ephemera
theory & politics in organization

ISSN 1473-2866 (Online)
ISSN 2052-1499 (Print)
www.ephemerajournal.org
volume 17(2): 377-397

Against a personalisation of the self

Renée Ridgway

abstract

With presently more than 3 billion search queries a day, Google is the most used search engine in the world. Since December 4, 2009 Google uses 'personalisation' where it captures users' data, logs users' histories and adapts previous search queries into real-time search results, even if one is not signed into a Google account. In exchange for data, users acquire 'tailored' advertising, turning themselves into commodities for advertisers and receiving free services. In order to gain a greater understanding of the complexities involved with data retention of online searching habits, I designed my own 'empirical' study in an attempt to circumvent personalisation and to determine whether one could be anonymous when searching online, and if so, how. In a critical and experimental (auto) ethnography of the self using specific keywords, I investigated if the 'anonymous' browser Tor (The Onion Browser) offered divergent search results from those of 'personalised' Google. The experiment proposes that Tor delivered divergent search results from Google's personalisation in two ways: the ranking of the results and the returned 'unique' URLs. Tor enables a degree of anonymity without exposing the identity of the user (IP address) and delivers 'relevant' search results, thereby offering an alternative to Google's personalisation.

Introduction

> He handed Mae a piece of paper, on which he had written, in crude capitals, a list of assertions under the headline *The rights of humans in a digital age*. Mae scanned it, catching passages: We all must have the right to anonymity. Not every human activity can be measured. The ceaseless pursuit of data to quantify the value of any endeavour 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*. (Eggers, 2013: 491)

Although personal data has been captured by governments and private entities for centuries, nowadays it has become a daily activity for citizens to give away their data to corporations in exchange for 'free services' in online activities, such as search queries. In the past couple of years I have been exploring the notion of personalisation. This process, where corporations deliver customised search results to users, stands at the core of the internet's power structures. In order to explore personalisation, I conducted a series of search experiments using chosen keywords in two different scenarios. On one computer (Apple), I allowed myself to be fully personalised by Google. On the other computer (PC), I searched with the same keywords using the anonymity network, Tor (The Onion Router). In other words, I became either a Google 'personalised subject' or a Tor 'anonymous user'. These experiments have led me to re-think how personalisation, anonymity and collectivity organise and control aspects of our quotidian lives.

Data collecting

On October 6, 2015 Max Schrems, a Viennese masters student of law, 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 whereby Silicon Valley companies were able to receive transfers of personal data from European citizens for data processing, such as data produced by online searches and social media usage. This decision stood at the end of a series of judicial procedures, which had started in 2008, when Schrems had requested to see the data Facebook had collated about him, including the posts he had deleted (eventually he obtained his data – all 1200 pages of information). It continued in 2013, when Schrems lodged the compliant about Facebook concerning EU data privacy restrictions, which eventually led to the aforementioned 2015 decision. Schrems felt that the responsibility should not be completely placed on the consumer (or user), as so many are not able to read the copious *Terms of Service* agreement, nor fathom exactly what it means. He also pointed out that in contrast to the U.S., data privacy in Europe is considered a fundamental right, and the ruling now renders data transfers illegal that only rely on the Safe Harbour self-certification. Edward Snowden's revelations regarding data surveillance by governments and corporations that were not in compliance with EU laws further motivated Schrems.

The 'Schrem suit' might indeed slow down the data transfer of European consumers to corporations located in the U.S. Ways of circumventing the law to enable certain types of data transfer will likely be implemented, but it reflects the consciousness of Europeans to have their data in their own hands in lieu of Silicon Valley. With the additional EU Commission's ruling in 2014 concerning the 'right

to be forgotten', EU citizens now have the option to go through the legal red tape, requesting Google delete information that the user deems embarrassing, even if this information is true (Thylstrup, 2014: 35). Both the 'EU's new directive confirming the right to be forgotten in the face of leaking machines that seem to remember forever', (*ibid.*: 36) and the invalidation of the Safe Harbour agreement function as temporary deterrents to Silicon Valley's international corporate governance of Europe.

However, in a situation where 'international data transfers are the lifeblood of the digital economy' (Levine, 2015), as stated by Thomas Feehan, chief executive of IAB Europe, which represents start-ups and Google, all of this does not even begin to scratch the surface concerning the ability of having access or control over one's data, let alone whether one can be anonymous online, so that one's data cannot be captured and assigned to a particular user: 'As more of our data, and the programs to manipulate and communicate this data, move online, there is a growing tension between the dynamics on the front (where users interact) and on the back (to which the owners have access)' (Stalder, 2012: 242). The user on the backend is a 'data shadow', (Thylstrup, 2014: 30) comprised of the bits and pieces of 'dividual' selves, dispersed as 'masses, samples, data, markets or banks' (Deleuze, 1992: 5-7). Without robust safeguards, multinational companies and governments organise our online experiences around advertising, data tracking and surveillance. 'There stands the nightmare of a "transparency society" in which the exposed life of individuals becomes "big data" in the hands of Internet companies and government intelligence agencies that, while remaining non-transparent themselves, collect and evaluate the traces that have been left behind by digital users' (Beyes and Pias, 2014: 111).

In previous centuries, analogue querying accounted for the collation of citizens' data and could be considered as the 'pre-history of search engines'. Census bureaus relegated government control through constructed statistics, such as King Philip II's 'elaciones topográficas', Louie XIV's administrator Jean-Baptiste Colbert's enquêtes and the harvesting of information by the Hapsburg dynasty (Tantner, 2014: 123). Human informants, such as maids, servants and journeymen, added to this, they had functions not too dissimilar to the present-day 'crawlers' of search engines (*ibid.*) that index information and then pass it on to interested parties. The 'office of address' in the 16th and 17th centuries in many cities of Europe (Paris, London, Amsterdam), collated information (and addresses) from advertisers as well as seekers in 'register books' – and here you can trace similarities to today's IP addresses.

