/afleveringen/2017-2018/mensen-goden-en-technologie.html

Harrington, Brooke. 2010. "Economic Sociology: The Sense of Dissonance–An Interview with David Stark". *The Society Pages*. Available here: https://thesocietypages.org/economicsociology/2010/04/14/the-sense-of-dissonance-an-interview-with-david-stark/

Harvard Business School. 2007. "Platform-Mediated Networks: Definitions and Core Concepts", Module Note, 9-807-049.

Hayles, Katherine. 1999. *How we Became Posthuman:* Virtual Bodies in Cybernetics, Literature, and Informatics. Chicago: University of Chicago Press.

Heffernan, Virginia. 2017. "Just Google It: A Short History of a Newfound Verb". *WIRED*. Available here: https://www.wired.com/story/just-google-it-a-short-history-of-a-newfound-verb/.

Heidegger, Martin. 1977. *The Question Concerning Technology and other essays*. (1954) New York: Garland Publishing.

Helmond, Anne. 2013. "The Algorithmization of the Hyperlink". *Computational Culture*. Volume 3. Available here: http://computationalculture.net/the-algorithmization-of-the-hyperlink/.

Hendler, James and Hugill, Andrew. 2013. "The syzygy surfer: (Ab)using the semantic web to inspire creativity". *International Journal of Creative Computing*. Volume 1, Issue 1. pp. 20–34

Hermann, Michael, Ren Zhang, Kai-Chun Ning, Claudia Diaz and Bart Preneel. 2014. "Censorship-Resistant and Privacy-Preserving Distributed Web Search". 14th IEEE International Conference on Peer-to-Peer Computing. *IEEE*. London, pp. 1-10. doi: 10.1109/P2P.2014.6934312.

Hersher, Rebecca. 2017. "What Happened When Dylann Roof Asked Google For Information About Race?" *NPR*. January 10, 2017. Available here: https://www.npr.org/sections/thetwo-way/2017/01/10/508363607/what-happened-when-dylann-roof-asked-google-for-information-about-race.

Hillis, Ken, Michael Petit and Kylie Jarret. 2013. *The Culture of Search*. New York: Routledge.

Hindman, Matthew. 2009. *The Myth of Digital Democracy*. Princeton: Princeton Press.

Hindman, Matthew. 2018. *The Internet Trap: How the Digital Economy builds monopolies and undermines democracy.* Princeton: Princeton University Press

Hindman, Matthew and Kostas, Tsioutsiouliklis, Judy A. Johnson. 2003. "Googlearchy: How a Few Heavily-Linked Sites Dominate Politics on the Web". Annual meeting of the Midwest Political Science Association 4, 1-33.

Hogan, Bernie. 2012. "Pseudonyms and the Rise of the Real-Name Web." In J. Hartley, J. Burgess and A. Bruns (Eds.), *A Companion to New Media Dynamics* (pp. 290–308). Chichester, UK: Blackwell Publishing Ltd.

Hoofnagle, Chris Jay, Jennifer King, Su Li and Joseph Turow. 2010. "How different are young adults from older adults when it comes to information privacy attitudes and policies?" *SSRN Electronic Journal*. Available here: http://www.ssrn.com/abstract=1589864.

Howe, Daniel C. and Nissenbaum, Helen. 2017. "Engineering Privacy and Protest: A Case Study of AdNauseum". Proceedings of the 3rd International Workshop on Privacy Engineering. Volume 1873, 57-64.

Hugill, Andrew, Hongji Yang, Fania Raczinski and James Sawle. 2013. "The pataphysics of creativity: developing a tool for creative search". *Digital Creativity*. 24:3, 237-251.

Hunger, Francis. 2017. "Artificial Des-Intelligence or Why machines will not take over the world. At least not now. Part I: There is no Artificial Intelligence". *Database Cultures*. November 15, 2017. Available here: http://databasecultures.irmielin.org/artificial-des-intelligence/#more-907

Hwang, Tim. 2020. *Subprime Attention Crisis: Advertising and the Time Bomb at the Heart of the Internet.* New York: FSG Originals

I

Introna, Lucas. 2007. "Maintaining the Reversibility of Foldings: Making the ethics (politics) of information technology visible." *Ethics and Information Technology,* 9(1): 11-25. Springer

Introna, Lucas. 2016. "The algorithmic choreography of the impressionable subject". In Seyfert, Robert and Roberge, Jonathan (Eds.), *Algorithmic Cultures: Essays on Meaning, Performance and New Technologies*. Abingdon-on-Thames: Routledge.

Introna, Lucas D. and Helen Nissenbaum. 2000. "Shaping the Web: Why the politics of search engines matters". *The Information Society*. 16 (3), 169-185. DOI: 10.1080/01972240050133634

Invisible Committee. 2014. "Fuck Off Google". In *To our friends*. Cambridge: MIT Press.

Irani, Lilly. 2013. "The cultural work of microwork". *New Media & Society*, 0(0) 1–21. Thousand Oaks: Sage Publishers

J

Jacobsen, Mikkel. 1991. *The Freudian Subject*. Stanford: Stanford University Press.

Jameson, Frederic. 1988. *Cognitive Mapping. In Marxism and the Interpretation of Culture*. Cary Nelson, Lawrence Grossberg. Eds. Chicago: University of Illinois Press

Jarrett, K. 2014. A Database of Intention? R. König & M. Rasch, (Eds), Society of the Query Reader #9: *Reflections on Web Search*, Amsterdam: Institute of Network Cultures.

Jiang, Min. 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, Vol. 16(2) 212-233.

K

Kadianakis, George. 2015. "Some statistics about onions". *Tor Project blog*: https://blog.torproject.org/some-statistics-about-onions. February 26, 2015.

Kallinokos, Jannis. 2003. "Work, Human Agency and Organizational Forms: An Anatomy of Fragmentation". Organization Studies. 24(4): 595–618. Thousand Oaks: SAGE Publications.

Kaplan, Frederic. 2014. "Linguistic Capitalism and Algorithmic Mediation". In *Representations*, Vol. 127 No. 1, Summer *2014*. Oakland: University of California Press.

Kim, Larry. 2017. "Has machine learning created a new model for SEO ranking?" *Search Engine Land*. February 22, 2017. Available here: http://searchengineland.com/new-model-seo-ranking-269419.

Kitchin, Rob. 2017. "Thinking critically about and researching algorithms". *Information, Communication & Society*, 20:1, 14-29.

Kittler, Friedrich A. 1999. *Gramaphone, Film, Typewriter*. Stanford. Stanford University Press.

Kleinberg, Jon M. 1997. "Authoritative Sources in a Hyperlinked Environment" Publication: SODA '98: Proceedings of the ninth annual ACM-SIAM symposium on Discrete algorithmsJanuary 1998 Pages 668–677

Knappenberger, Brian. 2014. *The Internet's Own Boy: The Story of Aaron Swartz*. Luminant Media, Unjustsus Films.

Kopstein, Joshua. 2016. "Confused Judge Says You Have No Expectation of Privacy When Using Tor." *Motherboard,* Available here: https://motherboard.vice.com/en_us/article/confused-judge-says-you-have-no-expectation-of-privacy-when-using-tor-playpen-fbi-michaud

Krajweski Markus. 2010. "Ask Jeeves: Servants as Search Engines". Grey Room 38, Winter 2010, pp. 6–19. Cambridge: Grey Room, Inc. and Massachusetts Institute of Technology.

Krajewski, Markus. 2011. *Paper Machines: About Cards & Catalogs*, 1548-1929. Cambridge: MIT Press.

L

Laaff, Meike. 2011. "Networked Knowledge, Decades Before Google". *Der Spiegel.* July 22, 2011. Available here: http://www.spiegel.de/international/world/internet-visionary-paul-otlet-networked-knowledge-decades-before-google-a-775951.html.

Langley, Paul and Leyshon, Andrew. 2016. "Platform capitalism: The intermediation and capitalisation of digital economic circulation". *Finance and Society* /2016, EarlyView: 1-21.

Lawtoo, Nidesh. 2013 *Phantom of the Ego* (Modernism and the Mimetic Unconscious). Michigan State university Press.

Lazzaroto, Maurizio. 2010. "Conversation with Maurizio Lazzaroto". Public Editing Session #3, June 23, 2010, from 'Exhausting Immaterial Labour in Performance'; joint issue of *Le Journal des Laboratoires and TkH Journal for Performing Arts Theory* (no. 17), October 2010

Lazzaroto, Maurizio. 2014. *Signs and Machines, Capitalism and the Production of Subjectivity*. Los Angeles: Semiotext(e) Foreign Agent series.

Leonard, Herman B. 1983. "Elicitation of Honest Preferences for the Assignment of Individuals to Positions". *Journal of Political Economy*, Vol. 91, No. 3 (Jun., 1983), pp. 461-479. Chicago: The University of Chicago Press.

Lepore, Jill. 2013. "The Prism". *New Yorker*. Available at: https://www.newyorker.com/magazine/2013/06/24/the-prism

Leshem, Dotan. 2016. "Retrospectives: What Did the Ancient Greeks Mean by Oikonomia?" *Journal of Economic Perspectives,* 30(1): 225-38

Levine, Robert. 2015. "Behind the European privacy ruling that's confounding Silicon Valley", *New York Times,* October 9. Available here: http://www.nytimes.com/2015/10/11/business/international/behind-the-european-privacy-ruling-thats-confounding-silicon-valley.html?ref=technology&_r=2.

Levy, Steve. 2009. "Secret of Googlenomics: Data-Fueled Recipe Brews Profitability." *WIRED*. May 22, 2009. Available at: https://www.wired.com/2009/05/nep-googlenomics/

Lewandowski, Dirk. 2014. "Why We Need an Independent Index of the Web". R. König & M. Rasch, (Eds), Society of the Query Reader #9: *Reflections on Web Search*, Amsterdam: Institute of Network Cultures.

Lewandowski, Dirk. 2015. Chapter 1: New Perspectives on Web Search Engine Research. In *Web Search Engine Research*. Published online: 08 Mar 2015; 1-16.

Lewandowski, Dirk. 2017a. Users' Understanding of Search Engine Advertisements. Journal of Information Science Theory and Practice. 5 (4):6-25 (2017).

Lewandowski, Dirk, Friederike Kerkmann, Sandra Rümmele and Sebastian Sünkler. 2017. “An Empirical Investigation on Search Engine ad Disclosure”. *Journal Of The Association For Information Science And Technology*, 69(3):420–437, 2018.

Lipinski, Tomas A. 2002. “To Speak or Not to Speak: Developing Legal Standards for Anonymous Speech on the Internet”. Informing Science InSITE - “Where Parallels Intersect” June 2002.

Lovink, Geert. 2009. “Society of the Query: The Googlization of our lives”. In K. Becker and F. Stalder (Eds). *Deep search: The politics of search beyond Google*. Innsbruck: Studien Verlag, pp. 45–53.

Lovink, Geert and Nathaniel Tkacz. 2015. “MoneyLab: Sprouting New Digital-Economic Forms”. In *Moneylab: An Intervention in Digital Economy*. Amsterdam: Institute of Networked Culture.

Lovink, Geert. 2017. NXS Interview with Geert Lovink: “Defining the Synthetic Self.” December 4, 2017. Available here: http://networkcultures.org/geert/2017/12/04/nxs-interview-with-geert-lovink-defining-the-synthetic-self/.

Lury, Celia and Wakeford, Nina 2012. *Inventive Methods: The Happening of the Social*. Abingdon: Routledge

Lynch, Richard. 2015. “CMU’s Software Engineering Institute Contract Renewed by Department of Defense for $1.73 Billion.” Press Release, Carnegie Mellon University. July 28th Available here: https://www.cmu.edu/news/stories/archives/2015/july/sei-contract-renewed.html.

Lyon, David. 2007. “Surveillance, Security and Social Sorting: Emerging Research Priorities”. *International Criminal Justice Review*, *17*(3), 161–170.

Lyotard, Jean-Francois. 1984. *The Postmodern Condition: A Report on Knowledge*. Minneapolis: University of Minnesota Press.

M

Madsen, Anders. Koed. 2012. “Web-Visions as Controversy-Lenses”. Interdisciplinary Science Reviews, 37(1), 51-68.

Mager, Astrid. 2012. “Algorithmic Ideology.” *Information, Communication & Society*, 15:5, 769-787.

Mager, Astrid. 2017. “Search engine imaginary: Visions and values in the co-production of search technology and Europe”. *Social Studies of Science*, Vol. 47(2) 240–262.

Maharaj, Sarat. 2009. “Know-how and No-How: stopgap notes on ‘method’ in visual art as knowledge production.” *Art & Research. A Journal of Ideas, Contexts and Methods*. Volume 2. No. 2. Spring 2009.

Manovich, Lev. 2001. *The Language of New Media*. Cambridge: MIT Press.

Marchiori, M. 1997. "The Quest for Correct Information on the Web: Hyper Search Engines." *Proceedings of the Sixth International World Wide Web Conference (WWW6)*, 1997. Available at https://www.w3.org/People/Massimo/papers/WWW6/.

Marres, Noortje. 2012. "The Experiment in Living". In *Inventive Methods: The Happening of the Social*. Celia Lury, Nina Wakeford. Eds. Abingdon: Routledge.

Marres, Noortje. 2015. "Why Map Issues? On Controversy Analysis as a Digital Method". *Science, Technology, & Human Values* 2015, Vol. 40(5) 655-686. Thousand Oaks: Sage Publishers.

Martin, Reinhold. 2013. *The Organizational Complex: Architecture, Media, and Corporate Space*. Cambridge: MIT Press.

Marx, Gary. 1999. "What's in a Name? Some Reflections on the Sociology of Anonymity". *The Information Society*, 15:2, 99-112.

Marx, Karl. 1993. *Grundrisse: Foundations of the Critique of Political Economy* (1857-8). Notebook VII – The Chapter on Capital. London: Penguin Classics.

Matzner, Tobias. 2018. "Privacy in digital media – an Arendtian approach". Preprint (16 June 2018). Published in French as: La vie privée à l'ère numerique. Perspectives „arend-tiennes". In: /La confiance à l'ère numérique/, ed. by Milad Doueihi and Jacopo Domenicucci. Boulogne-Billancourt: Berger-Levrault. 2017.

