The *Search* for Contextual Integrity

— A Google Story —

Kartikeya Saigal

Department of Media and Communication

Erasmus University Rotterdam

5th November 2021

kartikeyasaigal01@gmail.com

The Search for Contextual Integrity: A Google Story

The Google Search Engine is link by link, click by click, building possibly the most lasting, ponderous and significant cultural artefact in the history of humankind (Tene, p.1435). During the early days of the company and in the heyday of Silicon Valley when its founders might have been mistaken for young idealists keen to do good, the company was grappling with a lack of revenue — Google is a *free* service. But so powerful are its stores of information that once redirected from their original user centric yet financially underperforming model of simply providing the best *search*, to an almost single minded focus on financial gain through advertising, from 2001 - 2004, following the discovery and exploitation of behavioural surplus as a result of economic and institutional pressure Google conjured a revenue stream from what seemed to be thin air and produced a stunning 3,590 percent increase in revenue (Zuboff, p.165). Initially justified by its founders as a desperate measure to turn a profit, this state of exception, of exploiting users privacy and undermining their decision rights, was institutionalised, and today the company is valued at over a trillion dollars.

This article in its exploration and examination of the Google search engine as a cultural and digital artefact will do so by mentioning the positives of this technology but focusing on the concerns it raises with regards to contextual integrity of individual privacy and decision rights. The concept of contextual integrity will help us understand the Google search engine as a mutable product of human innovation and not the timeless, incontestable one it wants to be. It will do this by illustrating the differences between human interaction in the real world as compared to ones now mainstream in the digital realm, in shades of nuance. In doing so it appeals to the instinctual need for privacy and other social values correlated with civilisational progress.

The Need for a Search Engine

A lot is already in the public discourse about the benefits of Google's technology and this is in no short measure because of the actual genius and convenience provided by their invention but also because of their campaigning. For all of Google's technological prowess and computational talent, the real credit for its success goes to the radical social relations that the company declared as facts, beginning with its disregard for the boundaries of private human experience and the moral integrity of the autonomous individual (Zuboff, p.41). Therefore we shall sum up the pros quickly. Imagine the internet without search. It would exist as unexplored

treasure, the largest collection of human knowledge in history without the means by which to consistently access relevant information. One might hunt forever through the bookshelves of this infinite library. And so the need for a search engine. A digital program that would crawl and index each page of this ever- expanding collection and make it organised and accessible. This algorithm would function like a meticulous librarian, knowledgeable and capable of retrieving exactly the words you handed in. One senses the unlocking of human potential. Already this technology is magnificent. But Google does it better.

Google's librarian can not only retrieve information based on your search words but the longer you talk, the more you reveal, the more your browsing is personalised. No longer guessing at your requests, the librarian can offer you grammatical corrections, tell you what you might be interested in, try to pinpoint each query to your exact desire. This is no longer an ordinary digital librarian but one that has the means to become omniscient.

Inside Google Search

When a user performs a search, and especially if they have logged into their account, Google tracks all entered data, analyses its relevance in the users historical context as well as broader contexts such as language, age, sex etc. and delivers search results that progressively grow more accurate and relevant with use. In addition to key words, each Google search query also produces a wake of collateral data such as the number and pattern of search terms, how a query is phrased, spelling, punctuation, dwell times, click patterns, and location (Zuboff, p.129). It took me 0.60 seconds to get results for 'metadata' with suggestions for what other people ask in relation to it, top sites visited by previous searchers, a dutch translation because my IP address pinpoints me in the Netherlands, YouTube videos on the subject and so on. More interestingly, if I search for 'Apple price' it gives me the stock summary for Apple inc. It could just as well have offered the prices of apples at the nearest grocery store but it knows my browsing history and assumes this is what I'm looking for. Similarly, Tene writes 'the words "Paris Hilton video" might be entered by a user searching for accommodation in the French capital, or (perhaps more likely) by one eager to follow the celebrity heiress's latest antics (Tene, p.1450). To be the best search engine Google needs to know what you mean. Indeed, this element of predictive, intuitive functioning is crucial to the company and the success of its product. It was also the first design goal in Larry Page and Sergey Brin's PhD document 'The Anatomy of a Large-Scale Hypertextual Web Search Engine' which outlined the functioning and use of the Google search engine. "Our main goal is to improve the quality of web search

engines" (Brin & Page, p.2). And while this ended up being the most user oriented and valuable aspect of Google search it was actually founded on a mathematical problem. In their essay, the co-founders determine that the number of documents in their indices has been increasing by many orders of magnitude, but a user's ability to look at documents has not. People are still only willing to look at the first few tens of results. Because of this, as the collection size grows, they needed tools that have very high precision (number of relevant documents returned, say in the top tens of results). Indeed, they wanted their notion of "relevant" to only include the very best documents since there may be tens of thousands of slightly relevant documents. This very high precision is important even at the expense of recall (the total number of relevant documents the system is able to return) (Brin & Page, p.2).