The IP address is the numeric label assigned to any type of device that is connected to a network and that uses IP (Internet Protocol) for communication. The IP

address (as well as GPS) helps determine where you can be locally pinpointed and located. IP addresses are captured when users type in keywords with Google Search and serve as a tracking device. Google fills the word in for us, thereby offering suggestions with their 'autocomplete', which extends to their system of AdWords. Google would indeed 'do no evil', if this would simply speed up search and not direct us to 'favoured' search results. However, with the danger of pointing out the obvious, advertising is still Google's primary revenue model, providing 91% of their revenue.[1] Thus, autocomplete's main intention is to redirect our thoughts and to rather choose popular words that advertisers have paid for in AdWords: 'Google managed to transform this "linguistic capital" into actual money by organizing an algorithmic auction model for selling keywords' (Kaplan, 2014: 57). And it is for this reason that circumventing an IP address identification system is becoming increasingly difficult.

Google's personalisation

Since December 4, 2009 Google uses 'personalisation' where it captures users' data and logs users' histories and adapts previous search queries into real-time search results, even if one is not signed into a Google account. This search engine bias retains user data as algorithms, which gather, extract, filter and monitor our online behaviour, offering suggestions for subsequent search requests. In exchange for our data we receive 'tailored' advertising, making things fit, turning ourselves into commodities for advertisers and receiving free internet usage. Many users are generally aware of these data collating activities yet do not exercise their rights to opt out, or access and delete 'their' data if they can. Ostensibly most users agree to the hidden control of search algorithms and how they affect obtained results, whether for the production of knowledge, information retrieval or just surfing. This personalisation is the currency in the online marketing of our data, correlated through algorithmic technologies as our information (data) is acquired by marketers, or third parties (Ridgway, 2014).

In an attempt to understand how Google is personalising search results, 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'. Published on the *First Monday* blog in February 2011, the research was carried out with great difficulty in the preceding years. Google interfered with the testing while it was being conducted by blocking IP addresses and adding personalisation: '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

1 https://investor.google.com/earnings/2015/Q2_google_earnings.html.

statistical groups' (Feuz et al, 2011). Based on buying habits, search histories and so on, the user is first classified and assigned according to demographics, not as an individual, but rather with what one might call mass personalisation. The authors conclude 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' (Feuz et al., 2011).

This business model has serious side effects. One such side effect is the now notorious *Filter Bubble*, that is the 'distortion effects' of personalised filters:

> 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. In some ways it can act as a magnifying glass, helpfully expanding our view of a niche area of knowledge. (Pariser, 2012: 82-83)

At the same time, these filters limit what we are exposed to and therefore affect our ability to think and learn. In this way, personalisation has legitimised an online public sphere that is manipulated by algorithms. Through the lens of this 'filter bubble' we do not get information that diverges from our own, instead we receive recommendations from our social network and search histories (Pariser, 2012: 82). 'We are led – by algorithms and our own preference for the like-minded – into "filter bubbles", where we find only the news we expect and the political perspectives we already hold dear' (Gillespie, 2014: 88).

Tor's anonymisation

The ability to have control over personal information, deemed 'informational self-determinism', has been at the forefront of many research enquiries, which investigate whether this could even be possible because of the manifold ways in which information is constantly captured in an era of 'big data' (Mayer-Schönberger and Cukier, 2013). One way to escape such forms of commodified statistics and its side effects are tools that provide partial anonymity online. Much like the corporate search algorithms of Google, which are proprietary and their evaluative criteria and code obfuscated from the user, the user in turn, can find ways to obfuscate their online presence, hidden from the very algorithms that are designed by humans to be obscured *and* that obscure. The user could become much more like the algorithms, stealth and arcane, shrouded in (onion) layers of Tor instead of remaining inside the filter bubble of Google.

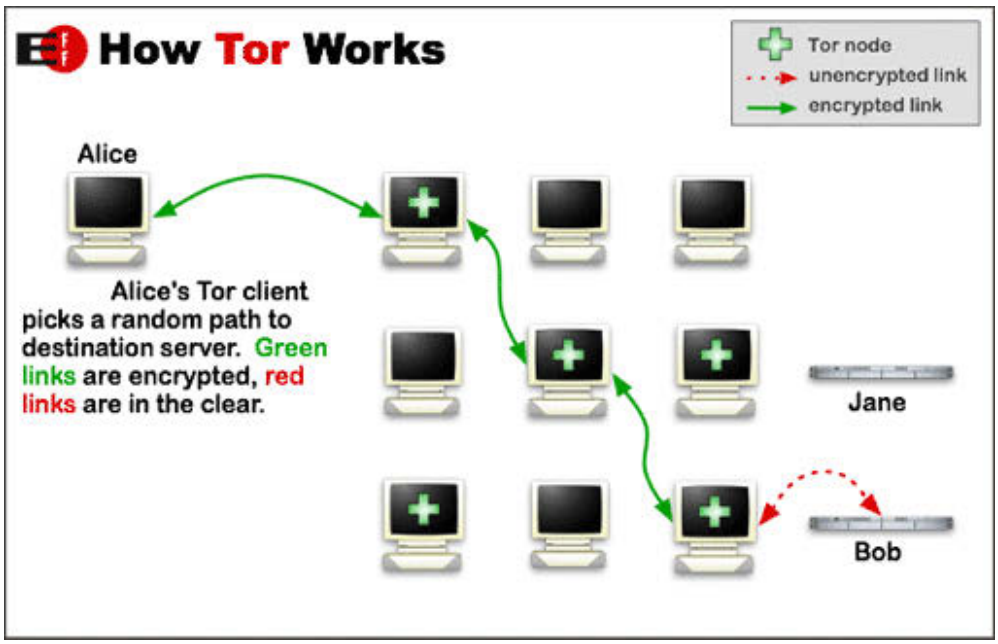


Image 1: Infographic about how Tor works from EFF (Electronic Frontier Foundation).

'Tor is a low-latency anonymity-preserving network that enables its users to protect their privacy online' (AlSabah et al., 2012: 1) and enables anonymous communication. The Tor p2p network is a mesh of proxy servers where the data is bounced through relays, or nodes. Presently more than 7,000 relays[2] enable the transferral of data, applying 'onion routing' as a tactic for anonymity (Spitters et al., 2014: 1). Onion routing was first developed and designed by the US Naval Research Laboratory (NRL) in order to secure online intelligence activities. It is structured by 3 relays (entry, middle, exit) that through a system of circuits transmit the communication, thereby not divulging the IP address of the user.[3]

2 https://torstatus.blutmagie.de.