Mayer-Schönberger, Viktor; Cukier, Kenneth. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. London: John Murray.

Metz, Cade. 2016. "AI Is Transforming Google Search. The Rest of the Web Is Next". *WIRED.* February 4, 2016. Available here: https://www.wired.com/2016/02/ai-is-changing-the-technology-behind-google-searches/.

Modderkolk, Huib. 2018. "Dutch agencies provide crucial intel about Russia's interference in US-elections". *De Volkskrant*. January 25, 2018. Available here: https://www.volkskrant.nl/wetenschap/dutch-agencies-provide-crucial-intel-about-russia-s-interference-in-us-elections~b4f8111b/.

Mirowski, Philip. 2013. *Never Let a Serious Crisis Go to Waste*. New York: Verso.

Montaigne, Michel de. 1588. *Essais*. Paris: Simon Millanges, Jean Richer.

Morozov, Evgeny. 2013. *To Save Everything, Click Here: The Folly of Technological Solutionism*. *New York, USA: Public Affairs.*

Morozov, Evgeny. 2017. "Big Tech een digital wereldrijk wankelt". *NRC.* Available here: https://www.nrc.nl/nieuws/2017/09/08/big-tech-een-digitaal-wereldrijk-wankelt-12901285-a1572789.

Moulier-Boutang, Yann. 2012. Cognitive Capitalism. Cambridge: Polity Press**.**

Mulligan, Deirdre K. and Daniel S Griffin. 2018. "Rescripting Search to Respect the Right to Truth". 2 *Georgetown Law Review Technology Review*. Rev. 557.

Munkholm, Johan Lau. 2019. Lecture 'Porous Boundaries: Privacy '. September 9 at Kopenhagen University.

Munro, Iain. 2016. "Organizational resistance as a vector of deterritorialization: The case of WikiLeaks and secrecy havens". *Organization* 23(4):567-587

Musk, Elon. 2018. "Interview with Joe Rogan." September 6, 2018. Available here: https://www.youtube.com/watch?v=Ra3fv8gl6NE

N

Nakamura, Lisa. 2014. "Afterword: Blaming, Shaming and the Feminization of Social." Shoshana Magnet and Rachel Dubrofsky (eds). *Media Feminist Surveillance Studies*. Duke University Press. 2015

Ng, Alfred. 2020. "Google is giving data to police based on search keywords, court docs show". *Cne*t. October 8, 2020. Available here: https://www.cnet.com/news/google-is-giving-data-to-police-based-on-search-keywords-court-docs-show/

Nissenbaum, Helen. 1999. "The Meaning of Anonymity in an Information Age". In Spinello, Richard A. and Tavani, Herman T. (eds.) *The Information Society*. 15:141-144. (Reprinted in Readings in CyberEthics 2001 Sudbury: Jones and Bartlett.)

Nissenbaum, Helen and Daniel C. Howe. 2009. (417) "TrackMeNot: Resisting Surveillance in Web Search". In Kerr,I., Lucock, C. and Steeves , C. (Eds). *'Lessons from the Identity Trail: Anonymity, Privacy, and Identity in a Networked Society,* Oxford: Oxford University Press.

Nissenbaum Helen, Daniel C. Howe, Mushon Zer-Aviv. 2014. *Ad Nauseum.* Available at: http://dhowe.github.io/AdNauseam

Noble, Safiya Umoja. 2018a. *Algorithms of Oppression, How Search Engines Reinforce Racism.* New York: New York University Press

Noble, Safia Umoja. 2018b. Algorisky. Season 2. Episode 5. IRL (Online Life is Real Life). https://irlpodcast.org/season2/. Published: March 5, 2018

Noys, Benjamin. 2019. "Rage against the machine? On techno-pessimism and digital resistance." Lecture. May 5, 2019. King's College, London.

O

Ohm, Paul. 2010. *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization,* UCLA Law Review 57: 1701–1777. U of Colorado Law Legal Studies Research Paper No. 9–12. (August 13, 2009) http://ssrn.com/abstract=1450006

Olsen, Stefanie. 2004 “California privacy law kicks in.” *Cnet.com*. July 6, 2004. Available here: http://news.cnet.com/alifornia-privacy-law-kicks-in/2100-1028_3-5258824.html

O’Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Penguin Random House.

O’Neil, Cathy. 2017. “The Ivory Tower Can’t Keep Ignoring Tech”. *New York Times*. November 14, 2017. Available here: https://www.nytimes.com/2017/11/14/opinion/academia-tech-algorithms.html

O’Neill, Patrick Howell. 2020. “How to search the dark net” *The Daily Dot*. Available here: https://www.dailydot.com/debug/how-to-search-the-deep-web/)

Oremus, Will. 2013. “Google’s Big Break”. *Slate*. October 13, 2013. Available here: 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

Ørmen, Jacob. 2016. “Googling the news: Opportunities and challenges in studying news events through Google Search. *Digital Journalism*, Vol. 4, No. 1, 2016, p. 107-124.
Orwell, George. 1949. *1984.* London: Secker & Warburg.

Otlet, Paul and Henri LaFontaine. 1911. *Mundaneum*.

P

Page, Lawrence. 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, Lawrence. 2004. “Method for Scoring Documents in a Linked Database.”
U.S. Patent 6,799,176, filed July 6, 2001, and issued September 28, 2004

Page, Lawrence, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. “The PageRank Citation Ranking: Bringing Order to the Web.” Technical Report, *Stanford InfoLab*. Available here: http://ilpubs.stanford.edu:8090/422/

Parikka, Jussi. 2012. *What is Media Archaeology?* Cambridge, UK: Polity

Pariser, Eli. 2012. *The Filter Bubble.* New York: Penguin Books.

Parker, Martin. 2000. “Manufacturing Bodies: Flesh, Organization, Cyborgs”. In Hassard, J., Holliday, R., Willmott, H. (Eds) *Body and Organization*. London: SAGE Publications. pp. 71-86.

Parker, Martin and Cooper, Robert. 2016b. “Cyborganization: Cinema as Nervous System”. In *For Robert Cooper: Collected Work*. Edited by Gibson Burrell and Martin Parker. New York; Routledge, pp. 236-252.

Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge: Harvard University Press.

Pasquinelli, Matteo. 2009. “Google’s PageRank Algorithm: A diagram of Cognitive Capitalism and the Rentier of the Common Intellect”. In: F. Stalder and K. Becker (eds). *Deep Search: The politics of search beyond Google.* Innsbruck: StudienVerlag.

Pasquinelli, Matteo. 2015. *All Watched over by Algorithms*. Transmediale 2015. Available here: https://transmediale.de/content/all-watched-over-by-algorithms

Peters, John Durham. 2015. *The Marvelous Clouds: Toward a Philosophy of Elemental Media.* Chicago: University of Chicago Press.

Plantin, Jean-Christophe, Carl Lagoze, Paul N. Edwards, Christian Sandvig. 2018. “Infrastructure studies meet platform studies in the age of Google and Facebook”. New Media & Society, Vol. 20(1) 293–310. SAGE Publishers.

Poitras, Laura. 2014. *Citizen Four.* Documentary

Polly, Jean Armour.1992. “Surfing the Internet: An Introduction. Version 2.0.2”. Project Project Gutenberg. Available here: http://www.gutenberg.org/cache/epub/49/pg49-images.html.

Power, Mike. 2014. “What happens when a software bot goes on a darknet shopping spree?” The Guardian. December 5, 2014. Available here: https://www.theguardian.com/technology/2014/dec/05/software-bot-darknet-shopping-spree-random-shopper

Pynchon, Thomas. 2013. *Bleeding Edge*. New York: Penguin.

R

Raley, Rita. 2013. “Dataveillance and Countervailance.” *In “Raw Data” is an Oxymoron*, edited by Lisa Gitelman, 121–45. Cambridge, Mass.: MIT Press.

Rall, Denise N. 2015. “Chapter 11: What Would Kant Think? Testing Truth Claims in Research Traditions, and Proposing Deeper Meanings for the Concept of ‘Search”. In *Web Search Engine Research*. Published online: 08 Mar 2015; 281-307.

Ranciere, Jacques. 2004. *The Politics of aesthetics: the distribution of the sensible*. London: Continuum.

Random Darknet Shopper 2016. MediabitnikGruppe: https://wwwwwwwwwwwwwwwwwwwww.bitnik.org/r/

Rankin, Jennifer. 2018. “Google fined £3.8bn by EU over Android antitrust violations”. *WIRED*. 18 Jul 2018. Available here: https://www.theguardian.com/business/2018/jul/18/google-faces-record-multibillion-fine-from-eu-over-android

Rajaraman, Anand and Jeffrey David Ullman. 2016. “Collaborative Filtering”. From the series ‘Mining of Massive Datasets: Recommender Systems’. Stanford University. Available here: https://www.youtube.com/watch?v=h9gpufJFF-0

Rath, Arun. 2014. *Going dark: the Internet behind the Internet. Available at: http://www.npr.org/blogs/alltechconsidered/2014/05/25/315821415/going-dark-the-internet-behind-the-internet.*

Rayward, W. Boyd. 2002. "Anticipating the Digital World: Paul Otlet and his Paper Internet," Bartels Lecture at the University of Leeds.

Ridgway, Renée. 2015. "Personalisation as currency". *APRJA* (A Peer-Reviewed Journal About). Available here: http://www.aprja.net/?p=2531.

Ridgway, Renée. 2017. "Against a Personalisation of the Self". *Ephemera*. Vol. 17, number 2.

Ridgway, Renée. 2017. "Who's Hacking Whom?" in Hacks, Leaks, and Breaches. Eds E. Gabriella Coleman, Christopher M. Kelty. Available here: http://limn.it/whos-hacking-whom/

Rieder, Bernhard. 2012. "What is in PageRank? A Historical and Conceptual Investigation of a Recursive Status Index".Computational Culture. Volume 2. Available here: http://computationalculture.net/what_is_in_pagerank/

Rieder, Bernhard and Sire, Guillaume. 2014. "Conflicts of interest and incentives to bias: A microeconomic critique of Google's tangled position on the Web". *New Media & Society*. Volume16, Issue 2. 195-211 Sage Publications.

Rogers, Richard. 2009. "The Googlization Question and the Inculpable Engine".In: F. Stalder and K. Becker (eds). *Deep Search: The politics of search beyond Google.* Innsbruck: StudienVerlag.

Rogers, Richard. 2013. *Digital Methods*. Cambridge: MIT Press.

Rosenberg, Daniel. 2013. "Data before the Fact". In L. Gitelman (Ed) *Raw Data Is an Oxymoron*. Cambridge, MIT Press.

Rosenblueth, Arturo, Julian Bigelow,and Norbert Wiener. 1943. "Behavior, Purpose and Teleology". *Philosophy of Science* 10 (Jan. 1943).

Rouvroy, Antoinette. 2013. "The end(s) of critique: Data behaviourism versus due process". In Hildebrandt, Mireille and De Vries, Katja (eds.) *Privacy, due process and the computational turn: the philosophy of law meets the philosophy of technology.* New York: Routledge, Taylor & Francis Group.

Rushkoff, Douglas. 2014. "Herovering van het nu". (Reclaiming the now) *Tegenlicht*. https://www.vpro.nl/programmas/tegenlicht/kijk/afleveringen/2013-2014/de-herovering-van-het-nu.html

S

Sarawagi, Shivang. 2020. “Google Databases: How Do Google Services Store Petabyte-Exabyte Scale Data?”. *8bitmen*. Available here: https://www.8bitmen.com/google-database-how-do-google-services-store-petabyte-exabyte-scale-data/.

Schwartz, Barry. 2016. “Google Penguin doesn’t penalize for bad links – or does it?” *Search Engine Land.* September 28. Available at: http://searchengineland.com/google-penguin-doesnt-penalize-bad-links-259981.

Scott, David. 2014. *Gilbert Simondon's Psychic and Collective Individuation: A Critical Introduction and Guide.* Edinburgh: Edinburgh University Press.

Senior, Jennifer. 2019. “You’re Not Alone When You’re on Google.” *New York Times.* May 17, 2019. Available here: https://www.nytimes.com/2019/05/17/opinion/sunday/google-privacy.html

Share Labs. “Browsing histories”. Available at: https://labs.rs/en/browsing-histories/

Sheikh, Simon. 2009. Objects of Study or Commodification of Knowledge? Remarks on Artistic Research. Art & Research. A Journal of Ideas, Contexts and Methods. Volume 2. No.2. Spring 2009.

Shepard, F. 1873. *Shepard’s Citations*. New York: Frank Shepard Company.

Shullenberger, Geoff. 2016. “Mimesis, Violence, and Facebook: Peter Thiel’s French Connection”. *The Society Pages*. August 13, 2016. Available here: https://thesocietypages.org/cyborgology/2016/08/13/mimesis-violence-and-facebook-peter-thiels-french-connection-full-essay/.

Siegert, Bernard. 1999. *Relays*. Stanford: Stanford University Press.

Smythe, Dallas. W. 1981. “On the audience commodity and its work”, in M. G. Durham and D. M. Kellner (eds) *Media and cultural studies* (230–56). Malden, MA: Blackwell.

Solove, Daniel. 2011. “The Virtues of Anonymity”. *New York Times*. Op-Ed. Available here: http://www.nytimes.com/roomfordebate/2011/06/21/youre-mad-youre-on-youtube/the-virtues-of-anonymity.

Soon, Winnie. 2016. “Executing Liveness: An examination of the live dimension of code inter-actions in software (art) practice”. PhD thesis. *School of Communication and Culture, Aarhus University*, 2016.

Spitters, Martijn, Stefan Verbruggen and Mark van Staalduinen. 2014. “Towards a Comprehensive Insight into the Thematic Organization of the Tor Hidden Services”, presented at 2014 *IEEE Joint Intelligence and Security Informatics Conference*. Los Angeles, CA, USA; 15 -17 Dec 2014.

Srnicek, Nick. 2016. *Platform Capitalism*. Cambridge: Polity Press.

Stackoverflow. 2020. “What database does Google use?” Available here: https://stackoverflow.com/questions/362956/what-database-does-google-use.

Stalder, Felix. 2013. Digital Solidarity. Mute. Berlin: Anagram Books

Stalder, Felix. 2010. “Autonomy and Control in the Era of Post-Privacy”. *Open. Cahier on Art and the Public Domain.* # 19: Beyond Privacy. New Notions of the Private and Public Domains.