Google search began this process by using link structure and anchor text but once it gained traction it could use whole populations worth of user search data and cross correlate information making it more useful and accessible. Already, in this scenario, privacy concerns are being ignored and yet information is still within the somewhat preferred behavioural value reinvestment cycle aimed at user and service benefit (Zuboff, p.135).

This is important because a central tenet of contextual integrity as analysed by Nissenbaum is that there are no arenas of life not governed by norms of information flow, no information or spheres of life for which "anything goes." Almost everything — things that we do, events that occur, transactions that take place — happens in a context not only of place but of politics, convention, and cultural expectation (Nissenbaum, p.137). And therefore considering the scope and unilateral power of the Google search engine one can begin to understand the extent of its infringements at scale.

Contextual Integrity and Google Search

Contextual integrity can be understood by the use of examples. As illustrated by Zimmer, it might be acceptable for one to approach a stranger and offer her a hug at a moving religious service, but not in the grocery store. A judge might willingly accept birthday gifts from colleagues, but would hesitate to accept one from a lawyer currently arguing a case in her courtroom. It is deemed appropriate for a physician to ask one their age, but not for a bank teller. Norms of behaviour and thus privacy, vary given their context. In terms of contextual integrity according to Nissenbaum, a privacy violation has occurred when either contextual norms of appropriateness or norms of flow have been breached. For her, norms of

appropriateness dictate what information about persons is appropriate, or fitting, to reveal in a particular context and norms of flow dictate whether its distribution, or flow, respects contextual norms of information flow (Nissenbaum, p.141).

Crucially, with Google being the central disseminating and analysing body of this information, all authority over user information lies with the corporation and is enforced by click- wrap agreements - lengthy, jargon laden, open to update and mandatory digital contracts that users must accept to make use of services - that bind them to the paradigm. One must note that it is almost impossible for societal, economic and political institutions to function without the services of these large technology companies and their products. Without companies like Google or Meta (Facebook, WhatsApp, Instagram) or Twitter the gears of our current information age institutions would grind to a halt. Hence, these corporations *dictate* fundamental rules and norms of social, political and economic functioning in the present day.

Now, as mentioned earlier, Google's success was made possible by its huge stores of user information and the machine capabilities and prediction models it could develop and sell as a result of them. It invented surveillance capitalism. Zuboff notes surveillance capitalism unilaterally claims human experience as free raw material for translation into behavioural data. She describes the functioning of surveillance capitalism as follows 'although some of these data are applied to product or service improvement, the rest are declared as a proprietary behavioural surplus, fed into advanced manufacturing processes known as "machine intelligence," and fabricated into prediction products that anticipate what you will do now, soon, and later. Finally, these prediction products are traded in a new kind of marketplace for behavioural predictions that she calls behavioural futures markets (Zuboff, p.21). By this definition it is clear that Google could not have achieved such lucrative monetisation of user information if understandings and practices of contextual integrity were enforced upon its activities. Users in the Google business model are not treated as individuals with privacy rights but as *sources* of data that are to be used to build *prediction products* for the company's real customers — *advertisers.*

To further expand their already entrenched systems of power any political or social debates about contextual integrity of information and any attempts to unbox their algorithmic blackboxes are actively fought by these companies. Larry Page has historically defended Google's unprecedented information power from public scrutiny with the following statement "In general, having the data present in companies like Google is better than having it in the

government with no due process to get that data, because we obviously care about our reputation. I'm not sure the government cares about that as much" (Zuboff, p.116-117). The statement demonstrates in frank terms Google's contempt for legislation and interference with its activities as well as the unilateral trust demanded from individual users and institutions.

In 2009, the public was made aware that that Google maintains search histories indefinitely. And that data available as raw-material supplies are also available to intelligence and law-enforcement agencies. When questioned about these practices, the corporation's former CEO Eric Schmidt said, "The reality is that search engines including Google do retain this information for some time" (Zuboff, p.33). Insightfully, and of relevance to any attempts to regulate this technological parasite, she reviews that statement by saying that in truth, search engines do not retain, but surveillance capitalism does.

The problems with and arguments for maintaining contextual integrity in the case of the Google search engine also go beyond 'just' infringing individual privacy and concentrating such enormous informational resource into the hands of a single institution. They are that such technology, used in a manner that ignores contextual integrity lays the foundations for systemic social, economic and political inequality, conflict and oppression, especially by those in positions of relative power. Privacy has value beyond its usefulness in helping the individual maintain his or her dignity or develop personal relationships. Most privacy scholars emphasise that the individual is better off if privacy exists and Nissenbaum argues that society is better off as well when privacy exists. She maintains that privacy serves not just individual interests but also common, public, and collective purposes (Nissenbaum, p.150).