3 'Tor is a low-latency anonymity network which is based on a client-server architecture model. Clients, known as Onion Proxies (OPs), periodically connect to directory servers to download information about the currently available Onion Routers (ORs), and information on how to contact them such as the OR IP and public keys. Then, clients use ORs to form paths, known as circuits, through the network to Internet destinations. By default, circuits are composed of three ORs, 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).

Data is sent using Tor through a proxy configuration adding a layer of encryption at every node whilst decrypting the data at every 'hop' and forwarding it to the next onion router.[4] '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). In this way the 'clear text' does not appear at the same time and thereby 'hides' the identity of the user and provides anonymity. At the end of a browsing session the user history is deleted along with the HTTP cookie. Although Tor is easy to download and install, the largest critique of Tor by users the past years is its latency, though the last two years it has become much quicker. Moreover, the more people use Tor, the higher the anonymity level becomes for users who are connected to the p2p network, where volunteers around the world provide servers and enable the Tor traffic to flow.

There is controversy surrounding the Tor network. Most of these controversies connect Tor to the so-called 'Dark Net' and 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). In 2014, members of the UK government suggested banning Tor or anonymity systems online and the Chinese government attempted to block and forbid it. Russia even offered a significant monetary award to challenge the anonymity of the Tor network (Çalışkan et al. 2015: 18). Therefore the risk involved in using Tor has become more pronounced. On the other hand, Tor today is an influential anti-censorship technology that allows people in oppressive regimes to access information without the fear of being blocked, tracked or monitored. Tor has often been accredited the past few years in protecting the anonymity of the user in areas of protest and freedom of speech: '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). All this has increased the risks involved in using Tor.

As shown in numerous studies (AlSabah et al., 2012: 1; Biryukov et al., 2013; Çalışkan et al., 2015: 18; Spitters et al., 2014: 1, and Winter et al., 2014: 1), different actors have compromised the Tor network, cracking its anonymity. These actors potentially include the NSA, authoritarian governments worldwide and

4 'A SOCKS proxy interfaces between user applications and the OP. When the application sends data through Tor, the OP divides the data to 512-byte fixed-sized cells, and adds a layer of encryption for every node on the forward path. Then, cells are source-routed through the established circuits. Every hop, on receiving a relay cell, looks up the corresponding circuit, decrypts the relay header and payload with the session key for that circuit, replaces the circuit ID of the header, and forwards the decrypted cell to the next OR' (AlSabah et al., 2012: 75).

multinational corporations – organisations that would like to discover the identity of users and their personal information. Specifically, it should not be disregarded that the Tor exit node operators have access to the traffic going through their exit nodes, whoever they are (Çalışkan et al., 2015: 29). In other words, Tor does not offer 100% anonymity since the exit node is in a position to capture any traffic passing through it, including IP addresses.[5] Other breaches of security include personal computers that might already be infected with malware or spyware and cybercafés that have keyloggers installed on their computers, which make anonymity for users more difficult and even dangerous. Applying a VPN (Virtual Privacy Network) all the way through the three relays could help to boost anonymity, along with using Tails (The Amnesic Incognito Live System). Although Tor's design and programming (along with various patches, etc.) have been added to enhance its security, everything must be perfectly configured. In conclusion, Tor anonymises the origin of the traffic, and ensures encryption inside the Tor network, but it 'does not magically encrypt all traffic throughout the Internet' (Çalışkan et al., 2015: 30).

Besides governmental actors in the security industries, activists, dissidents, journalists and whistleblowers using Tor, there are those who wish to search regions of the internet that have not yet been indexed by Google to form the 'surface web'. This user group includes myself, as I desire to experience the serendipity of finding alternative results. This 'freedom to surf collections without the constraints of disciplinary institutions and freedom to contribute to the construction and curation of one's own past' is my goal in establishing a method to be anonymous online (Thylstrup, 2014: 36). Self-determination, self-governance of one's own data and being free of corporate search strategies are key issues that need to be addressed for such an endeavour. However, 'more research is urgently needed to develop a wider understanding of the social and cultural implications of personalisation of web search in people's everyday life' (Feuz et al., 2011). In order to gain a greater understanding of the complexities involved with data retention of online searching habits, I designed my own empirical experiment in an attempt to circumvent personalisation and to determine whether one could be anonymous when searching online, and if so, how. In a critical and experimental (auto) ethnography of the self I investigate if the 'anonymous' browser Tor offers divergent search results from those of 'personalised' Google.

5 Tails warning: [https://tails.boum.org/doc/about/warning/index.en.html].

The personalised subject versus the anonymised user

The study compares searching keywords on a 'hacker approved' PC that runs Debian using the Tor browser[6] with a completely personalised Apple with an OS Yosemite operating system using Google Search,[7] where Google applies its algorithms to offer relevance and recommendations. Whilst conducting the research online, the search results affect, in turn, the research through the URLs (Uniform Resource Locator) obtained, but also offline behaviour through being personalised. Recursive in spirit regarding the research of search, this study will concomitantly attempt to answer the call for 'a poetics as such for this mysterious new machinic space' (Galloway, 2011: 11).