Stalder, Felix. 2012. “Between Democracy and Spectacle: The Front-End and the Back-End of the Social Web”. In M. Mandiberg (ed.), *The Social Media Reader,* (pp. 242-256). New York: New York University Press.

Stalder, Felix and Christine Mayer. 2009. “Second Index”. In: F. Stalder and K. Becker (eds). *Deep Search: The politics of search beyond Google.* Innsbruck: StudienVerlag.

Stark, David. 2009. *The Sense of Dissonance*. Princeton: Princeton University Press.

Stephens-Davidowitz, Seth. 2017. *Everybody Lies: What the Internet Can Tell Us About Who We Really Are.* New York: Harper Collins

Stephens-Davidowitz, Seth. 2017b. “Everybody lies: how Google search reveals our darkest secrets”. *The Guardian*. 9 July 2017. Available here: https://www.theguardian.com/technology/2017/jul/09/everybody-lies-how-google-reveals-darkest-secrets-seth-stephens-davidowitz.

Stiegler, Bernhard. 1998. *Technics and Time, 1. The Fall of Epimetheus*. Stanford: Stanford University Press.

Suchman, Lucy. 2012. “Configuration”. In *Inventive Methods: The Happening of the Social*. Celia Lury, Nina Wakeford. Eds. Abingdon: Routledge.

Suler John. 2005. “The Online Disinhibition Effect”. *Contemporary Media Forum International Journal of Applied Psychoanalytic Studies*, Vol. 2, No. 2. Whurr Publishers Ltd.

Sullivan, Danny. 2010. “Dear Bing, We have 10000 Ranking Signals to your 1000. Love Google.” *Search Engine Land.* November 11. Available here: http://searchengineland.com/bing-10000-ranking-signals-google-55473

Sullivan, Danny. 2016. “FAQ: All about the Google RankBrain algorithm”. *Search Engine Land.* June 23, 2016. Available here: http://searchengineland.com/faq-all-about-the-new-google-rankbrain-algorithm-234440.

Sweeney, Lantaya. 2002. “k-anonymity: a model for protecting privacy”. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10. 2002; 557-570.

T

Tait, John I. 2007. “Karen Spärck Jones”. *Computational Linguistics* .Volume 33, Number 3. Cambridge: MIT Press.

Tantner, Anton. 2014. “Before Google: A Pre-History of Search Engines in Analogue Times”. René König & Miriam Rasch, (Eds), Society of the Query Reader #9: *Reflections on Web Search*, Amsterdam: Institute of Network Cultures.

Tantner, Anton. 2015. *Die ersten Suchmaschinen: Adressbüros, Fragämter, Intelligenz-Comptoirs*. Berlin: Verlag Klaus Wagenbach.

The Tor Project. 2017. Available here: https://www.torproject.org/

Thylstrup, Nanna. 2014. “Archival Shadows in the Digital Age”. *Nordisk Tidsskrift for Informationsvidenskab og Kulturformidling*, årg. 3, nr. 2/3, 2014.

Thylstrup, Nanna. 2019. *The Politics of Mass Digitization.* Cambridge: MIT Press

Tiqqun. 2001. “The Cybernetic Hypothesis”. Intervention Series, 28.
Semiotext(e). Cambridge: MIT Press.

Trevisan, Fillipo. 2014. “Search Engines-From social science objects to academic inquiry tools also provides inspiration”. *First Monday.* Volume 19, Number 11 - 3 November 2014. Available here:http://firstmonday.org/ojs/index.php/fm/rt/printerFriendly/5237/4157Racial

Tripodi, Francesca. 2017. “Googling for Truth. Chapter in ‚Analyzing Scriptural Inference in Conservative News Practices”. *Data & Society Research* Institute. Available here: https://datasociety.net/research/media-manipulation/

Tufekci, Zeynep. 2015. “Algorithmic Harms Beyond Facebook And Google: Emergent Challenges of Computational Agency”. *Journal on Telecommunication & High Technology Law*. Volume 13, 203.

Tungate, Mark. 2007. *“Pioneers of Persuasion—The Duly Authorized agent” Adland: A Global History of Advertising. Kogan. March, 2007. Page. pp. 7–8.*

Turk, Victoria. 2015. “When Algorithms Are Sexist.” *Motherboard.* March 20. Available at: https://motherboard.vice.com/en_us/article/ezvkee/when-algorithms-are-sexist

Turkle, Sherry. 2005. *The Second Self*. Cambridge Mass: MIT Press.

Turner, Fred. 2006. *From Counterculture to Cyberculture*: *Stewart Brand, the Whole Earth Network, and the Rise of Digital Utopianism*. Chicago. University of Chicago Press.

U

Ullman, Ellen. 1997. *Close to the Machine: Technophilia and Its Discontents*. San Francisco, CA: City Lights Books.

V

Vaidhyanathan, Siva. 2011. *Googlization of everything (And why we should worry).* Oakland: University of California Press.

Van Couvering, Elizabeth. 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 der Nagel, Emily and Frith, Jonathan. 2015. "Anonymity, pseudonymity, and the agency of online identity: Examining the social practices of r/Gonewild". *First Monday*. Volume 20, Number 3 - 2 March 2015.

Vismann, Cornelia. 2013. "Cultural Techniques and Sovereignty". *Theory, Culture & Society*. 30(6):83-93.

W

Wang, Chia-Ling. 2008. **"**What is "critical" in critical ethnography?" Presented at *Philosophy of Education Society* Great Britain annual conference (28th-30th March). Oxford: New College.

Warren, Carol and Barbara Laslett. 1977. "Privacy and Secrecy: A Conceptual Comparison." *Journal of Social Issues* 33: 43–51.

Warren, Samuel D. & Brandeis, Louis D. 1890. "Right to Privacy". *Harvard Law Review*, Vol. 4, No. 5. (Dec. 15, 1890), pp. 193-220.

Waters, Richard. 2014. "FT interview with Google co-founder and CEO Larry Page". *Financial Times*. October 31, 2014. Available here: https://www.ft.com/content/3173f19e-5fbc-11e4-8c27-00144feabdc0.

Weiser, Mark. 1991. "The Computer for the 21st Century". *Scientific American* Special Issue on Communications, Computers, and Networks, September, 1991

Weiskopf, Richard. 2018. "Algorithmic decision making, 'ghostly demarcations' and the space of ethics" Paper. *EGOS Conference*, Tallinn.

Wells, H.G. 1937. "World Brain: The Idea of a Permanent World Encyclopaedia". Encyclopédie Française, August, 1937. Available here: https://sherlock.ischool.berkeley.edu/wells/world_brain.html

Weltevrede, Esther; Helmond, Anne; Carolin Gerlitz. 2014. "The Politics of Real-time: A Device Perspective on Social Media Platforms and Search Engines". *Theory, Culture and Society*. Vol 31, Issue 6.

Wiener, Norbert. 1961. *Cybernetics or control and communication in the animal and the machine*. Cambridge: MIT Press. Second Edition.

Winter, Philipp, Richard Köwer, Martin Mulazzani, Markus Huber. Sebastian Schrittwieser Stefan Lindskog, Edgar Weippl. 2014. "Spoiled Onions: Exposing Malicious Tor Exit Relays." In: *Privacy Enhancing Technologies Symposium*. New York City: Springer

Z

Zaunseder, Andreas. 2018. “The darknet is not a hellhole, it’s an answer to internet privacy”. *The Conversation*. August 16, 2018. Available here: https://theconversation.com/the-darknet-is-not-a-hellhole-its-an-answer-to-internet-privacy-101420?

Zetter, Kim. 2011. “*New Yorker* Sheds New Light on NSA’s Warrantless Wiretapping and Data Mining”. *WIRED*. May 5, 2011. Available here: https://www.wired.com/2011/05/new-yorker-on-thomas-drake/.

Zetter, Kim. 2015. “Darpa Is Developing a Search Engine for the Dark Web”. *WIRED*. Oct 2, 2015. http://www.wired.com/2015/02/darpa-memex-dark-web/.

Zimmerman, Phil. 1991. “Why I Wrote PGP”. Available here: https://www.philzimmermann.com/EN/essays/WhyIWrotePGP.html

Zizek, Slavoj. 2009. *The Ticklish Subject: The Absent Centre of Political Ontology.* Brooklyn: Verso Books

Zuboff, Shoshana. 1981. “The Psychological and Organizational Implications of Computer-Mediated Work”. MIT Working Paper.

Zuboff, Shoshana. 2015. “Big Other: surveillance capitalism and the prospects of an information civilization”. *Journal of Information Technology* 30:75-89.

Zuboff, Shoshana. 2017. “Secrets of Surveillance Capitalism”. Lecture. Queens University, Canada. November 2017.

TITLER I PH.D.SERIEN:

2004

1. Martin Grieger
Internet-based Electronic Marketplaces and Supply Chain Management

2. Thomas Basbøll
LIKENESS
A Philosophical Investigation

3. Morten Knudsen
Beslutningens vaklen
En systemteoretisk analyse of moderniseringen af et amtskommunalt sundhedsvæsen 1980-2000

4. Lars Bo Jeppesen
Organizing Consumer Innovation
A product development strategy that is based on online communities and allows some firms to benefit from a distributed process of innovation by consumers

5. Barbara Dragsted
SEGMENTATION IN TRANSLATION AND TRANSLATION MEMORY SYSTEMS
An empirical investigation of cognitive segmentation and effects of integrating a TM system into the translation process

6. Jeanet Hardis
Sociale partnerskaber
Et socialkonstruktivistisk casestudie af partnerskabsaktørers virkelighedsopfattelse mellem identitet og legitimitet

7. Henriette Hallberg Thygesen
System Dynamics in Action

8. Carsten Mejer Plath
Strategisk Økonomistyring

9. Annemette Kjærgaard
Knowledge Management as Internal Corporate Venturing
– a Field Study of the Rise and Fall of a Bottom-Up Process

10. Knut Arne Hovdal
De profesjonelle i endring
Norsk ph.d., ej til salg gennem Samfundslitteratur

11. Søren Jeppesen
Environmental Practices and Greening Strategies in Small Manufacturing Enterprises in South Africa
– A Critical Realist Approach

12. Lars Frode Frederiksen
Industriel forskningsledelse
– på sporet af mønstre og samarbejde i danske forskningsintensive virksomheder

13. Martin Jes Iversen
The Governance of GN Great Nordic
– in an age of strategic and structural transitions 1939-1988

14. Lars Pynt Andersen
The Rhetorical Strategies of Danish TV Advertising
A study of the first fifteen years with special emphasis on genre and irony

15. Jakob Rasmussen
Business Perspectives on E-learning

16. Sof Thrane
The Social and Economic Dynamics of Networks
– a Weberian Analysis of Three Formalised Horizontal Networks

17. Lene Nielsen
Engaging Personas and Narrative Scenarios – a study on how a user-centered approach influenced the perception of the design process in the e-business group at AstraZeneca

18. S.J Valstad
Organisationsidentitet
Norsk ph.d., ej til salg gennem Samfundslitteratur

19. Thomas Lyse Hansen
Six Essays on Pricing and Weather risk in Energy Markets

20. Sabine Madsen
Emerging Methods – An Interpretive Study of ISD Methods in Practice

21. Evis Sinani
The Impact of Foreign Direct Investment on Efficiency, Productivity Growth and Trade: An Empirical Investigation

22. Bent Meier Sørensen
Making Events Work Or, How to Multiply Your Crisis

23. Pernille Schnoor
Brand Ethos
Om troværdige brand- og virksomhedsidentiteter i et retorisk og diskursteoretisk perspektiv

24. Sidsel Fabech
Von welchem Österreich ist hier die Rede?
Diskursive forhandlinger og magtkampe mellem rivaliserende nationale identitetskonstruktioner i østrigske pressediskurser

25. Klavs Odgaard Christensen
Sprogpolitik og identitetsdannelse i flersprogede forbundsstater
Et komparativt studie af Schweiz og Canada

26. Dana B. Minbaeva
Human Resource Practices and Knowledge Transfer in Multinational Corporations

27. Holger Højlund
Markedets politiske fornuft
Et studie af velfærdens organisering i perioden 1990-2003

28. Christine Mølgaard Frandsen
A.s erfaring
Om mellemværendets praktik i en transformation af mennesket og subjektiviteten

29. Sine Nørholm Just
The Constitution of Meaning – A Meaningful Constitution?
Legitimacy, identity, and public opinion in the debate on the future of Europe

2005

1. Claus J. Varnes
Managing product innovation through rules – The role of formal and structured methods in product development

2. Helle Hedegaard Hein
Mellem konflikt og konsensus – Dialogudvikling på hospitalsklinikker

3. Axel Rosenø
Customer Value Driven Product Innovation – A Study of Market Learning in New Product Development

4. Søren Buhl Pedersen
Making space
An outline of place branding

5. Camilla Funck Ellehave
Differences that Matter
An analysis of practices of gender and organizing in contemporary workplaces

6. Rigmor Madeleine Lond
Styring af kommunale forvaltninger

7. Mette Aagaard Andreassen
Supply Chain versus Supply Chain Benchmarking as a Means to Managing Supply Chains

8. Caroline Aggestam-Pontoppidan
From an idea to a standard
The UN and the global governance of accountants' competence

9. Norsk ph.d.

10. Vivienne Heng Ker-ni
An Experimental Field Study on the

Effectiveness of Grocer Media Advertising
Measuring Ad Recall and Recognition, Purchase Intentions and Short-Term Sales

11. Allan Mortensen
Essays on the Pricing of Corporate Bonds and Credit Derivatives

12. Remo Stefano Chiari
Figure che fanno conoscere
Itinerario sull'idea del valore cognitivo e espressivo della metafora e di altri tropi da Aristotele e da Vico fino al cognitivismo contemporaneo

13. Anders McIlquham-Schmidt
Strategic Planning and Corporate Performance
An integrative research review and a meta-analysis of the strategic planning and corporate performance literature from 1956 to 2003