There are many things that must be taken into consideration when analysing data shared with third parties. One must know things about those parties such as their such as their social roles, their capacity to affect the lives of data subjects, and their intentions with regard to subjects (Nissenbaum, p.155). These notions and how they relate to privacy are fundamentally challenged by the activities of Google via its search engine. Users are only vaguely, if at all, aware of where and whom their data is being shared with outside of their activity. It is important to ask whether the information practice under consideration harms subjects; interferes with their self-determination; or amplifies undesirable inequalities in status, power, and wealth (Nissenbaum, p.155). This has many inferences and its results can range from exploitation and publication of sensitive information such as credit scores and health records to authoritarian government suppression of dissent and protest via information gleaned from Google search

legally or illegally. As the public learned from the National Security Agency (NSA) leaks in 2013, the capabilities to tap into such reservoirs of information are well within reach of motivated and funded actors. One might perhaps even be forgiven for asking why they have to try so hard because public-private partnerships whose legality and function remain outside the realm of lay awareness have created a legal ecosystem of sharing information with or without consent from users. Another problem also arises with data at scale being unable to accurately portray the intentions or context of the individuals creating it. For example, an individual's google searches relating to a murder documentary might be inferred as criminal intent when actually there could be none. The *context* is important.

Google's original privacy policy states 'Google may share information about users with advertisers, business partners, sponsors, and other third parties. However, we only talk about our users in aggregate, not as individuals. For example, we may disclose how frequently the average Google user visits Google, or which other query words are most often used with the query word "Microsoft." This is misleading. To begin with Google isn't even saying its anonymous data, only that it is aggregate.

In his book Data and Goliath, Schneier elaborates how most of us underestimate just how easy it has become to identify us using data that we consider anonymous (Schneier, p.11). For large sets of anonymous data we might naïvely think that there are so many of us that it's easy to hide in the volume of information. Or that most of our data is anonymous. That's not true. Most techniques for anonymising data don't work, and the data can be de-anonymised with surprisingly little information (Schneier, p.47).

He explains how by using public anonymous data from a 1990 United States census, computer scientist Latanya Sweeney found that 87% of the population in the United States, 216 million of 248 million people, could likely be uniquely identified by their five-digit ZIP code combined with their gender and date of birth. For about half, just a city, town, or municipality name was sufficient. Other researchers reported similar results using 2000 census data. (Schneier, p.47). He continues to argue that Google, with its database of users' Internet searches, could de-anonymise a public database of Internet purchases, or zero in on searches of medical terms to de-anonymise a public health database. And conversely, merchants who maintain detailed customer and purchase information could use their data to partially de-anonymise any large search engine's search data. Likewise a data broker holding databases of several companies might be able to de-anonymise most of the records in those databases. In terms then of our

artefact, not only does it generate volumes of data, that data itself can be used to correlate and make sense of other information. All of these examples violate the essential norms of appropriateness and flow that contextual integrity emphasises.

According to their privacy policy Google will share these easily manipulatable and deanonymizable data sets with advertisers, business partners, sponsors, and other third parties. As an example of how ingrained the previously mentioned public-private-partnership is, in an early and surprisingly contested version of their privacy policy (dated 1999 it does not appear on Google's website but is accessible via the Internet Archive's Wayback Machine which is a database of over 623 billion web pages saved over time) Google has stated quite clearly that they 'will release specific personal information about you if required to do so in order to comply with any valid legal process such as a search warrant, subpoena, statute, or court order' (Hoback, 2013). As their products and services have expanded outwards from the search engine their privacy policy has also gotten more lengthy and layered in euphemism.

Ensuring we maintain a broader perspective to this discussion about the Google search engine one must take into account a few other concepts, such as social constructivism and interpretive flexibility. These ideas are linked to each other. And by reviewing relevant literature with regards to Google's search engine one begins to see the true scope of the philosophical, sociological, technological and legislative task posed to our species *only by Google search*.

Other Concepts

Since the 1980s and 1990s, many of the views about media and information technologies advanced within STS and media studies have been more broadly adopted among communication researchers dissatisfied with the implicit technological determinism of media effects research and the language of "impacts" of new technologies on society, behaviour, and culture (Boczkowski & Lievrouw, 2008). Thus, the role of social interactions in the development of technology and the inherent ambiguities of engineering design are an important way in which to evaluate Google search (Kilker & Gay, 1999). The social constructivist view posits that technology exists in a symbiotic relationship with its users and like a relationship it creates meaning through subjective and interpretive interaction. As new innovations are developed and introduced, their inherent '*interpretive flexibility*' will allow various social groupings to associate different meanings to the artefact (Bijker, 1995). One would like highlight this concept to acknowledge how Google search is interpreted differently by

developers, users and so on because of the inherent way in which technical artefacts represent '*different things to different actors*' (Doherty, Coombs, & Loan-Clarke, 2006). Further, this essay notes that Google's rhetoric and imposition of new social norms has contributed to the way in which individuals interact and understand the end goals of this technology which almost certainly differ vastly from the company's own.