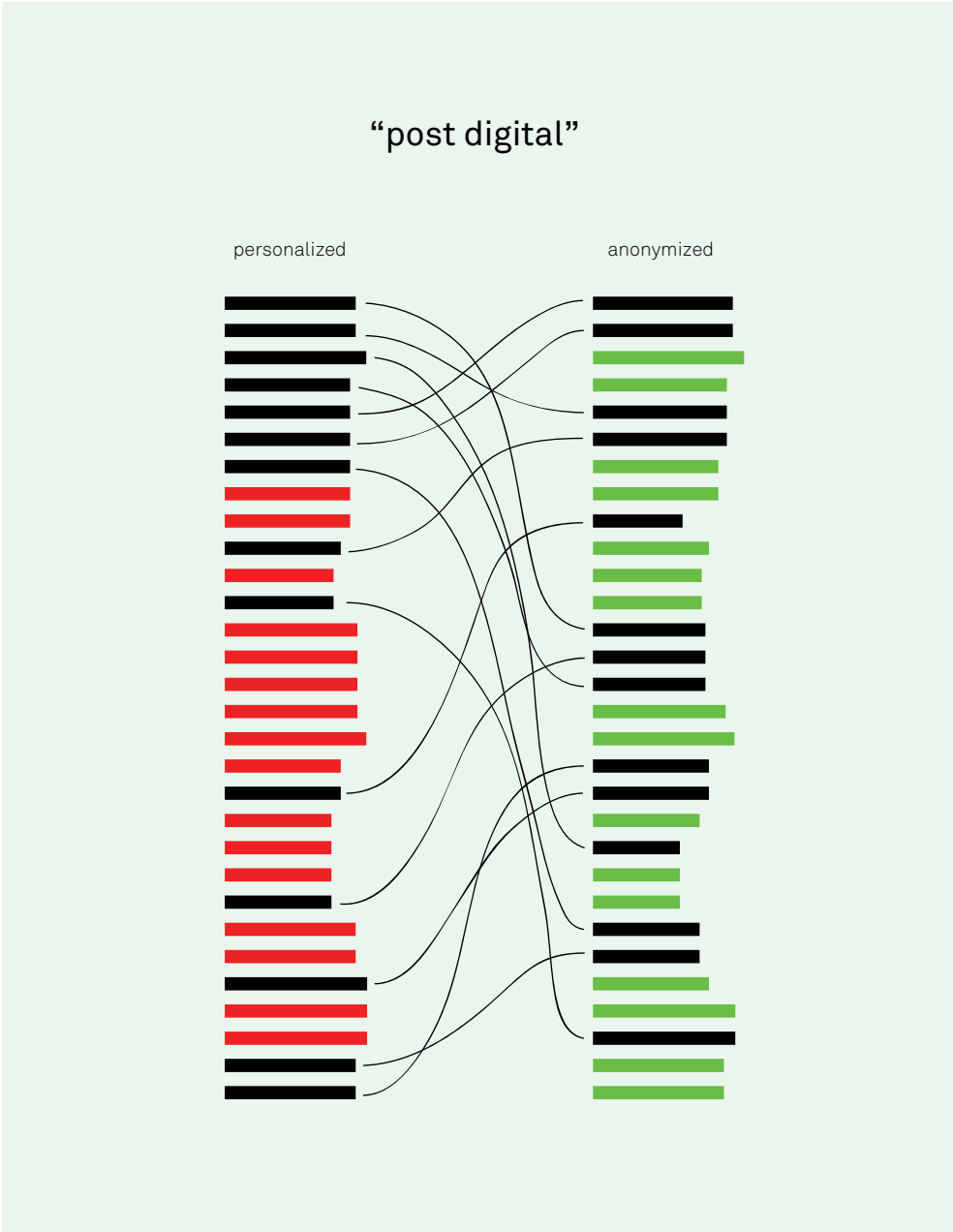
In order to carry out the study in a secure and parallel manner, I received permission from the technical service department at CBS (Copenhagen Business School) to have a router installed inside my office with multiple ports. Each computer was connected to the router by a cable with the router coupled to the CBS internet, allowing almost simultaneous querying, or at least within the same time frame.[8] While my hypothesis was simple – that I would obtain divergent search results on the two computers – I also wanted to find a way to show *how* they differed. In order to do so, I decided to search with the same keywords, same router, same internet connection with cable, same time stamp (same hour), on two different computers and two different browsers. I gathered data on each computer by capturing the entire web page of the 1st page of results, along with the 10th, 20th, 30th, 40th and 50th pages for the data set. I saved these web pages and

6 Obtaining a so-called 'clean' computer was a concerted effort, even convincing trusted hackers to allow me to purchase one of their computers. The PC (a Lenovo Think Pad from about 7 years ago) has only the Tor Browser installed and according to the hackers who set it up for me, is clean. In other words there is no chance of a 'backdoor' when made in China as it has been taken apart, checked and now recycled for this experiment.

7 Given to me to use during my 3-year PhD at Leuphana University, this 13-inch Powerbook is from the end of 2013 with a Retina display, 2,6 GHz Intel Core i5, 16 GB 1600 MHz DDR3 of Memory and an Intel Iris 1536 MB Graphics card. I use Firefox as my browser to search. I have installed no plug-ins (Ghostery, AdBlocker, etc.) for higher privacy and instead, the preferences are set to allow Google 'to have its way with me'.

8 As with most technical setups, this was not without unexpected obstacles en route. Denmark is a trust society, which means everyone in my department has the same master's key to all offices and can let outsiders access offices during business hours. One day, a telephone installer changed all the plugs around, knocking my router off the internet so the technical service had to reconfigure the router, which meant bureaucracy and loss of time. This happened twice up until now. With security updates, the technical service at CBS has reconfigured my router, three times so far, unbeknownst to me, until I discover I have no connection.

clicked through the page numbers at the bottom, and continued to the next page of results. I then engaged the services of a graphic designer, and the data visualisations included here are speculative results. 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.