14. Jens Geersbro
The TDF – PMI Case
Making Sense of the Dynamics of Business Relationships and Networks

15 Mette Andersen
Corporate Social Responsibility in Global Supply Chains
Understanding the uniqueness of firm behaviour

16. Eva Boxenbaum
Institutional Genesis: Micro – Dynamic Foundations of Institutional Change

17. Peter Lund-Thomsen
Capacity Development, Environmental Justice NGOs, and Governance: The Case of South Africa

18. Signe Jarlov
Konstruktioner af offentlig ledelse

19. Lars Stæhr Jensen
Vocabulary Knowledge and Listening Comprehension in English as a Foreign Language
An empirical study employing data elicited from Danish EFL learners

20. Christian Nielsen
Essays on Business Reporting
Production and consumption of strategic information in the market for information

21. Marianne Thejls Fischer
Egos and Ethics of Management Consultants

22. Annie Bekke Kjær
Performance management i Proces-innovation
– belyst i et social-konstruktivistisk perspektiv

23. Suzanne Dee Pedersen
GENTAGELSENS METAMORFOSE
Om organisering af den kreative gøren i den kunstneriske arbejdspraksis

24. Benedikte Dorte Rosenbrink
Revenue Management
Økonomiske, konkurrencemæssige & organisatoriske konsekvenser

25. Thomas Riise Johansen
Written Accounts and Verbal Accounts
The Danish Case of Accounting and Accountability to Employees

26. Ann Fogelgren-Pedersen
The Mobile Internet: Pioneering Users' Adoption Decisions

27. Birgitte Rasmussen
Ledelse i fællesskab – de tillidsvalgtes fornyende rolle

28. Gitte Thit Nielsen
Remerger
– skabende ledelseskræfter i fusion og opkøb

29. Carmine Gioia
A MICROECONOMETRIC ANALYSIS OF MERGERS AND ACQUISITIONS

30. Ole Hinz
Den effektive forandringsleder: pilot, pædagog eller politiker?
Et studie i arbejdslederes meningstilskrivninger i forbindelse med vellykket gennemførelse af ledelsesinitierede forandringsprojekter

31. Kjell-Åge Gotvassli
Et praksisbasert perspektiv på dynamiske
læringsnettverk i toppidretten
Norsk ph.d., ej til salg gennem Samfundslitteratur

32. Henriette Langstrup Nielsen
Linking Healthcare
An inquiry into the changing performances of web-based technology for asthma monitoring

33. Karin Tweddell Levinsen
Virtuel Uddannelsespraksis
Master i IKT og Læring – et casestudie i hvordan proaktiv proceshåndtering kan forbedre praksis i virtuelle læringsmiljøer

34. Anika Liversage
Finding a Path
Labour Market Life Stories of Immigrant Professionals

35. Kasper Elmquist Jørgensen
Studier i samspillet mellem stat og erhvervsliv i Danmark under
1. verdenskrig

36. Finn Janning
A DIFFERENT STORY
Seduction, Conquest and Discovery

37. Patricia Ann Plackett
Strategic Management of the Radical Innovation Process
Leveraging Social Capital for Market Uncertainty Management

2006

1. Christian Vintergaard
Early Phases of Corporate Venturing

2. Niels Rom-Poulsen
Essays in Computational Finance

3. Tina Brandt Husman
Organisational Capabilities, Competitive Advantage & Project-Based Organisations
The Case of Advertising and Creative Good Production

4. Mette Rosenkrands Johansen
Practice at the top
– how top managers mobilise and use non-financial performance measures

5. Eva Parum
Corporate governance som strategisk kommunikations- og ledelsesværktøj

6. Susan Aagaard Petersen
Culture's Influence on Performance Management: The Case of a Danish Company in China

7. Thomas Nicolai Pedersen
The Discursive Constitution of Organizational Governance – Between unity and differentiation
The Case of the governance of environmental risks by World Bank environmental staff

8. Cynthia Selin
Volatile Visions: Transactons in Anticipatory Knowledge

9. Jesper Banghøj
Financial Accounting Information and Compensation in Danish Companies

10. Mikkel Lucas Overby
Strategic Alliances in Emerging High-Tech Markets: What's the Difference and does it Matter?

11. Tine Aage
External Information Acquisition of Industrial Districts and the Impact of Different Knowledge Creation Dimensions

A case study of the Fashion and Design Branch of the Industrial District of Montebelluna, NE Italy

12. Mikkel Flyverbom
Making the Global Information Society Governable
On the Governmentality of Multi-Stakeholder Networks

13. Anette Grønning
Personen bag
Tilstedeværer i e-mail som interaktionsform mellem kunde og medarbejder i dansk forsikringskontekst

14. Jørn Helder
One Company – One Language?
The NN-case

15. Lars Bjerregaard Mikkelsen
Differing perceptions of customer value
Development and application of a tool for mapping perceptions of customer value at both ends of customer-supplier dyads in industrial markets

16. Lise Granerud
Exploring Learning
Technological learning within small manufacturers in South Africa

17. Esben Rahbek Pedersen
Between Hopes and Realities: Reflections on the Promises and Practices of Corporate Social Responsibility (CSR)

18. Ramona Samson
The Cultural Integration Model and European Transformation.
The Case of Romania

2007

1. Jakob Vestergaard
Discipline in The Global Economy
Panopticism and the Post-Washington Consensus

2. Heidi Lund Hansen
Spaces for learning and working
A qualitative study of change of work, management, vehicles of power and social practices in open offices

3. Sudhanshu Rai
Exploring the internal dynamics of software development teams during user analysis
A tension enabled Institutionalization Model; "Where process becomes the objective"

4. Norsk ph.d.
Ej til salg gennem Samfundslitteratur

5. Serden Ozcan
EXPLORING HETEROGENEITY IN ORGANIZATIONAL ACTIONS AND OUTCOMES
A Behavioural Perspective

6. Kim Sundtoft Hald
Inter-organizational Performance Measurement and Management in Action
– An Ethnography on the Construction of Management, Identity and Relationships

7. Tobias Lindeberg
Evaluative Technologies
Quality and the Multiplicity of Performance

8. Merete Wedell-Wedellsborg
Den globale soldat
Identitetsdannelse og identitetsledelse i multinationale militære organisationer

9. Lars Frederiksen
Open Innovation Business Models
Innovation in firm-hosted online user communities and inter-firm project ventures in the music industry
– A collection of essays

10. Jonas Gabrielsen
Retorisk toposlære – fra statisk 'sted' til persuasiv aktivitet

11. Christian Moldt-Jørgensen
Fra meningsløs til meningsfuld evaluering.
Anvendelsen af studentertilfredshedsmålinger på de korte og mellemlange videregående uddannelser set fra et psykodynamisk systemperspektiv

12. Ping Gao
Extending the application of actor-network theory
Cases of innovation in the telecommunications industry

13. Peter Mejlby
Frihed og fængsel, en del af den samme drøm?
Et phronetisk baseret casestudie af frigørelsens og kontrollens sameksistens i værdibaseret ledelse!

14. Kristina Birch
Statistical Modelling in Marketing

15. Signe Poulsen
Sense and sensibility:
The language of emotional appeals in insurance marketing

16. Anders Bjerre Trolle
Essays on derivatives pricing and dynamic asset allocation

17. Peter Feldhütter
Empirical Studies of Bond and Credit Markets

18. Jens Henrik Eggert Christensen
Default and Recovery Risk Modeling and Estimation

19. Maria Theresa Larsen
Academic Enterprise: A New Mission for Universities or a Contradiction in Terms?
Four papers on the long-term implications of increasing industry involvement and commercialization in academia

20. Morten Wellendorf
Postimplementering af teknologi i den offentlige forvaltning
Analyser af en organisations kontinuerlige arbejde med informationsteknologi

21. Ekaterina Mhaanna
Concept Relations for Terminological Process Analysis

22. Stefan Ring Thorbjørnsen
Forsvaret i forandring
Et studie i officerers kapabiliteter under påvirkning af omverdenens forandringspres mod øget styring og læring

23. Christa Breum Amhøj
Det selvskabte medlemskab om managementstaten, dens styringsteknologier og indbyggere

24. Karoline Bromose
Between Technological Turbulence and Operational Stability
– An empirical case study of corporate venturing in TDC

25. Susanne Justesen
Navigating the Paradoxes of Diversity in Innovation Practice
– A Longitudinal study of six very different innovation processes – in practice

26. Luise Noring Henler
Conceptualising successful supply chain partnerships
– Viewing supply chain partnerships from an organisational culture perspective

27. Mark Mau
Kampen om telefonen
Det danske telefonvæsen under den tyske besættelse 1940-45

28. Jakob Halskov
The semiautomatic expansion of existing terminological ontologies using knowledge patterns discovered

on the WWW – an implementation and evaluation

29. Gergana Koleva
European Policy Instruments Beyond Networks and Structure: The Innovative Medicines Initiative

30. Christian Geisler Asmussen
Global Strategy and International Diversity: A Double-Edged Sword?

31. Christina Holm-Petersen
Stolthed og fordom
Kultur- og identitetsarbejde ved skabelsen af en ny sengeafdeling gennem fusion

32. Hans Peter Olsen
Hybrid Governance of Standardized States
Causes and Contours of the Global Regulation of Government Auditing

33. Lars Bøge Sørensen
Risk Management in the Supply Chain

34. Peter Aagaard
Det unikkes dynamikker
De institutionelle mulighedsbetingelser bag den individuelle udforskning i professionelt og frivilligt arbejde

35. Yun Mi Antorini
Brand Community Innovation
An Intrinsic Case Study of the Adult Fans of LEGO Community

36. Joachim Lynggaard Boll
Labor Related Corporate Social Performance in Denmark
Organizational and Institutional Perspectives

2008

1. Frederik Christian Vinten
Essays on Private Equity

2. Jesper Clement
Visual Influence of Packaging Design on In-Store Buying Decisions

3. Marius Brostrøm Kousgaard
Tid til kvalitetsmåling?
– Studier af indrulleringsprocesser i forbindelse med introduktionen af kliniske kvalitetsdatabaser i speciallægepraksissektoren

4. Irene Skovgaard Smith
Management Consulting in Action
Value creation and ambiguity in client-consultant relations

5. Anders Rom
Management accounting and integrated information systems
How to exploit the potential for management accounting of information technology

6. Marina Candi
Aesthetic Design as an Element of Service Innovation in New Technology-based Firms

7. Morten Schnack
Teknologi og tværfaglighed
– en analyse af diskussionen omkring indførelse af EPJ på en hospitalsafdeling

8. Helene Balslev Clausen
Juntos pero no revueltos – un estudio sobre emigrantes norteamericanos en un pueblo mexicano

9. Lise Justesen
Kunsten at skrive revisionsrapporter. En beretning om forvaltningsrevisionens beretninger

10. Michael E. Hansen
The politics of corporate responsibility: CSR and the governance of child labor and core labor rights in the 1990s

11. Anne Roepstorff
Holdning for handling – en etnologisk undersøgelse af Virksomheders Sociale Ansvar/CSR

12. Claus Bajlum
Essays on Credit Risk and Credit Derivatives

13. Anders Bojesen
The Performative Power of Competence – an Inquiry into Subjectivity and Social Technologies at Work

14. Satu Reijonen
Green and Fragile
A Study on Markets and the Natural Environment

15. Ilduara Busta
Corporate Governance in Banking
A European Study

16. Kristian Anders Hvass
A Boolean Analysis Predicting Industry Change: Innovation, Imitation & Business Models
The Winning Hybrid: A case study of isomorphism in the airline industry

17. Trine Paludan
De uvidende og de udviklingsparate
Identitet som mulighed og restriktion blandt fabriksarbejdere på det aftayloriserede fabriksgulv

18. Kristian Jakobsen
Foreign market entry in transition economies: Entry timing and mode choice

19. Jakob Elming
Syntactic reordering in statistical machine translation

20. Lars Brømsøe Termansen
Regional Computable General Equilibrium Models for Denmark
Three papers laying the foundation for regional CGE models with agglomeration characteristics

21. Mia Reinholt
The Motivational Foundations of Knowledge Sharing

22. Frederikke Krogh-Meibom
The Co-Evolution of Institutions and Technology
– A Neo-Institutional Understanding of Change Processes within the Business Press – the Case Study of Financial Times

23. Peter D. Ørberg Jensen
OFFSHORING OF ADVANCED AND HIGH-VALUE TECHNICAL SERVICES: ANTECEDENTS, PROCESS DYNAMICS AND FIRMLEVEL IMPACTS

24. Pham Thi Song Hanh
Functional Upgrading, Relational Capability and Export Performance of Vietnamese Wood Furniture Producers

25. Mads Vangkilde
Why wait?
An Exploration of first-mover advantages among Danish e-grocers through a resource perspective

26. Hubert Buch-Hansen
Rethinking the History of European Level Merger Control
A Critical Political Economy Perspective

2009

1. Vivian Lindhardsen
From Independent Ratings to Communal Ratings: A Study of CWA Raters' Decision-Making Behaviours

2. Guðrið Weihe
Public-Private Partnerships: Meaning and Practice

3. Chris Nøkkentved
Enabling Supply Networks with Collaborative Information Infrastructures
An Empirical Investigation of Business Model Innovation in Supplier Relationship Management

4. Sara Louise Muhr
Wound, Interrupted – On the Vulnerability of Diversity Management

5. Christine Sestoft
Forbrugeradfærd i et Stats- og Livsformsteoretisk perspektiv

6. Michael Pedersen
Tune in, Breakdown, and Reboot: On the production of the stress-fit self-managing employee

7. Salla Lutz
Position and Reposition in Networks – Exemplified by the Transformation of the Danish Pine Furniture Manufacturers

8. Jens Forssbæck
Essays on market discipline in commercial and central banking

9. Tine Murphy
Sense from Silence – A Basis for Organised Action
How do Sensemaking Processes with Minimal Sharing Relate to the Reproduction of Organised Action?

10. Sara Malou Strandvad
Inspirations for a new sociology of art: A sociomaterial study of development processes in the Danish film industry

11. Nicolaas Mouton
On the evolution of social scientific metaphors:
A cognitive-historical enquiry into the divergent trajectories of the idea that collective entities – states and societies, cities and corporations – are biological organisms.