To conclude, this analysis with regards to the Google search engine notes how almost every aspect of the *products of its existence* which are *analysed behavioural data* are used in ways that disregard the concept of contextual integrity. One can argue that this is a result of Google's business model, which is surveillance capitalism, and that the company has knowingly promoted, advertised and *enforced* its surveillance gathering mechanisms to make such infringements of privacy commonplace and accepted in public perception. There is, however, no doubt as to the fundamental need for and role a search engine plays in the analysis and making sense of knowledge. But Google's mission statement 'to organise the world's information and make it universally accessible' must be verifiably made to conclude with the words 'responsibly and with individuals in mind'.

Conclusion

While the Google search engine is the principle object of this essay, it is easy to understand the relevance and importance of developing tangible legislation keeping in mind contextual integrity in an attempt to regulate institutional power in favour of users, especially in a world of increasingly frequent digital interactions, not to mention Facebook co-founder Mark Zuckerberg-esque notions of a metaverse.

The task of developing functional legislation to inculcate and ensure contextual integrity within the framework of existing technologies like Google search is going to, therefore, require reworking foundational aspects of how these technologies and their creators function and behave. But the real problem, one that has proven difficult to overlook by technologists and their companies, is not that it is impossible to imagine alternative versions of technology that serve people and their privacy better but that none may be as profitable.

References

Boczkowski, P., & Lievrouw, L. A. (2008). Bridging STS and Communication Studies: Scholarship on Media and Information Technologies. *The Handbook of Science and Technology Studies*, 949–969.

Bijker, WE (1995) Of Bicycles, Bakelites and Bulbs. MIT Press, Cambridge.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, *30*(1–7), 107–117. https://doi.org/10.1016/s0169-7552(98)00110-x

Doherty, N. F., Coombs, C. R., & Loan-Clarke, J. (2006). A re-conceptualization of the interpretive flexibility of information technologies: redressing the balance between the social and the technical. *European Journal of Information Systems*, *15*(6), 569–582. https://doi.org/10.1057/palgrave.ejis.3000653

Google: Privacy Policy. (1999). Internet Archive. https://web.archive.org/web/19991012225420/http://google.com/privacy.html

Hoback, C. (Director). (2013). *Terms and Conditions May Apply*. Hyrax Films.

Kilker, J, Gay, G. (1999) The Social Construction of a Digital Library: A Case Study Examining Implications for Evaluation. *Information Technology and Libraries*, 60–70.

Nissenbaum, H. (2004). Privacy as Contextual Integrity. *Washington Law Review*, *79*.
Privacy Policy â. (1999). Privacy & Terms Â Google.

https://policies.google.com/privacy/archive/19990609?hl=en-US

Schneier, B. (2021). The Hidden Battles to Collect Your Data and Control Your World Data and Goliath (Hardback) - Common. W. W. Norton & Company.

Tene, O. (2007). What Google Knows: Privacy and Internet Search Engines. *SSRN Electronic Journal*. Published. https://doi.org/10.2139/ssrn.1021490

Wakabayashi, D. (2020, January 16). *Trillion-Dollar Company: Google Reaches Milestone in Market Value*. The New York Times. https://www.nytimes.com/2020/01/16/technology/google-trillion-dollar-market-cap.html

Zimmer, M. (2008). Privacy on Planet Google: Using the Theory of "Contextual Integrity" to Clarify the Privacy Threats of Google's Quest for the Perfect Search Engine. *Journal of Business & Technology Law*, *3*(1).

Zuboff, S. (2020). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power (Illustrated ed.). PublicAffairs.

Hoback, C. (Director). (2013). *Terms and Conditions May Apply*. Hyrax Films.

Wakabayashi, D. (2020, January 16). *Trillion-Dollar Company: Google Reaches Milestone in Market Value*. The New York Times. https://www.nytimes.com/2020/01/16/technology/google-trillion-dollar-market-cap.html

Zimmer, M. (2008). Privacy on Planet Google: Using the Theory of "Contextual Integrity" to Clarify the Privacy Threats of Google's Quest for the Perfect Search Engine. *Journal of Business & Technology Law*, *3*(1).

Zuboff, S. (2020). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power (Illustrated ed.). PublicAffairs.