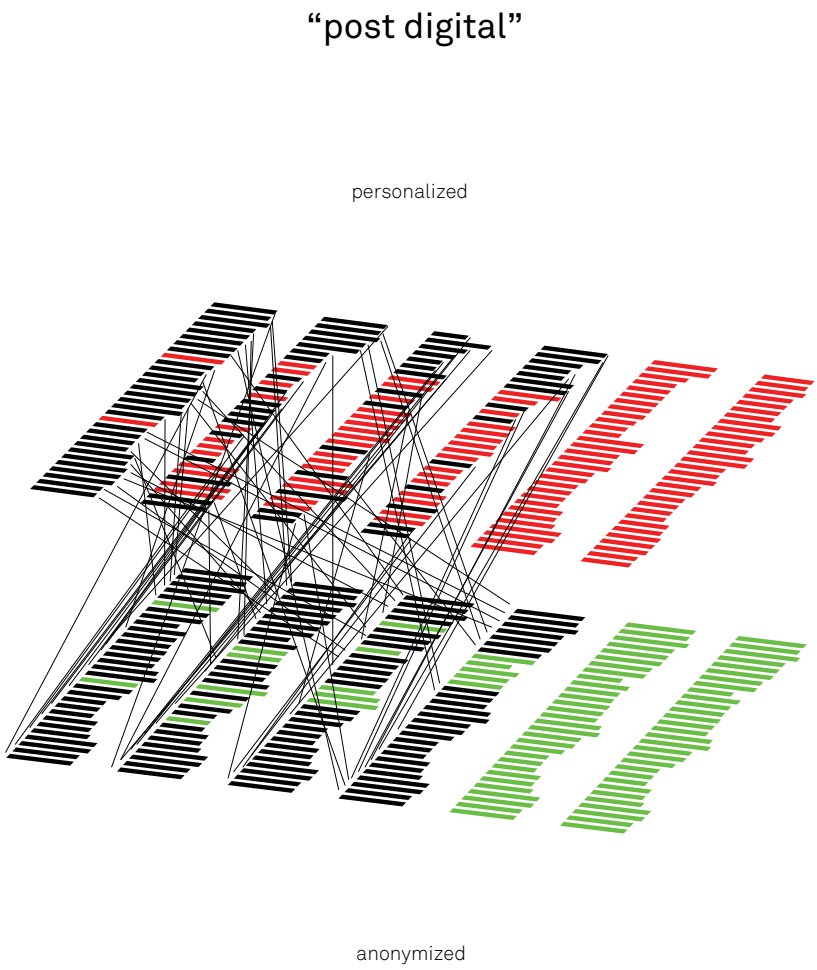


Image 2 and 3: Comparison of imagined ‘personalized’ and ‘anonymized’ search results with keyword ‘postdigital’. Concept: Renée Ridgway. Data visualisation: Richard Vijgen.

Actually the keyword ‘post-digital’ in my findings show that Google does not go beyond the 87th page and Tor not past the 60th page. Whilst searching on the personalised Apple after so many results (35th page), Google asked me if I wanted to search again without having the redundancies eliminated, so I did, with the result of receiving more results, which I used instead. The images emphasise the

links that are unique for a specific configuration – those which only appear with 'personalisation' or those which only appear when 'anonymised'. This 'small data' test shows that the more results obtained, the larger the amount of difference between the two browsers. Numerous speculations exist why there are divergent results; the most obvious is that locative data is included in personalisation, which affects the results based on country and language.

As a next step, I decided to build a larger data set, with more 'keywords' that reflect the vocabularies I came across in my research that were not just 'trending' on Google or had a high currency for AdWords. I used specific keywords, terminology in contemporary art, new media and digital aesthetics, in order to find potential undiscovered texts or projects about these very notions. *Re:search – Terms of art* reflects the 'epistemological gain' measured by their URLs (Uniform Resource Locator) and consists of the following keywords: Accelerationism, Aesthetic Turn, Anthropocene, Artistic Research, Contemporaneity, Creative Industries, Cultural Entrepreneurship, New Aesthetic, Object Oriented Ontology, Performativity, Post Digital, Post Humanism, Post Internet, Post Media, Transmedia.[9] Over the course of a couple of months (October-November in 2015) I searched more and more keywords, in other words carrying out more 'qualitative interviews' with algorithms that gave me 'answers' or results, in an environment that was in constant flux. In regard to labour, I manually conducted the search queries, conducting Mechanical Turking of sorts. The labour was repetitive, incredibly time-consuming and required full-concentration in order to save every web page and gather the data. At a certain moment Google sent me a message to the effect of 'looks like you are a machine' and I had to start over again (earlier, I had also received CAPTCHAs from Tor to test whether I was a machine.).

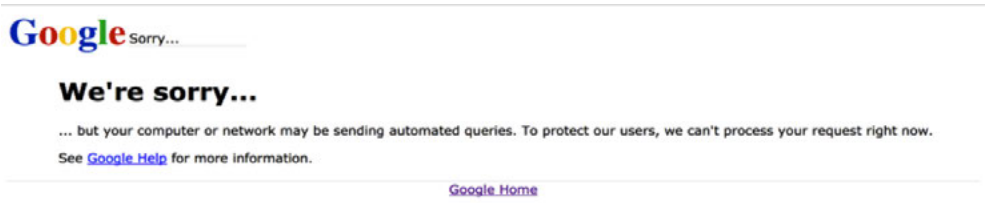


Image 4: Screenshot of 'Googlesorry'

9 I should mention that the term prefix 'post' is problematic. Not only is it now the 'term of the year' (2016) with 'post-truth' society, but with it's double meaning. On the one hand it means 'mail delivery services' and I received many URLs that referred to the various Danish postal systems. On the other hand, many art terms use the term 'post' to make a distinction between various eras or movements. Even a 'post post' whatever is commonplace nowadays.

My data collection so far only involved saving webpages and making screenshots. Initially, I was manually extracting the URLs I obtained. In order to save time, I then started to apply a Python script (provided by an 'anonymous' colleague at Leuphana University and written specifically for this experiment) in order to extract the URLs, which I then extrapolated and placed in an Excel file. My graphic designer used these Excel files to visualise my results – helping me to see another 'view' to understand the results and to compare these two types of online querying. This method, which I am calling 'data visualisation as transcription', allowed me to interpret and analyse my results more efficiently and in a completely new way.[10]

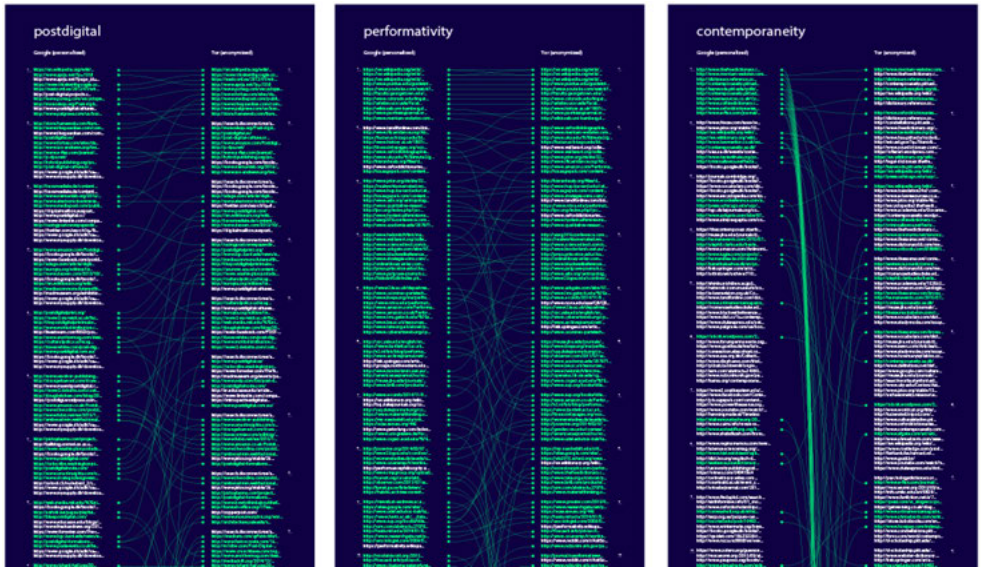


Image 5: Comparison of Goodle Search 'personalized' and Tor Browser 'anonymized' search results with keywords 'postdigital', 'performativity' and 'contemporaneity' Green represents 'identical' URLs. White represents 'unique' URLs