12. Lars Andreas Knutsen
Mobile Data Services:
Shaping of user engagements

13. Nikolaos Theodoros Korfiatis
Information Exchange and Behavior
A Multi-method Inquiry on Online Communities

14. Jens Albæk
Forestillinger om kvalitet og tværfaglighed på sygehuse
– skabelse af forestillinger i læge- og plejegrupperne angående relevans af nye idéer om kvalitetsudvikling gennem tolkningsprocesser

15. Maja Lotz
The Business of Co-Creation – and the Co-Creation of Business

16. Gitte P. Jakobsen
Narrative Construction of Leader Identity in a Leader Development Program Context

17. Dorte Hermansen
"Living the brand" som en brandorienteret dialogisk praxis:
Om udvikling af medarbejdernes brandorienterede dømmekraft

18. Aseem Kinra
Supply Chain (logistics) Environmental Complexity

19. Michael Nørager
How to manage SMEs through the transformation from non innovative to innovative?

20. Kristin Wallevik
Corporate Governance in Family Firms
The Norwegian Maritime Sector

21. Bo Hansen Hansen
Beyond the Process
Enriching Software Process Improvement with Knowledge Management

22. Annemette Skot-Hansen
Franske adjektivisk afledte adverbier, der tager præpositionssyntagmer indledt med præpositionen à som argumenter
En valensgrammatisk undersøgelse

23. Line Gry Knudsen
Collaborative R&D Capabilities
In Search of Micro-Foundations

24. Christian Scheuer
Employers meet employees
Essays on sorting and globalization

25. Rasmus Johnsen
The Great Health of Melancholy
A Study of the Pathologies of Performativity

26. Ha Thi Van Pham
Internationalization, Competitiveness Enhancement and Export Performance of Emerging Market Firms:
Evidence from Vietnam

27. Henriette Balieu
Kontrolbegrebets betydning for kausativalternationen i spansk
En kognitiv-typologisk analyse

2010

1. Yen Tran
Organizing Innovationin Turbulent Fashion Market
Four papers on how fashion firms create and appropriate innovation value

2. Anders Raastrup Kristensen
Metaphysical Labour
Flexibility, Performance and Commitment in Work-Life Management

3. Margrét Sigrún Sigurdardottir
Dependently independent
Co-existence of institutional logics in the recorded music industry

4. Ásta Dis Óladóttir
Internationalization from a small domestic base:
An empirical analysis of Economics and Management

5. Christine Secher
E-deltagelse i praksis – politikernes og forvaltningens medkonstruktion og konsekvenserne heraf

6. Marianne Stang Våland
What we talk about when we talk about space:
End User Participation between Processes of Organizational and Architectural Design

7. Rex Degnegaard
Strategic Change Management
Change Management Challenges in the Danish Police Reform

8. Ulrik Schultz Brix
Værdi i rekruttering – den sikre beslutning
En pragmatisk analyse af perception og synliggørelse af værdi i rekrutterings- og udvælgelsesarbejdet

9. Jan Ole Similä
Kontraktsledelse
Relasjonen mellom virksomhetsledelse og kontraktshåndtering, belyst via fire norske virksomheter

10. Susanne Boch Waldorff
Emerging Organizations: In between local translation, institutional logics and discourse

11. Brian Kane
Performance Talk
Next Generation Management of Organizational Performance

12. Lars Ohnemus
Brand Thrust: Strategic Branding and Shareholder Value
An Empirical Reconciliation of two Critical Concepts

13. Jesper Schlamovitz
Håndtering af usikkerhed i film- og byggeprojekter

14. Tommy Moesby-Jensen
Det faktiske livs forbindtlighed
Førsokratisk informeret, ny-aristotelisk ἦθος*-tænkning hos Martin Heidegger*

15. Christian Fich
Two Nations Divided by Common Values
French National Habitus and the Rejection of American Power

16. Peter Beyer
Processer, sammenhængskraft og fleksibilitet
Et empirisk casestudie af omstillingsforløb i fire virksomheder

17. Adam Buchhorn
Markets of Good Intentions
Constructing and Organizing Biogas Markets Amid Fragility and Controversy

18. Cecilie K. Moesby-Jensen
Social læring og fælles praksis
Et mixed method studie, der belyser læringskonsekvenser af et lederkursus for et praksisfællesskab af offentlige mellemledere

19. Heidi Boye
Fødevarer og sundhed i senmodernismen
– En indsigt i hyggefænomenet og de relaterede fødevarepraksisser

20. Kristine Munkgård Pedersen
Flygtige forbindelser og midlertidige mobiliseringer
Om kulturel produktion på Roskilde Festival

21. Oliver Jacob Weber
Causes of Intercompany Harmony in Business Markets – An Empirical Investigation from a Dyad Perspective

22. Susanne Ekman
Authority and Autonomy
Paradoxes of Modern Knowledge Work

23. Anette Frey Larsen
Kvalitetsledelse på danske hospitaler
– Ledelsernes indflydelse på introduktion og vedligeholdelse af kvalitetsstrategier i det danske sundhedsvæsen

24. Toyoko Sato
Performativity and Discourse: Japanese Advertisements on the Aesthetic Education of Desire

25. Kenneth Brinch Jensen
Identifying the Last Planner System
Lean management in the construction industry

26. Javier Busquets
Orchestrating Network Behavior for Innovation

27. Luke Patey
The Power of Resistance: India's National Oil Company and International Activism in Sudan

28. Mette Vedel
Value Creation in Triadic Business Relationships. Interaction, Interconnection and Position

29. Kristian Tørning
Knowledge Management Systems in Practice – A Work Place Study

30. Qingxin Shi
An Empirical Study of Thinking Aloud Usability Testing from a Cultural Perspective

31. Tanja Juul Christiansen
Corporate blogging: Medarbejderes kommunikative handlekraft

32. Malgorzata Ciesielska
Hybrid Organisations.
A study of the Open Source – business setting

33. Jens Dick-Nielsen
Three Essays on Corporate Bond Market Liquidity

34. Sabrina Speiermann
Modstandens Politik
Kampagnestyring i Velfærdsstaten.
En diskussion af trafikkampagners styringspotentiale

35. Julie Uldam
Fickle Commitment. Fostering political engagement in 'the flighty world of online activism'

36. Annegrete Juul Nielsen
Traveling technologies and transformations in health care

37. Athur Mühlen-Schulte
Organising Development
Power and Organisational Reform in the United Nations Development Programme

38. Louise Rygaard Jonas
Branding på butiksgulvet
Et case-studie af kultur- og identitetsarbejdet i Kvickly

2011

1. Stefan Fraenkel
Key Success Factors for Sales Force Readiness during New Product Launch
A Study of Product Launches in the Swedish Pharmaceutical Industry

2. Christian Plesner Rossing
International Transfer Pricing in Theory and Practice

3. Tobias Dam Hede
Samtalekunst og ledelsesdisciplin – en analyse af coachingsdiskursens genealogi og governmentality

4. Kim Pettersson
Essays on Audit Quality, Auditor Choice, and Equity Valuation

5. Henrik Merkelsen
The expert-lay controversy in risk research and management. Effects of institutional distances. Studies of risk definitions, perceptions, management and communication

6. Simon S. Torp
Employee Stock Ownership:
Effect on Strategic Management and Performance

7. Mie Harder
Internal Antecedents of Management Innovation

8. Ole Helby Petersen
Public-Private Partnerships: Policy and Regulation – With Comparative and Multi-level Case Studies from Denmark and Ireland

9. Morten Krogh Petersen
'Good' Outcomes. Handling Multiplicity in Government Communication

10. Kristian Tangsgaard Hvelplund
Allocation of cognitive resources in translation - an eye-tracking and key-logging study

11. Moshe Yonatany
The Internationalization Process of Digital Service Providers

12. Anne Vestergaard
Distance and Suffering
Humanitarian Discourse in the age of Mediatization

13. Thorsten Mikkelsen
Personligsheds indflydelse på forretningsrelationer

14. Jane Thostrup Jagd
Hvorfor fortsætter fusionsbølgen udover "the tipping point"?
– en empirisk analyse af information og kognitioner om fusioner

15. Gregory Gimpel
Value-driven Adoption and Consumption of Technology: Understanding Technology Decision Making

16. Thomas Stengade Sønderskov
Den nye mulighed
Social innovation i en forretningsmæssig kontekst

17. Jeppe Christoffersen
Donor supported strategic alliances in developing countries

18. Vibeke Vad Baunsgaard
Dominant Ideological Modes of Rationality: Cross functional

integration in the process of product innovation

19. Throstur Olaf Sigurjonsson
Governance Failure and Icelands's Financial Collapse

20. Allan Sall Tang Andersen
Essays on the modeling of risks in interest-rate and inflation markets

21. Heidi Tscherning
Mobile Devices in Social Contexts

22. Birgitte Gorm Hansen
Adapting in the Knowledge Economy Lateral Strategies for Scientists and Those Who Study Them

23. Kristina Vaarst Andersen
Optimal Levels of Embeddedness The Contingent Value of Networked Collaboration

24. Justine Grønbæk Pors
Noisy Management A History of Danish School Governing from 1970-2010

25. Stefan Linder
Micro-foundations of Strategic Entrepreneurship Essays on Autonomous Strategic Action

26. Xin Li
Toward an Integrative Framework of National Competitiveness An application to China

27. Rune Thorbjørn Clausen
Værdifuld arkitektur Et eksplorativt studie af bygningers rolle i virksomheders værdiskabelse

28. Monica Viken
Markedsundersøkelser som bevis i varemerke- og markedsføringsrett

29. Christian Wymann
Tattooing The Economic and Artistic Constitution of a Social Phenomenon

30. Sanne Frandsen
Productive Incoherence A Case Study of Branding and Identity Struggles in a Low-Prestige Organization

31. Mads Stenbo Nielsen
Essays on Correlation Modelling

32. Ivan Häuser
Følelse og sprog Etablering af en ekspressiv kategori, eksemplificeret på russisk

33. Sebastian Schwenen
Security of Supply in Electricity Markets

2012

1. Peter Holm Andreasen
The Dynamics of Procurement Management - A Complexity Approach

2. Martin Haulrich
Data-Driven Bitext Dependency Parsing and Alignment

3. Line Kirkegaard
Konsulenten i den anden nat En undersøgelse af det intense arbejdsliv

4. Tonny Stenheim
Decision usefulness of goodwill under IFRS

5. Morten Lind Larsen
Produktivitet, vækst og velfærd Industrirådet og efterkrigstidens Danmark 1945 - 1958

6. Petter Berg
Cartel Damages and Cost Asymmetries

7. Lynn Kahle
Experiential Discourse in Marketing A methodical inquiry into practice and theory

8. Anne Roelsgaard Obling
Management of Emotions in Accelerated Medical Relationships

9. Thomas Frandsen
Managing Modularity of Service Processes Architecture

10. Carina Christine Skovmøller
CSR som noget særligt
Et casestudie om styring og meningsskabelse i relation til CSR ud fra en intern optik

11. Michael Tell
Fradragsbeskæring af selskabers finansieringsudgifter
En skatteretlig analyse af SEL §§ 11, 11B og 11C

12. Morten Holm
Customer Profitability Measurement Models
Their Merits and Sophistication across Contexts

13. Katja Joo Dyppel
Beskatning af derivater
En analyse af dansk skatteret

14. Esben Anton Schultz
Essays in Labor Economics
Evidence from Danish Micro Data

15. Carina Risvig Hansen
"Contracts not covered, or not fully covered, by the Public Sector Directive"

16. Anja Svejgaard Pors
Iværksættelse af kommunikation
- patientfigurer i hospitalets strategiske kommunikation

17. Frans Bévort
Making sense of management with logics
An ethnographic study of accountants who become managers

18. René Kallestrup
The Dynamics of Bank and Sovereign Credit Risk

19. Brett Crawford
Revisiting the Phenomenon of Interests in Organizational Institutionalism
The Case of U.S. Chambers of Commerce

20. Mario Daniele Amore
Essays on Empirical Corporate Finance

21. Arne Stjernholm Madsen
The evolution of innovation strategy
Studied in the context of medical device activities at the pharmaceutical company Novo Nordisk A/S in the period 1980-2008

22. Jacob Holm Hansen
Is Social Integration Necessary for Corporate Branding?
A study of corporate branding strategies at Novo Nordisk

23. Stuart Webber
Corporate Profit Shifting and the Multinational Enterprise

24. Helene Ratner
Promises of Reflexivity
Managing and Researching Inclusive Schools

25. Therese Strand
The Owners and the Power: Insights from Annual General Meetings

26. Robert Gavin Strand
In Praise of Corporate Social Responsibility Bureaucracy

27. Nina Sormunen
Auditor's going-concern reporting
Reporting decision and content of the report

28. John Bang Mathiasen
Learning within a product development working practice:
- an understanding anchored in pragmatism

29. Philip Holst Riis
Understanding Role-Oriented Enterprise Systems: From Vendors to Customers

30. Marie Lisa Dacanay
Social Enterprises and the Poor
Enhancing Social Entrepreneurship and Stakeholder Theory

31. Fumiko Kano Glückstad
Bridging Remote Cultures: Cross-lingual concept mapping based on the information receiver's prior-knowledge

32. Henrik Barslund Fosse
Empirical Essays in International Trade

33. Peter Alexander Albrecht
Foundational hybridity and its reproduction
Security sector reform in Sierra Leone

34. Maja Rosenstock
CSR - hvor svært kan det være? Kulturanalytisk casestudie om udfordringer og dilemmaer med at forankre Coops CSR-strategi

35. Jeanette Rasmussen
Tweens, medier og forbrug
Et studie af 10-12 årige danske børns brug af internettet, opfattelse og forståelse af markedsføring og forbrug

36. Ib Tunby Gulbrandsen
'This page is not intended for a US Audience'
A five-act spectacle on online communication, collaboration & organization.