There were constant incidents en route that 'messed' with my searching methods. Google started returning less SERPS (Search Engine Result Page), which means I received less search results. Eventually Tor started to do the same, though Google always delivered more SERPS than Tor, at least with my chosen keywords. As someone who went through all of the webpages of the given results, I was able to see exactly how many pages (and therefore number of results) were actually returned. Moreover, once more I would receive a message from Google around the 35-39th page stating that they had eliminated redundancies and asking if I wished to search again. I kept this data set to the first returns, with the consciousness that I was personalising myself if I were to repeat the keyword during the data capture,

10 'While the URLs are shortened in the print version for legibility reasons, it is the full URL that is being tested for matches'. Richard Vijgen, graphic designer.

stacking the deck as it were.[11] Tor followed suit and seemed to be mimicking Google results with the amount of pages they delivered, however it was consistently less than Google.

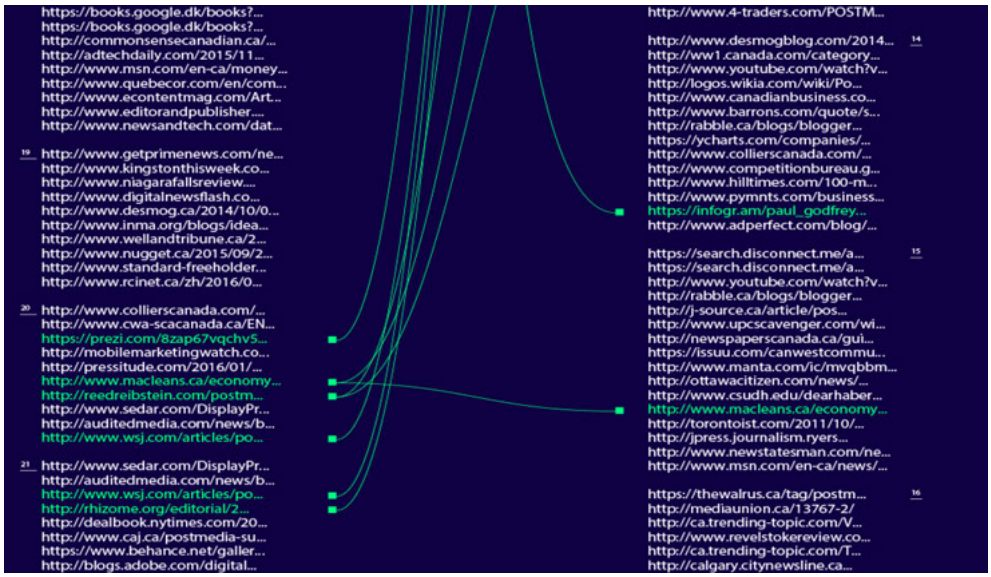


Image 6: Detail: Comparison of Google Search 'personalized' and Tor Browser 'anonymized' search results with keyword 'postmedia'. Green represents 'identical' URLs. White represents 'unique' URLs.

The results were never exactly the same. The major difference between the search results is that even though the URL is the same, the ranking of Google is not the same as Tor. If we compare various keywords, the same URLs are represented by green yet there are also unique results as shown by the white URLs: both Tor and Google delivered unique results. It wasn't until I started digging deeper (searching for answers with Google Search) that I discovered Tor had changed its default browser. 'Startpage.com' used to be Tor's default search engine, yet since March 2015 Tor incorporated 'Disconnect Search' in its browser bundle as its default search engine. Tor stated on their website that 'Disconnect provides private Google search results to Tor users without CAPTCHAs or bans'.[12] I also started to obtain 'ads' from Disconnect Search in my results, which skewed the data as they changed the order of how many results were delivered per page (even though I had set them both at 10 returns per page). Disconnect declared that it does 'detect non-personally identifiable geo-location information to optimize our services, but

11 I did not click on the URL during any time when I was capturing data, as this would have affected the results for personalisation and even added to it. Many of the words were 'first time' search terms.

12 https://blog.torproject.org/blog/tor-browser-45-released.

[unlike Google] we definitely don't collect your precise geo-location or associate geo-location information with a particular user' (*ibid*.). After my experiment, the situation changed once more, as Disconnect was delivering results from Bing and no longer delivering Google search results and as of June 2016, DuckDuckGo is the default search engine for Tor.[13]

Looking back on my 'small data search experiment,' which executed many search requests in a given time frame, I carried out 'qualitative interviews' with algorithms that gave me 'answers' or results. The process itself necessitated negotiating the technical anxieties of attempting to carry out 'empirical' research in an environment where the infrastructure is invisible and the algorithms are in constant flux. As much as Tor is changing its browser bundle and choice of the default search engine, Google is constantly changing its search algorithm. With Tor my IP address was hidden – there is no pinpointing locative data – and for these moments I felt I was able to gather data online 'anonymously'. The ability to be anonymous online, on the 'clear net' without Google's personalisation gives one a sense of freedom and control over one's autonomy. I witnessed a completely different user experience using Tor (and a PC) than searching with Google Search, as Google has a seamless interface that makes searching effortless, and suggests past searches with autocomplete.