37. Kasper Aalling Teilmann
Interactive Approaches to Rural Development

38. Mette Mogensen
The Organization(s) of Well-being and Productivity
(Re)assembling work in the Danish Post

39. Søren Friis Møller
From Disinterestedness to Engagement
Towards Relational Leadership In the Cultural Sector

40. Nico Peter Berhausen
Management Control, Innovation and Strategic Objectives – Interactions and Convergence in Product Development Networks

41. Balder Onarheim
Creativity under Constraints
Creativity as Balancing 'Constrainedness'

42. Haoyong Zhou
Essays on Family Firms

43. Elisabeth Naima Mikkelsen
Making sense of organisational conflict
An empirical study of enacted sensemaking in everyday conflict at work

2013

1. Jacob Lyngsie
Entrepreneurship in an Organizational Context

2. Signe Groth-Brodersen
Fra ledelse til selvet
En socialpsykologisk analyse af forholdet imellem selvledelse, ledelse og stress i det moderne arbejdsliv

3. Nis Høyrup Christensen
Shaping Markets: A Neoinstitutional Analysis of the Emerging Organizational Field of Renewable Energy in China

4. Christian Edelvold Berg
As a matter of size
THE IMPORTANCE OF CRITICAL MASS AND THE CONSEQUENCES OF SCARCITY FOR TELEVISION MARKETS

5. Christine D. Isakson
Coworker Influence and Labor Mobility
Essays on Turnover, Entrepreneurship and Location Choice in the Danish Maritime Industry

6. Niels Joseph Jerne Lennon
Accounting Qualities in Practice
Rhizomatic stories of representational faithfulness, decision making and control

7. Shannon O'Donnell
Making Ensemble Possible
How special groups organize for collaborative creativity in conditions of spatial variability and distance

8. Robert W. D. Veitch
Access Decisions in a Partly-Digital World
Comparing Digital Piracy and Legal Modes for Film and Music

9. Marie Mathiesen
Making Strategy Work
An Organizational Ethnography

10. Arisa Shollo
The role of business intelligence in organizational decision-making

11. Mia Kaspersen
The construction of social and environmental reporting

12. Marcus Møller Larsen
The organizational design of offshoring

13. Mette Ohm Rørdam
EU Law on Food Naming
The prohibition against misleading names in an internal market context

14. Hans Peter Rasmussen
GIV EN GED!
Kan giver-idealtyper forklare støtte til velgørenhed og understøtte relationsopbygning?

15. Ruben Schachtenhaufen
Fonetisk reduktion i dansk

16. Peter Koerver Schmidt
Dansk CFC-beskatning
I et internationalt og komparativt perspektiv

17. Morten Froholdt
Strategi i den offentlige sektor
En kortlægning af styringsmæssig kontekst, strategisk tilgang, samt anvendte redskaber og teknologier for udvalgte danske statslige styrelser

18. Annette Camilla Sjørup
Cognitive effort in metaphor translation
An eye-tracking and key-logging study

19. Tamara Stucchi
The Internationalization of Emerging Market Firms:
A Context-Specific Study

20. Thomas Lopdrup-Hjorth
"Let's Go Outside":
The Value of Co-Creation

21. Ana Alačovska
Genre and Autonomy in Cultural Production
The case of travel guidebook production

22. Marius Gudmand-Høyer
Stemningssindssygdommenes historie i det 19. århundrede
Omtydningen af melankolien og manien som bipolære stemningslidelser i dansk sammenhæng under hensyn til dannelsen af det moderne følelseslivs relative autonomi.
En problematiserings- og erfarings-analytisk undersøgelse

23. Lichen Alex Yu
Fabricating an S&OP Process
Circulating References and Matters of Concern

24. Esben Alfort
The Expression of a Need
Understanding search

25. Trine Pallesen
Assembling Markets for Wind Power
An Inquiry into the Making of Market Devices

26. Anders Koed Madsen
Web-Visions
Repurposing digital traces to organize social attention

27. Lærke Højgaard Christiansen
BREWING ORGANIZATIONAL RESPONSES TO INSTITUTIONAL LOGICS

28. Tommy Kjær Lassen
EGENTLIG SELVLEDELSE
En ledelsesfilosofisk afhandling om selvledelsens paradoksale dynamik og eksistentielle engagement

29. Morten Rossing
Local Adaption and Meaning Creation in Performance Appraisal

30. Søren Obed Madsen
Lederen som oversætter
Et oversættelsesteoretisk perspektiv på strategisk arbejde

31. Thomas Høgenhaven
Open Government Communities
Does Design Affect Participation?

32. Kirstine Zinck Pedersen
Failsafe Organizing?
A Pragmatic Stance on Patient Safety

33. Anne Petersen
Hverdagslogikker i psykiatrisk arbejde
En institutionsetnografisk undersøgelse af hverdagen i psykiatriske organisationer

34. Didde Maria Humle
Fortællinger om arbejde

35. Mark Holst-Mikkelsen
Strategieksekvering i praksis
– barrierer og muligheder!

36. Malek Maalouf
Sustaining lean
Strategies for dealing with organizational paradoxes

37. Nicolaj Tofte Brenneche
Systemic Innovation In The Making
The Social Productivity of Cartographic Crisis and Transitions in the Case of SEEIT

38. Morten Gylling
The Structure of Discourse
A Corpus-Based Cross-Linguistic Study

39. Binzhang YANG
Urban Green Spaces for Quality Life
- Case Study: the landscape architecture for people in Copenhagen

40. Michael Friis Pedersen
Finance and Organization:
The Implications for Whole Farm Risk Management

41. Even Fallan
Issues on supply and demand for environmental accounting information

42. Ather Nawaz
Website user experience
A cross-cultural study of the relation between users´ cognitive style, context of use, and information architecture of local websites

43. Karin Beukel
The Determinants for Creating Valuable Inventions

44. Arjan Markus
External Knowledge Sourcing and Firm Innovation
Essays on the Micro-Foundations of Firms' Search for Innovation

2014

1. Solon Moreira
Four Essays on Technology Licensing and Firm Innovation

2. Karin Strzeletz Ivertsen
Partnership Drift in Innovation Processes
A study of the Think City electric car development

3. Kathrine Hoffmann Pii
Responsibility Flows in Patient-centred Prevention

4. Jane Bjørn Vedel
Managing Strategic Research
An empirical analysis of science-industry collaboration in a pharmaceutical company

5. Martin Gylling
Processuel strategi i organisationer
Monografi om dobbeltheden i tænkning af strategi, dels som vidensfelt i organisationsteori, dels som kunstnerisk tilgang til at skabe i erhvervsmæssig innovation

6. Linne Marie Lauesen
Corporate Social Responsibility in the Water Sector: How Material Practices and their Symbolic and Physical Meanings Form a Colonising Logic

7. Maggie Qiuzhu Mei
LEARNING TO INNOVATE: The role of ambidexterity, standard, and decision process

8. Inger Høedt-Rasmussen
Developing Identity for Lawyers Towards Sustainable Lawyering

9. Sebastian Fux
Essays on Return Predictability and Term Structure Modelling

10. Thorbjørn N. M. Lund-Poulsen
Essays on Value Based Management

11. Oana Brindusa Albu
Transparency in Organizing: A Performative Approach

12. Lena Olaison
Entrepreneurship at the limits

13. Hanne Sørum
DRESSED FOR WEB SUCCESS? An Empirical Study of Website Quality in the Public Sector

14. Lasse Folke Henriksen
Knowing networks How experts shape transnational governance

15. Maria Halbinger
Entrepreneurial Individuals Empirical Investigations into Entrepreneurial Activities of Hackers and Makers

16. Robert Spliid
Kapitalfondenes metoder og kompetencer

17. Christiane Stelling
Public-private partnerships & the need, development and management of trusting A processual and embedded exploration

18. Marta Gasparin
Management of design as a translation process

19. Kåre Moberg
Assessing the Impact of Entrepreneurship Education From ABC to PhD

20. Alexander Cole
Distant neighbors Collective learning beyond the cluster

21. Martin Møller Boje Rasmussen
Is Competitiveness a Question of Being Alike? How the United Kingdom, Germany and Denmark Came to Compete through their Knowledge Regimes from 1993 to 2007

22. Anders Ravn Sørensen
Studies in central bank legitimacy, currency and national identity Four cases from Danish monetary history

23. Nina Bellak
Can Language be Managed in International Business? Insights into Language Choice from a Case Study of Danish and Austrian Multinational Corporations (MNCs)

24. Rikke Kristine Nielsen
Global Mindset as Managerial Meta-competence and Organizational Capability: Boundary-crossing Leadership Cooperation in the MNC The Case of 'Group Mindset' in Solar A/S.

25. Rasmus Koss Hartmann
User Innovation inside government Towards a critically performative foundation for inquiry

26. Kristian Gylling Olesen
Flertydig og emergerende ledelse i folkeskolen
Et aktør-netværksteoretisk ledelses-studie af politiske evalueringsreformers betydning for ledelse i den danske folkeskole

27. Troels Riis Larsen
Kampen om Danmarks omdømme 1945-2010
Omdømmearbejde og omdømmepolitik

28. Klaus Majgaard
Jagten på autenticitet i offentlig styring

29. Ming Hua Li
Institutional Transition and Organizational Diversity:
Differentiated internationalization strategies of emerging market state-owned enterprises

30. Sofie Blinkenberg Federspiel
IT, organisation og digitalisering:
Institutionelt arbejde i den kommunale digitaliseringsproces

31. Elvi Weinreich
Hvilke offentlige ledere er der brug for når velfærdstænkningen flytter sig – er Diplomuddannelsens lederprofil svaret?

32. Ellen Mølgaard Korsager
Self-conception and image of context in the growth of the firm – A Penrosian History of Fiberline Composites

33. Else Skjold
The Daily Selection

34. Marie Louise Conradsen
The Cancer Centre That Never Was
The Organisation of Danish Cancer Research 1949-1992

35. Virgilio Failla
Three Essays on the Dynamics of Entrepreneurs in the Labor Market

36. Nicky Nedergaard
Brand-Based Innovation
Relational Perspectives on Brand Logics and Design Innovation Strategies and Implementation

37. Mads Gjedsted Nielsen
Essays in Real Estate Finance

38. Kristin Martina Brandl
Process Perspectives on Service Offshoring

39. Mia Rosa Koss Hartmann
In the gray zone
With police in making space for creativity

40. Karen Ingerslev
Healthcare Innovation under The Microscope
Framing Boundaries of Wicked Problems

41. Tim Neerup Themsen
Risk Management in large Danish public capital investment programmes

2015

1. Jakob Ion Wille
Film som design
Design af levende billeder i film og tv-serier

2. Christiane Mossin
Interzones of Law and Metaphysics
Hierarchies, Logics and Foundations of Social Order seen through the Prism of EU Social Rights

3. Thomas Tøth
TRUSTWORTHINESS: ENABLING GLOBAL COLLABORATION
An Ethnographic Study of Trust, Distance, Control, Culture and Boundary Spanning within Offshore Outsourcing of IT Services

4. Steven Højlund
Evaluation Use in Evaluation Systems – The Case of the European Commission

5. Julia Kirch Kirkegaard
AMBIGUOUS WINDS OF CHANGE – OR FIGHTING AGAINST WINDMILLS IN CHINESE WIND POWER
A CONSTRUCTIVIST INQUIRY INTO CHINA'S PRAGMATICS OF GREEN MARKETISATION MAPPING CONTROVERSIES OVER A POTENTIAL TURN TO QUALITY IN CHINESE WIND POWER

6. Michelle Carol Antero
A Multi-case Analysis of the Development of Enterprise Resource Planning Systems (ERP) Business Practices

Morten Friis-Olivarius
The Associative Nature of Creativity

7. Mathew Abraham
New Cooperativism:
A study of emerging producer organisations in India

8. Stine Hedegaard
Sustainability-Focused Identity: Identity work performed to manage, negotiate and resolve barriers and tensions that arise in the process of constructing or ganizational identity in a sustainability context

9. Cecilie Glerup
Organizing Science in Society – the conduct and justification of resposible research

10. Allan Salling Pedersen
Implementering af ITIL® IT-governance - når best practice konflikter med kulturen Løsning af implementerings-problemer gennem anvendelse af kendte CSF i et aktionsforskningsforløb.

11. Nihat Misir
A Real Options Approach to Determining Power Prices

12. Mamdouh Medhat
MEASURING AND PRICING THE RISK OF CORPORATE FAILURES

13. Rina Hansen
Toward a Digital Strategy for Omnichannel Retailing

14. Eva Pallesen
In the rhythm of welfare creation
A relational processual investigation moving beyond the conceptual horizon of welfare management

15. Gouya Harirchi
In Search of Opportunities: Three Essays on Global Linkages for Innovation

16. Lotte Holck
Embedded Diversity: A critical ethnographic study of the structural tensions of organizing diversity

17. Jose Daniel Balarezo
Learning through Scenario Planning

18. Louise Pram Nielsen
Knowledge dissemination based on terminological ontologies. Using eye tracking to further user interface design.

19. Sofie Dam
PUBLIC-PRIVATE PARTNERSHIPS FOR INNOVATION AND SUSTAINABILITY TRANSFORMATION
An embedded, comparative case study of municipal waste management in England and Denmark

20. Ulrik Hartmyer Christiansen
Follwoing the Content of Reported Risk Across the Organization

21. Guro Refsum Sanden
Language strategies in multinational corporations. A cross-sector study of financial service companies and manufacturing companies.

22. Linn Gevoll
Designing performance management for operational level
- A closer look on the role of design choices in framing coordination and motivation

23. Frederik Larsen
Objects and Social Actions – on Second-hand Valuation Practices

24. Thorhildur Hansdottir Jetzek
The Sustainable Value of Open Government Data
Uncovering the Generative Mechanisms of Open Data through a Mixed Methods Approach

25. Gustav Toppenberg
Innovation-based M&A – Technological-Integration Challenges – The Case of Digital-Technology Companies

26. Mie Plotnikof
Challenges of Collaborative Governance
An Organizational Discourse Study of Public Managers' Struggles with Collaboration across the Daycare Area

27. Christian Garmann Johnsen
Who Are the Post-Bureaucrats? A Philosophical Examination of the Creative Manager, the Authentic Leader and the Entrepreneur

28. Jacob Brogaard-Kay
Constituting Performance Management A field study of a pharmaceutical company

29. Rasmus Ploug Jenle
Engineering Markets for Control: Integrating Wind Power into the Danish Electricity System

30. Morten Lindholst
Complex Business Negotiation: Understanding Preparation and Planning

31. Morten Grynings
TRUST AND TRANSPARENCY FROM AN ALIGNMENT PERSPECTIVE

32. Peter Andreas Norn
Byregimer og styringsevne: Politisk lederskab af store byudviklingsprojekter

33. Milan Miric
Essays on Competition, Innovation and Firm Strategy in Digital Markets

34. Sanne K. Hjordrup
The Value of Talent Management Rethinking practice, problems and possibilities

35. Johanna Sax
Strategic Risk Management – Analyzing Antecedents and Contingencies for Value Creation

36. Pernille Rydén
Strategic Cognition of Social Media

37. Mimmi Sjöklint
The Measurable Me - The Influence of Self-tracking on the User Experience

38. Juan Ignacio Staricco
Towards a Fair Global Economic Regime? A critical assessment of Fair Trade through the examination of the Argentinean wine industry

39. Marie Henriette Madsen
Emerging and temporary connections in Quality work

40. Yangfeng CAO
Toward a Process Framework of Business Model Innovation in the Global Context
Entrepreneurship-Enabled Dynamic Capability of Medium-Sized Multinational Enterprises

41. Carsten Scheibye
Enactment of the Organizational Cost Structure in Value Chain Configuration A Contribution to Strategic Cost Management

2016

1. Signe Sofie Dyrby
Enterprise Social Media at Work

2. Dorte Boesby Dahl
The making of the public parking attendant
Dirt, aesthetics and inclusion in public service work

3. Verena Girschik
Realizing Corporate Responsibility Positioning and Framing in Nascent Institutional Change

4. Anders Ørding Olsen
IN SEARCH OF SOLUTIONS
Inertia, Knowledge Sources and Diversity in Collaborative Problem-solving

5. Pernille Steen Pedersen
Udkast til et nyt copingbegreb
En kvalifikation af ledelsesmuligheder for at forebygge sygefravær ved psykiske problemer.