The key result – that Tor offered 'relevant' (ostensibly Google) search results – without exposing the identity of the user because of hidden IP address, offers an alternative to Google's personalisation. The experiment also confirms that Tor delivered divergent search results from Google's personalisation in two ways: first, the ranking of the results and the fact that 'unique' URLs were returned. Moreover,

13 https://disconnect.me/privacy. DuckDuckGo also does not share data [https://duckduckgo.com/]. During the past two years Tor has become much faster, which has to do with more relay and exit node operators, the increasing amount of bandwidth available to users, and the fact that file sharing over Tor is less common now that many other services exist for transfer of larger files or storage in the cloud. I can only infer that Tor would like to have quality search results 'as good as Google' yet it does not allow IP addresses to be collated nor sell user data to third parties. Tor states on their blog: 'For a while now Disconnect has no access to Google search results anymore, which we used in Tor Browser. Disconnect being more a meta search engine which allows users to choose between different search providers fell back to delivering Bing search results which were basically unacceptable quality-wise. While Disconnect is still trying to fix the situation we asked them to change the fall back to DuckDuckGo as their search results are strictly better than the ones Bing delivers.' [https://blog.torproject.org/blog/tor-browser-60-released] It is still possible for Tor users to specify they wish to search via Bing (or Yahoo) via Disconnect. But Google is not currently an option, although many users would like Disconnect Search to restore access to Google Search.' [https://techcrunch.com/2016/05/31/tor-switches-to-duckduckgo-search-results-by-default/].

it proposes that Google delivers customised (personalised) search but cannot show the criteria of how Google Search ranks the results, nor how it 'personalises' users and to what degree. I postulate that Google assigns users to 'people like them' as shown in the previous experiment I reference in this text (Feuz et al., 2011). On the other hand, it now seems to me as if I am, when searching via Tor, collaboratively filtered (assigned) on my computer as a 'Tor user' by every web server who sees my IP address.[14]

Preliminary conclusions

If we assume for now 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, or people like me (an assumption that 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 simple personalised versus anonymised search, I would have had, in fact, conducted, on the one hand, search that is collective-via-users-like-me, versus, on the other hand, search that is collective-via-all-Tor-users. At stake are two collectives. These two collectives 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 assign me into this or that collective. I have no access, no knowledge and no agency in regards to the collectives, which I am made part of via Google. Both the forces that sort me into a collective and the collectives that I am sorted into, i.e. the clusters or groups that Google sets up, are not transparent to me. Meanwhile, Google still collects my individual search activities, and in future scenarios Google will probably individualise search even further based on this data collated in the past and present. Tor's collective, on the other hand, is at least partially known to me. Of course I do not know who is in it (after all it is an anonymised network) all the time but I can look at the 'exit address' list, which is constantly updated that shows who is using it and their IP address).[15] The key difference perhaps is that we decide to be in the 'anonymous Tor' collective,

14 'Any web server will get which public IP address a user is coming from. It's kind of like watching people on public square: it is easy to see what's the street they came from with extra difference that internet users, if they would be on public square, would also wear the big sign saying what's their public IP address…one could easily imagine that the same way a user gets annotated by their country of origin (via geo-IP-database) they also get annotated as "Tor country" (matching IP address of Tor exit nodes). That's something any web server can do (unlike internet service providers who would need to (arguably) put in a lot of effort in order to "capture" Tor traffic from that (beginning) to end)'. Email correspondence with anonymous hacker (12.02.2017).

15 [https://check.torproject.org/exit-addresses].

whereas Google assigns us to particular groups through their non-transparent process of collaborative filtering.

Both search collectives, e.g. the one determined by Google algorithms as well the one created by the decision to use Tor, add to specific filter bubbles. But once more, the filter bubbles are structurally different: in the case of the bubble produced by Google's algorithms, Google uses the data of its users, tweaks its algorithms and feeds this back in the loop. When I search different things, I would just be merged into different clusters with *other* people like me. I would then add to the feedback loop by continuously adding to my own personalisation by clicking on the links that are fed to me. I do not have access to the Google cluster itself – I would be switched into a different cluster by an algorithmically organised process that I have no control over.[16] 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).

What changes is what Tor uses as their default search engine (Startpage, Disconnect Search or presently DuckDuckGo) and if this default uses Google. When I use Tor I am part of an anonymity p2p 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 in the constant flux of algorithmic clustering of personalisation. To partake anonymously in a p2p-collective individuates me more than personalisation does. At stake is an *individuation* in the sense of Bernard Stiegler's reading of Simondon – an individuation that is marked by being collective and psychic alike, which forms the opposite to the *individualisation* of the pseudo-autonomous objects of Google's personalisation.

When Introna and Nissenbaum wrote their seminal text, *Shaping the web: Why the politics of search engines matters* in 2000, the World Wide Web was a growing space of websites that were not necessarily interconnected. For some it 'was a new medium, a democratizing force that will give voice to diverse social, economic, and cultural groups, [and] to members of society not frequently heard in the public sphere. It will empower the traditionally disempowered, giving them access both to typically unreachable nodes of power and to previously inaccessible troves of information' (Introna and Nissenbaum, 2000: 177). There was also the belief that the web and searching wasn't only about information retrieval but knowledge exploration. Written at the dawn of the development of 'gateway platforms' for the

16 I have tried to obtain answers from various Google employees but they all sign 'non-disclosure' agreements when hired.

internet, one of their concerns regards access, for 'those with something to say and offer, as well as those wishing to hear and find' (Introna and Nissenbaum, 2000: 169). The concept of serendipity, or the discovery of websites and connecting linkages one didn't know occurred through surfing the net, by just clicking on hyperlinks and not knowing where these would lead, was the modus operandi for users. This utopian vision of the World Wide Web was that we would search and find information, which we could then share, and hence the world would become comprehensible in all its diversity. This was, of course, an illusion.

What we can now see is that one of the reasons for the end of this illusion lies in the way this 'democratic information space' was conceptualised. The web was thought of as emerging out of myriads of individualised website creators and surfers, whose joint activities would then add up miraculously to a new structured, yet democratic and open space. What was absent was a systematic approach to the need of organising *collectives* to systematise the processes that enables us to navigate this space. This absence, and with it the inherent ideology of individualism at the heart of the World Wide Web, was the opportunity search engines seized. They answered the question of how to navigate and organise such a space by creating algorithmically determined collectives, calling these processes of clustering, in a rather interesting twist: 'personalisation'. In doing so, they became a force that not only enabled accessibility, but also commodified and monopolised access to information, stifled psychic and collective co-individuation and pushed instead the individualisation of the web even further. Whereas early net programmers and users with their 'bulletin board' postings, chat rooms or networks in the 1990s envisioned a 'digital democracy', instead a new form of censorship within political discourse emerged, creating what Matthew Hindman (2009) describes as 'Googlearchy'.[17] The tragedy of the web is that 'deliberative democracy' has been prohibited by a flaw in the World Wide Web's very own structure, recently elucidated by the UK referendum and the US election in 2016.

The experience of setting up these experiments has opened up a *view* on what search could look like, offering 'relevant' search results with Tor, without the user being a 'personalised subject'. After I conducted my research, the Tor browser switched its bundle to DuckDuckGo as its default search engine option and it is not clear whether it uses Google search results by default. However it offers privacy browsing and doesn't track users because data (user IP) is not collected nor do they collect precise geo-location and assign it to a particular user. Tor then provides online protection in the form of anonymity, even though there are still risks to using Tor as it has been and could potentially be compromised by malicious actors in the future. Aside from its other merits in terms of challenging surveillance by

17 Those most heavily linked 'rule', in other words.

state actors, using Tor is also one, albeit not the only, strategy to challenge the internet's very own malevolent power structures. As one of the few alternatives to personalised search it offers to the 'anonymous user' a chance to actually *explore* the internet in an on-going and almost impossible experiment in anonymity.

references

AlSabah, M., K. Bauer and I. Goldberg (2012) 'Enhancing Tor's performance using real-time traffic classification', paper presented at CCS '12, Raleigh, North Carolina, USA, October 16-18.

Beyes, T. and C. Pias (2014) 'Transparenz und Geheimnis', *Zeitschrift für Kulturwissenschaft* 2/2014: 111; quoting from B. Han (2012) *Transparenzgesellschaft*. Berlin: Matthes & Seitz.

Biryukov, A., I. Pustogarov and R.-P. Weinmann (2013) 'Trawling for Tor hidden services: Detection, measurement, deanonymization', paper presented at 2013 IEEE Symposium on Security and Privacy, San Francisco, California, USA, May 19-22.

Çalışkan, E., T. Minárik and A.-M. Osula (2015) 'Technical and legal overview of the Tor anonymity network', Tallin: CCDCOE, NATO Cooperative Cyber Defence Centre of Excellence.

Deleuze, G. (1992) 'Postscript on the societies of control', *October,* 59: 3-7.

Eggers, D. (2013) *The circle*. San Francisco: McSweeney's.

Feuz, M., M. Fuller and F. Stalder (2011) 'Personal web searching in the age of semantic capitalism: Diagnosing the mechanics of personalisation', *First Monday,* 16(2-7), February 2011.
[http://firstmonday.org/article/view/3344/2766]

Galloway, A. (2011) 'Are some things unrepresentable?', *Theory, Culture & Society,* 28(7-8): 85-102.

Gillespie, T. (2014) 'The relevance of algorithms', in T. Gillespie, P. Boczkowski and K. Foot (eds.) *Media technologies: Essays on communication, materiality, and society*. Cambridge, MA: MIT Press.

Hindman, M. (2009) *The myth of digital democracy*. Princeton: Princeton University Press.

Kaplan, F. (2014) 'Linguistic capitalism and algorithmic mediation', *Representations*, 127(1): 57-63.

Introna, L.D. and H. Nissenbaum (2000) 'Shaping the web: Why the politics of search engines matters', *The Information Society,* 16(3): 169-185.

Levine, R. (2015) 'Behind the European privacy ruling that's confounding Silicon Valley', *New York Times,* October 9. [http://www.nytimes.com/2015/10/11/business/international/behind-the-european-privacy-ruling-thats-confounding-silicon-valley.html?ref=technology&_r=2]

Mayer-Schönberger, V. and K. Cukier (2013) *Big data: A revolution that will transform how we live, work and think*. London: John Murray.

Pariser, E. (2012) *The filter bubble*. New York: Penguin Books.

Ridgway, R. (2014) 'Personalisation as currency', *A Peer-Reviewed Journal About (APRJA)*. [http://www.aprja.net/?p=2531]

Spitters, M., S. Verbruggen and M. van Staalduinen (2014) 'Towards a comprehensive insight into the thematic organization of the Tor hidden services', paper presented at 2014 IEEE Joint Intelligence and Security Informatics Conference, Los Angeles, CA, USA, Dec 15 -17.

Stalder, F. (2012) 'Between democracy and spectacle: The front-end and the back-end of the social web', in M. Mandiberg (ed.) *The social media reader*. New York: New York University Press.

Tantner, A. (2014) 'Before Google: A pre-history of search engines in analogue times' in R. König and M Rasch (eds.) *Society of the query reader: Reflections on web search*. Amsterdam: Institute of Networked Cultures.

The Tor Project (2017). [https://www.torproject.org/docs/faq.html]

Thylstrup, N. (2014) 'Archival shadows in the digital age', *Nordisk Tidsskrift for Informationsvidenskab og Kulturformidling*, 3(2/3): 29-39.

Winter, P., R. Köwer, M. Mulazzani, M. Huber, S. Schrittwieser, S. Lindskog and E. Weippl (2014) 'Spoiled onions: Exposing malicious Tor exit relays', paper presented at Privacy Enhancing Technologies Symposium, Amsterdam, Netherlands, July 16-18.

Appendix: Interview with an Anonymous hacker (AH)

Renée Ridgway (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 that you have purchased it 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 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 is an operating system that is installed on a USB stick that you boot with your computer. Using Tails in combination with Tor 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 deletion and anonymity.

the author

Renée Ridgway is a PhD fellow at Copenhagen Business School (Management, Politics and Philosophy department) and a research affiliate with the Digital Cultures Research Lab (DCRL), Leuphana University, Lüneburg. Her current research merges artistic and curatorial practice with digital economies in regard to online remuneration along with investigating the conceptual as well as technological implications of 'search'. Recent contributions to publications include Hacking Habitat, Money Labs (INC), OPEN!, APRJA and Disrupting Business.
Email: rr@reneeridgway.net