6. Kerli Kant Hvass
Weaving a Path from Waste to Value: Exploring fashion industry business models and the circular economy

7. Kasper Lindskow
Exploring Digital News Publishing Business Models – a production network approach

8. Mikkel Mouritz Marfelt
The chameleon workforce: Assembling and negotiating the content of a workforce

9. Marianne Bertelsen
Aesthetic encounters
Rethinking autonomy, space & time in today's world of art

10. Louise Hauberg Wilhelmsen
EU PERSPECTIVES ON INTERNATIONAL COMMERCIAL ARBITRATION

11. Abid Hussain
On the Design, Development and Use of the Social Data Analytics Tool (SODATO): Design Propositions, Patterns, and Principles for Big Social Data Analytics

12. Mark Bruun
Essays on Earnings Predictability

13. Tor Bøe-Lillegraven
BUSINESS PARADOXES, BLACK BOXES, AND BIG DATA: BEYOND ORGANIZATIONAL AMBIDEXTERITY

14. Hadis Khonsary-Atighi
ECONOMIC DETERMINANTS OF DOMESTIC INVESTMENT IN AN OIL-BASED ECONOMY: THE CASE OF IRAN (1965-2010)

15. Maj Lervad Grasten
Rule of Law or Rule by Lawyers?
On the Politics of Translation in Global Governance

16. Lene Granzau Juel-Jacobsen
SUPERMARKEDETS MODUS OPERANDI – en hverdagssociologisk undersøgelse af forholdet mellem rum og handlen og understøtte relationsopbygning?

17. Christine Thalsgård Henriques
In search of entrepreneurial learning – Towards a relational perspective on incubating practices?

18. Patrick Bennett
Essays in Education, Crime, and Job Displacement

19. Søren Korsgaard
Payments and Central Bank Policy

20. Marie Kruse Skibsted
Empirical Essays in Economics of Education and Labor

21. Elizabeth Benedict Christensen
The Constantly Contingent Sense of Belonging of the 1.5 Generation Undocumented Youth
An Everyday Perspective

22. Lasse J. Jessen
Essays on Discounting Behavior and Gambling Behavior

23. Kalle Johannes Rose
Når stifterviljen dør…
Et retsøkonomisk bidrag til 200 års juridisk konflikt om ejendomsretten

24. Andreas Søeborg Kirkedal
Danish Stød and Automatic Speech Recognition

25. Ida Lunde Jørgensen
Institutions and Legitimations in Finance for the Arts

26. Olga Rykov Ibsen
An empirical cross-linguistic study of directives: A semiotic approach to the sentence forms chosen by British, Danish and Russian speakers in native and ELF contexts

27. Desi Volker
Understanding Interest Rate Volatility

28. Angeli Elizabeth Weller
Practice at the Boundaries of Business Ethics & Corporate Social Responsibility

29. Ida Danneskiold-Samsøe
Levende læring i kunstneriske organisationer
En undersøgelse af læringsprocesser mellem projekt og organisation på Aarhus Teater

30. Leif Christensen
Quality of information – The role of internal controls and materiality

31. Olga Zarzecka
Tie Content in Professional Networks

32. Henrik Mahncke
De store gaver
- Filantropiens gensidighedsrelationer i teori og praksis

33. Carsten Lund Pedersen
Using the Collective Wisdom of Frontline Employees in Strategic Issue Management

34. Yun Liu
Essays on Market Design

35. Denitsa Hazarbassanova Blagoeva
The Internationalisation of Service Firms

36. Manya Jaura Lind
Capability development in an off-shoring context: How, why and by whom

37. Luis R. Boscán F.
Essays on the Design of Contracts and Markets for Power System Flexibility

38. Andreas Philipp Distel
Capabilities for Strategic Adaptation: Micro-Foundations, Organizational Conditions, and Performance Implications

39. Lavinia Bleoca
The Usefulness of Innovation and Intellectual Capital in Business Performance: The Financial Effects of Knowledge Management vs. Disclosure

40. Henrik Jensen
Economic Organization and Imperfect Managerial Knowledge: A Study of the Role of Managerial Meta-Knowledge in the Management of Distributed Knowledge

41. Stine Mosekjær
The Understanding of English Emotion Words by Chinese and Japanese Speakers of English as a Lingua Franca An Empirical Study

42. Hallur Tor Sigurdarson
The Ministry of Desire - Anxiety and entrepreneurship in a bureaucracy

43. Kätlin Pulk
Making Time While Being in Time
A study of the temporality of organizational processes

44. Valeria Giacomin
Contextualizing the cluster Palm oil in Southeast Asia in global perspective (1880s–1970s)

45. Jeanette Willert
Managers' use of multiple Management Control Systems: The role and interplay of management control systems and company performance

46. Mads Vestergaard Jensen
Financial Frictions: Implications for Early Option Exercise and Realized Volatility

47. Mikael Reimer Jensen
Interbank Markets and Frictions

48. Benjamin Faigen
Essays on Employee Ownership

49. Adela Michea
Enacting Business Models An Ethnographic Study of an Emerging Business Model Innovation within the Frame of a Manufacturing Company.

50. Iben Sandal Stjerne
Transcending organization in temporary systems Aesthetics' organizing work and employment in Creative Industries

51. Simon Krogh
Anticipating Organizational Change

52. Sarah Netter
Exploring the Sharing Economy

53. Lene Tolstrup Christensen
State-owned enterprises as institutional market actors in the marketization of public service provision: A comparative case study of Danish and Swedish passenger rail 1990–2015

54. Kyoung(Kay) Sun Park
Three Essays on Financial Economics

2017

1. Mari Bjerck
Apparel at work. Work uniforms and women in male-dominated manual occupations.

2. Christoph H. Flöthmann
Who Manages Our Supply Chains? Backgrounds, Competencies and Contributions of Human Resources in Supply Chain Management

3. Aleksandra Anna Rzeźnik
Essays in Empirical Asset Pricing

4. Claes Bäckman
Essays on Housing Markets

5. Kirsti Reitan Andersen
Stabilizing Sustainability in the Textile and Fashion Industry

6. Kira Hoffmann
Cost Behavior: An Empirical Analysis of Determinants and Consequences of Asymmetries

7. Tobin Hanspal
Essays in Household Finance

8. Nina Lange
Correlation in Energy Markets

9. Anjum Fayyaz
Donor Interventions and SME Networking in Industrial Clusters in Punjab Province, Pakistan

10. Magnus Paulsen Hansen
Trying the unemployed. Justification and critique, emancipation and coercion towards the 'active society'. A study of contemporary reforms in France and Denmark

11. Sameer Azizi
Corporate Social Responsibility in Afghanistan – a critical case study of the mobile telecommunications industry

12. Malene Myhre
The internationalization of small and medium-sized enterprises: A qualitative study

13. Thomas Presskorn-Thygesen
The Significance of Normativity – Studies in Post-Kantian Philosophy and Social Theory

14. Federico Clementi
Essays on multinational production and international trade

15. Lara Anne Hale
Experimental Standards in Sustainability Transitions: Insights from the Building Sector

16. Richard Pucci
Accounting for Financial Instruments in an Uncertain World Controversies in IFRS in the Aftermath of the 2008 Financial Crisis

17. Sarah Maria Denta
Kommunale offentlige private partnerskaber Regulering I skyggen af Farumsagen

18. Christian Östlund
Design for e-training

19. Amalie Martinus Hauge
Organizing Valuations – a pragmatic inquiry

20. Tim Holst Celik
Tension-filled Governance? Exploring the Emergence, Consolidation and Reconfiguration of Legitimatory and Fiscal State-crafting

21. Christian Bason
Leading Public Design: How managers engage with design to transform public governance

22. Davide Tomio
Essays on Arbitrage and Market Liquidity

23. Simone Stæhr
Financial Analysts' Forecasts Behavioral Aspects and the Impact of Personal Characteristics

24. Mikkel Godt Gregersen
Management Control, Intrinsic Motivation and Creativity – How Can They Coexist

25. Kristjan Johannes Suse Jespersen
Advancing the Payments for Ecosystem Service Discourse Through Institutional Theory

26. Kristian Bondo Hansen
Crowds and Speculation: A study of crowd phenomena in the U.S. financial markets 1890 to 1940

27. Lars Balslev
Actors and practices – An institutional study on management accounting change in Air Greenland

28. Sven Klingler
Essays on Asset Pricing with Financial Frictions

29. Klement Ahrensbach Rasmussen
Business Model Innovation The Role of Organizational Design

30. Giulio Zichella
Entrepreneurial Cognition. Three essays on entrepreneurial behavior and cognition under risk and uncertainty

31. Richard Ledborg Hansen
En forkærlighed til det eksisterende – mellemlederens oplevelse af forandringsmodstand i organisatoriske forandringer

32. Vilhelm Stefan Holsting
Militært chefvirke: Kritik og retfærdiggørelse mellem politik og profession

33. Thomas Jensen
Shipping Information Pipeline: An information infrastructure to improve international containerized shipping

34. Dzmitry Bartalevich
Do economic theories inform policy? Analysis of the influence of the Chicago School on European Union competition policy

35. Kristian Roed Nielsen
Crowdfunding for Sustainability: A study on the potential of reward-based crowdfunding in supporting sustainable entrepreneurship

36. Emil Husted
There is always an alternative: A study of control and commitment in political organization

37. Anders Ludvig Sevelsted
Interpreting Bonds and Boundaries of Obligation. A genealogy of the emergence and development of Protestant voluntary social work in Denmark as shown through the cases of the Copenhagen Home Mission and the Blue Cross (1850 – 1950)

38. Niklas Kohl
Essays on Stock Issuance

39. Maya Christiane Flensborg Jensen
BOUNDARIES OF PROFESSIONALIZATION AT WORK An ethnography-inspired study of care workers' dilemmas at the margin

40. Andreas Kamstrup
Crowdsourcing and the Architectural Competition as Organisational Technologies

41. Louise Lyngfeldt Gorm Hansen
Triggering Earthquakes in Science, Politics and Chinese Hydropower - A Controversy Study

2018

1. Vishv Priya Kohli
Combatting Falsifi cation and Counterfeiting of Medicinal Products in the E uropean Union – A Legal Analysis

2. Helle Haurum
Customer Engagement Behavior in the context of Continuous Service Relationships

3. Nis Grünberg
The Party -state order: Essays on China's political organization and political economic institutions

4. Jesper Christensen
A Behavioral Theory of Human Capital Integration

5. Poula Marie Helth
Learning in practice

6. Rasmus Vendler Toft-Kehler
Entrepreneurship as a career? An investigation of the relationship between entrepreneurial experience and entrepreneurial outcome

7. Szymon Furtak
Sensing the Future: Designing sensor-based predictive information systems for forecasting spare part demand for diesel engines

8. Mette Brehm Johansen
Organizing patient involvement. An ethnographic study

9. Iwona Sulinska
Complexities of Social Capital in Boards of Directors

10. Cecilie Fanøe Petersen
Award of public contracts as a means to conferring State aid: A legal analysis of the interface between public procurement law and State aid law

11. Ahmad Ahmad Barirani
Three Experimental Studies on Entrepreneurship

12. Carsten Allerslev Olsen
Financial Reporting Enforcement: Impact and Consequences

13. Irene Christensen
New product fumbles – Organizing for the Ramp-up process

14. Jacob Taarup-Esbensen
Managing communities – Mining MNEs' community risk management practices

15. Lester Allan Lasrado
Set-Theoretic approach to maturity models

16. Mia B. Münster
Intention vs. Perception of Designed Atmospheres in Fashion Stores

17. Anne Sluhan
Non-Financial Dimensions of Family Firm Ownership: How Socioemotional Wealth and Familiness Influence Internationalization

18. Henrik Yde Andersen
Essays on Debt and Pensions

19. Fabian Heinrich Müller
Valuation Reversed – When Valuators are Valuated. An Analysis of the Perception of and Reaction to Reviewers in Fine-Dining

20. Martin Jarmatz
Organizing for Pricing

21. Niels Joachim Christfort Gormsen
Essays on Empirical Asset Pricing

22. Diego Zunino
Socio-Cognitive Perspectives in Business Venturing

23. Benjamin Asmussen
Networks and Faces between Copenhagen and Canton, 1730-1840

24. Dalia Bagdziunaite
Brains at Brand Touchpoints A Consumer Neuroscience Study of Information Processing of Brand Advertisements and the Store Environment in Compulsive Buying

25. Erol Kazan
Towards a Disruptive Digital Platform Model

26. Andreas Bang Nielsen
Essays on Foreign Exchange and Credit Risk

27. Anne Krebs
Accountable, Operable Knowledge Toward Value Representations of Individual Knowledge in Accounting

28. Matilde Fogh Kirkegaard
A firm- and demand-side perspective on behavioral strategy for value creation: Insights from the hearing aid industry

29. Agnieszka Nowinska
SHIPS AND RELATION-SHIPS Tie formation in the sector of shipping intermediaries in shipping

30. Stine Evald Bentsen
The Comprehension of English Texts by Native Speakers of English and Japanese, Chinese and Russian Speakers of English as a Lingua Franca. An Empirical Study.

31. Stine Louise Daetz
Essays on Financial Frictions in Lending Markets

32. *Christian Skov Jensen*
Essays on Asset Pricing

33. Anders Kryger
Aligning future employee action and corporate strategy in a resource-scarce environment

34. Maitane Elorriaga-Rubio
The behavioral foundations of strategic decision-making: A contextual perspective

35. Roddy Walker
Leadership Development as Organisational Rehabilitation: Shaping Middle-Managers as Double Agents

36. Jinsun Bae
Producing Garments for Global Markets Corporate social responsibility (CSR) in Myanmar's export garment industry 2011–2015

37. Queralt Prat-i-Pubill
Axiological knowledge in a knowledge driven world. Considerations for organizations.

38. Pia Mølgaard
Essays on Corporate Loans and Credit Risk

39. Marzia Aricò
Service Design as a Transformative Force: Introduction and Adoption in an Organizational Context

40. Christian Dyrlund Wåhlin-Jacobsen
Constructing change initiatives in workplace voice activities Studies from a social interaction perspective

41. Peter Kalum Schou
Institutional Logics in Entrepreneurial Ventures: How Competing Logics arise and shape organizational processes and outcomes during scale-up

42. Per Henriksen
Enterprise Risk Management Rationaler og paradokser i en moderne ledelsesteknologi

43. Maximilian Schellmann
The Politics of Organizing Refugee Camps

44. Jacob Halvas Bjerre
Excluding the Jews: The Aryanization of Danish-German Trade and German Anti-Jewish Policy in Denmark 1937-1943

45. Ida Schrøder
Hybridising accounting and caring: A symmetrical study of how costs and needs are connected in Danish child protection work

46. Katrine Kunst
Electronic Word of Behavior: Transforming digital traces of consumer behaviors into communicative content in product design

47. Viktor Avlonitis
Essays on the role of modularity in management: Towards a unified perspective of modular and integral design

48. Anne Sofie Fischer
Negotiating Spaces of Everyday Politics: -An ethnographic study of organizing for social transformation for women in urban poverty, Delhi, India

2019

1. Shihan Du
   ESSAYS IN EMPIRICAL STUDIES BASED ON ADMINISTRATIVE LABOUR MARKET DATA

2. Mart Laatsit
   Policy learning in innovation policy: A comparative analysis of European Union member states

3. Peter J. Wynne
   Proactively Building Capabilities for the Post-Acquisition Integration of Information Systems

4. Kalina S. Staykova
   Generative Mechanisms for Digital Platform Ecosystem Evolution

5. Ieva Linkeviciute
   Essays on the Demand-Side Management in Electricity Markets

6. Jonatan Echebarria Fernández
   Jurisdiction and Arbitration Agreements in Contracts for the Carriage of Goods by Sea – Limitations on Party Autonomy

7. Louise Thorn Bøttkjær
   Votes for sale. Essays on clientelism in new democracies.

8. Ditte Vilstrup Holm
   The Poetics of Participation: the organizing of participation in contemporary art

9. Philip Rosenbaum
   Essays in Labor Markets – Gender, Fertility and Education

10. Mia Olsen
    Mobile Betalinger - Succesfaktorer og Adfærdsmæssige Konsekvenser

11. Adrián Luis Mérida Gutiérrez
    Entrepreneurial Careers: Determinants, Trajectories, and Outcomes

12. Frederik Regli
    Essays on Crude Oil Tanker Markets

13. Cancan Wang
    Becoming Adaptive through Social Media: Transforming Governance and Organizational Form in Collaborative E-government

14. Lena Lindbjerg Sperling
    Economic and Cultural Development: Empirical Studies of Micro-level Data

15. Xia Zhang
    Obligation, face and facework: An empirical study of the communicative act of cancellation of an obligation by Chinese, Danish and British business professionals in both L1 and ELF contexts

16. Stefan Kirkegaard Sløk-Madsen
    Entrepreneurial Judgment and Commercialization

17. Erin Leitheiser
    The Comparative Dynamics of Private Governance
    The case of the Bangladesh Ready-Made Garment Industry

18. Lone Christensen
    STRATEGIIMPLEMENTERING: STYRINGSBESTRÆBELSER, IDENTITET OG AFFEKT

19. Thomas Kjær Poulsen
    Essays on Asset Pricing with Financial Frictions

20. Maria Lundberg
    Trust and self-trust in leadership identity constructions: A qualitative exploration of narrative ecology in the discursive aftermath of heroic discourse

21. Tina Joanes
Sufficiency for sustainability Determinants and strategies for reducing clothing consumption

22. Benjamin Johannes Flesch
Social Set Visualizer (SoSeVi): Design, Development and Evaluation of a Visual Analytics Tool for Computational Set Analysis of Big Social Data

23. Henriette Sophia Groskopff
Tvede Schleimann
Creating innovation through collaboration – Partnering in the maritime sector

24. Kristian Steensen Nielsen
The Role of Self-Regulation in Environmental Behavior Change

25. Lydia L. Jørgensen
Moving Organizational Atmospheres

26. Theodor Lucian Vladasel
Embracing Heterogeneity: Essays in Entrepreneurship and Human Capital

27. Seidi Suurmets
Contextual Effects in Consumer Research: An Investigation of Consumer Information Processing and Behavior via the Applicati on of Eye-tracking Methodology

28. Marie Sundby Palle Nickelsen
Reformer mellem integritet og innovation: Reform af reformens form i den danske centraladministration fra 1920 til 2019

29. Vibeke Kristine Scheller
The temporal organizing of same-day discharge: A tempography of a Cardiac Day Unit

30. Qian Sun
Adopting Artificial Intelligence in Healthcare in the Digital Age: Perceived Challenges, Frame Incongruence, and Social Power

31. Dorthe Thorning Mejlhede
Artful change agency and organizing for innovation – the case of a Nordic fintech cooperative

32. Benjamin Christoffersen
Corporate Default Models: Empirical Evidence and Methodical Contributions

33. Filipe Antonio Bonito Vieira
Essays on Pensions and Fiscal Sustainability

34. Morten Nicklas Bigler Jensen
Earnings Management in Private Firms: An Empirical Analysis of Determinants and Consequences of Earnings Management in Private Firms

2020

1. Christian Hendriksen
Inside the Blue Box: Explaining industry influence in the International Maritime Organization

2. Vasileios Kosmas
Environmental and social issues in global supply chains: Emission reduction in the maritime transport industry and maritime search and rescue operational response to migration

3. Thorben Peter Simonsen
The spatial organization of psychiatric practice: A situated inquiry into 'healing architecture'

4. Signe Bruskin
The infinite storm: An ethnographic study of organizational change in a bank

5. Rasmus Corlin Christensen
Politics and Professionals: Transnational Struggles to Change International Taxation

6. Robert Lorenz Törmer
The Architectural Enablement of a Digital Platform Strategy

7. Anna Kirkebæk Johansson Gosovic
Ethics as Practice: An ethnographic study of business ethics in a multi-national biopharmaceutical company

8. Frank Meier
Making up leaders in leadership development

9. Kai Basner
Servitization at work: On proliferation and containment

10. Anestis Keremis
Anti-corruption in action: How is anti-corruption practiced in multinational companies?

11. Marie Larsen Ryberg
Governing Interdisciolinarity: Stakes and translations of interdisciplinarity in Danish high school education.

12. Jannick Friis Christensen
Queering organisation(s): Norm-critical orientations to organising and researching diversity

13. Thorsteinn Sigurdur Sveinsson
Essays on Macroeconomic Implications of Demographic Change

14. Catherine Casler
Reconstruction in strategy and organization: For a pragmatic stance

15. Luisa Murphy
Revisiting the standard organization of multi-stakeholder initiatives (MSIs): The case of a meta-MSI in Southeast Asia

16. Friedrich Bergmann
Essays on International Trade

17. Nicholas Haagensen
European Legal Networks in Crisis: The Legal Construction of Economic Policy

18. Charlotte Biil
Samskabelse med en sommerfugle-model: Hybrid ret i forbindelse med et partnerskabsprojekt mellem 100 selvejende daginstitutioner, deres paraplyorganisation, tre kommuner og CBS

19. Andreas Dimmelmeier
The Role of Economic Ideas in Sustainable Finance: From Paradigms to Policy

20. Maibrith Kempka Jensen
Ledelse og autoritet i interaktion - En interaktionsbaseret undersøgelse af autoritet i ledelse i praksis

21. Thomas Burø
LAND OF LIGHT: Assembling the Ecology of Culture in Odsherred 2000-2018

22. Prins Marcus Valiant Lantz
Timely Emotion: The Rhetorical Framing of Strategic Decision Making

23. Thorbjørn Vittenhof Fejerskov
Fra værdi til invitationer - offentlig værdiskabelse gennem affekt, potentialitet og begivenhed

24. Lea Acre Foverskov
Demographic Change and Employment: Path dependencies and institutional logics in the European Commission

25. Anirudh Agrawal
A Doctoral Dissertation

26. Julie Marx
Households in the housing market

27. Hadar Gafni
Alternative Digital Methods of Providing Entrepreneurial Finance

28. Mathilde Hjerrild Carlsen
Ledelse af engagementer: En undersøgelse af samarbejde mellem folkeskoler og virksomheder i Danmark

29. Suen Wang
Essays on the Gendered Origins and Implications of Social Policies in the Developing World

30. Stine Hald Larsen
The Story of the Relative: A Systems-Theoretical Analysis of the Role of the Relative in Danish Eldercare Policy from 1930 to 2020

31. Christian Casper Hofma
Immersive technologies and organizational routines: When head-mounted displays meet organizational routines

32. Jonathan Feddersen
The temporal emergence of social relations: An event-based perspective of organising

33. Nageswaran Vaidyanathan
ENRICHING RETAIL CUSTOMER EXPERIENCE USING AUGMENTED REALITY

2021

1. Vanya Rusinova
The Determinants of Firms' Engagement in Corporate Social Responsibility: Evidence from Natural Experiments

2. Lívia Lopes Barakat
Knowledge management mechanisms at MNCs:
The enhancing effect of absorptive capacity and its effects on performance and innovation

3. Søren Bundgaard Brøgger
Essays on Modern Derivatives Markets

4. Martin Friis Nielsen
Consuming Memory: Towards a conceptualization of social media platforms as organizational technologies of consumption

05. Fei Liu
Emergent Technology Use in Consumer Decision Journeys: A Process-as-Propensity Approach

06. Jakob Rømer Barfod
Ledelse i militære højrisikoteams

07. Elham Shafiei Gol
Creative Crowdwork Arrangements

08. Árni Jóhan Petersen
Collective Imaginary as (Residual) Fantasy: A Case Study of the Faroese Oil Bonanza

09. Søren Bering
"Manufacturing, Forward Integration and Governance Strategy"

10. Lars Oehler
Technological Change and the Decomposition of Innovation:
Choices and Consequences for Latecomer Firm Upgrading: The Case of China's Wind Energy Sector

11. Lise Dahl Arvedsen
Leadership in interaction in a virtual context:
A study of the role of leadership processes in a complex context, and how such processes are accomplished in practice

12. Jacob Emil Jeppesen
Essays on Knowledge networks, scientific impact and new knowledge adoption

13. Kasper Ingeman Beck
Essays on Chinese State-Owned Enterprises:
Reform, Corporate Governance and Subnational Diversity

14. Sönnich Dahl Sönnichsen
Exploring the interface between public demand and private supply for implementation of circular economy principles

15. Benjamin Knox
Essays on Financial Markets and Monetary Policy

16. Anita Eskesen
Essays on Utility Regulation: Evaluating Negotiation-Based Approaches inthe Context of Danish Utility Regulation

17. Agnes Guenther
Essays on Firm Strategy and Human Capital

18. Sophie Marie Cappelen
Walking on Eggshells: The balancing act of temporal work in a setting of culinary change

19. Manar Saleh Alnamlah
About Gender Gaps in Entrepreneurial Finance

20. Kirsten Tangaa Nielsen
Essays on the Value of CEOs and Directors

21. Renée Ridgway
Re:search - the Personalised Subject vs. the Anonymous User

TITLER I ATV PH.D.-SERIEN

1992

1. Niels Kornum
Servicesamkørsel – organisation, økonomi og planlægningsmetode

1995

2. Verner Worm
Nordiske virksomheder i Kina
Kulturspecifikke interaktionsrelationer ved nordiske virksomhedsetableringer i Kina

1999

3. Mogens Bjerre
Key Account Management of Complex Strategic Relationships
An Empirical Study of the Fast Moving Consumer Goods Industry

2000

4. Lotte Darsø
Innovation in the Making
Interaction Research with heterogeneous Groups of Knowledge Workers creating new Knowledge and new Leads

2001

5. Peter Hobolt Jensen
Managing Strategic Design Identities
The case of the Lego Developer Network

2002

6. Peter Lohmann
The Deleuzian Other of Organizational Change – Moving Perspectives of the Human

7. Anne Marie Jess Hansen
To lead from a distance: The dynamic interplay between strategy and strategizing – A case study of the strategic management process

2003

8. Lotte Henriksen
Videndeling
– om organisatoriske og ledelsesmæssige udfordringer ved videndeling i praksis

9. Niels Christian Nickelsen
Arrangements of Knowing: Coordinating Procedures Tools and Bodies in Industrial Production – a case study of the collective making of new products

2005

10. Carsten Ørts Hansen
Konstruktion af ledelsesteknologier og effektivitet

TITLER I DBA PH.D.-SERIEN

2007

1. Peter Kastrup-Misir
Endeavoring to Understand Market Orientation – and the concomitant co-mutation of the researched, the re searcher, the research itself and the truth

2009

1. Torkild Leo Thellefsen
Fundamental Signs and Significance effects
A Semeiotic outline of Fundamental Signs, Significance-effects, Knowledge Profiling and their use in Knowledge Organization and Branding

2. Daniel Ronzani
When Bits Learn to Walk Don't Make Them Trip. Technological Innovation and the Role of Regulation by Law in Information Systems Research: the Case of Radio Frequency Identification (RFID)

2010

1. Alexander Carnera
Magten over livet og livet som magt
Studier i den biopolitiske ambivalens