How Google Creates What We Know

Clarissa Calderon

University of Hamburg

Hammerbrookstr. 42b 1.06A

20097 Hamburg

+49 15773176349

clarissa.calderon9@gmail.com

Abstract

How does the Google Search Engine influence our contemporary practices of knowing? In the past, there have always existed gatekeepers, or individuals who control what information the public has access to. The difference today is the misperception that because algorithms are applied invariably, these new algorithmic gatekeepers are therefore "value-neutral". Instead of being value-neutral I argue that the Google Search Engine, and the algorithms of which it is comprised, actually reinforces a dominant narrative and thus exacerbates extant power structures. In order to argue this point, I ground my analysis in an interdisciplinary theoretical framework developed by Judith Simon, which combines insights from the fields of Social Epistemology, Philosophy of Computing and Feminist Theory. Together, these three fields relate the social to the technological to the relationship between power and knowledge, and allow for an analysis of what it means to assume epistemic responsibility in today's digital age. I attempt to give this analysis some form by showing that the Google Search Engine does not merely disseminate knowledge, but actively shapes that which we deem to be knowledge. I do this, by focusing on two types of information which are particularly susceptible to this kind of shaping: information regarding ongoing news stories and information regarding essentially contested concepts. Assuming epistemic responsibility today entails providing evidence for what we claim to know and revising our beliefs in the light of new evidence, while acknowledging the entangled system of human and non-human agents in which we engage in our practices of knowing.

Keywords: epistemology, google, performativity, responsibility, algorithms, knowledge, technology

I Introduction

Today's view of the plurality of perspectives offered by the internet is often contrasted with media structures of the past when there existed only a handful of broadcasters who disseminated information to the public. These broadcasters and their editors were viewed as information gatekeepers because they held central positions in the network of information of any society, thus the public they broadcasted to in many cases was entirely dependent on them for information regarding the world outside of their immediate experience. As such, they wielded tremendous power in shaping the politics and culture of any society to which they broadcasted.

From the onset, there was a certain optimism associated with the internet. This was largely due to the fact that it was seen as decentralizing the epistemic power that previously resided only in the hands of a few. In the early 2000s, many scholars believed that cyberspace would reduce barriers to information and thus enable the realization of the ideals of democracy (Agre 2002, Barlow 1996). By giving unmediated access to knowledge, it was expected to support a pluralistic diversity of intersecting public spheres. In particular it was the turn toward user-generated web design, what we now call Web 2.0, which brought with it the hope for the democratization of knowledge and information. Rather than a population being subject to the assertions of one source like in conventional broadcasting, the many could now communicate with the many.

However, the view that technology would help us realize our democratic ideals seems to take for granted that each voice would be given an equal vote in the public sphere that is the internet. While each person certainly has the capacity to voice their view, this does not imply that

they have an equal capacity to attract the attention of others. In analyzing commercial news markets (as opposed to states with a strong public service media) Hendricks observes that there is not necessarily exist a greater diversity of news stories. The same goes for the diversity of the news sources that attract attention. He explains that this is because online attention follows a power law distribution.[1]

> These conditions exist because online attention does not follow a normal distribution, but rather a power law distribution: A few players get the bulk of the attention... while everyone else has to fight over the very limited attention at the tail of distribution. Much like the world economy: 1% has 50% of the world's wealth, and the other 99% have to fight over the remaining 50%. (Hendricks, 2019: 29)

Hence, while each person has an equal capacity to voice their opinion, it would be improper to describe their 'vote' as being equal.

Still, another reason why online media is seen to be democratizing is because results on the web are the product of universally applied algorithms rather than the product of a news editor susceptible to all the cognitive or political biases to which humans are liable. However, this view rests on the assumption that algorithms themselves are not biased. Because they are artificial entities they of course do not suffer from cognitive biases like humans, but they do have the capacity to amplify biases formed at the collective level. Within the field of computer science the

[1] In statistics, a power law is a functional relationship between two quantities where a relative change in one quantity results in a proportional relative change in the other quantity (Wikipedia: https://en.wikipedia.org/wiki/Power_law [last accessed: 12.02.19])

adage “garbage in, garbage out” is well known: flawed input data produces a flawed output. Applied to the information market on the web, we see the same phenomenon emerge where software designers intentionally or unintentionally write code that magnifies social biases. While an algorithm can be quite useful in its intended function of quantifying the qualitative inevitably cuts must be made about what is relevant and what is not, and what represents relevance and what does not. However, in the offline world, many scholars of feminist epistemology have called for greater scrutiny over the ethics inherent in epistemology, emphasizing the situated nature of epistemic agents and the societal implications of who is deemed an epistemic authority and who is not. As the data entered into these algorithmic gatekeepers reflect notions of epistemic trustworthiness rooted in our collective imagination, the output will not only exhibit those same notions, but actually compound it.

Therefore, far from being a democratizing force, there is good reason to believe that the internet actually reinforces extant power structures. If this is true, what options does that leave us for being responsible in the ways that we know in our hyperconnected and increasingly digitized world? Judith Simon, a professor of Ethics in Information Technology, has developed a promising interdisciplinary framework that provides guidance on what it means to assume epistemic responsibility today. She combines insights from the fields of Social Epistemology, Philosophy of Computing and Feminist Theory in order to relate the social to the technological to the relationship between power and knowledge, and thus allows for an analysis of what it means to assume epistemic responsibility in today's digital age. First I will delve a bit deeper into her framework. Then, in order to give her theory some form, I will examine the Google Search

Engine and show how it does not simply disseminate information, but actually reinforces extant power structures.

II The Sociality of Knowledge

Social Epistemology is the field of philosophy which investigates how social interactions affect knowledge. A central topic within this field is testimony, where philosophers are particularly concerned with identifying the conditions under which the assertions of others can generate knowledge on the part of the hearer. As such, debates concerning the justifiability of testimony are normally centered around the relationship between hearer and speaker. Considering the sociality of knowledge is sometimes seen as challenging traditional epistemology, a field which focuses on how an individual can be rationally justified in their beliefs in abstraction from their social environment. However, because much of what we claim to know is based on the assertions of others, understanding the conditions for justifiably believing in testimony is a prominent task for anyone who wishes to develop a normative standard of knowledge.

Yet, once we start to consider the sociality of knowledge, we must also recognize its concomitant power relations and the potentials for epistemic injustice they introduce. A few philosophers have ruminated over the relationship between power and knowledge (Fricker 2007, Barad 2003, Simon 2015). Miranda Fricker has developed two conceptions of epistemic injustice that help to elucidate this point. First, testimonial injustice occurs when a speaker suffers a credibility deficit specifically because the hearer has prejudices about the social group to which the speaker belongs (Fricker 2007: 20). Fricker points to an example in the book *To Kill a*

Mockingbird in which an all-white jury refuses to believe the black defendant's testimony due to racial prejudice. It's important to note that rather than being an incidental case of credibility deficiency, testimonial injustice is systematic in that the speaker's credibility deficiency follows him throughout many different aspects of life including political, legal and economic dimensions. That is, a philosopher of science who is not taken as seriously as a scientific researcher at a particular conference is a case of credibility deficiency due to identity prejudice. But because it is highly localized and does not follow its victim into other aspects of social life, it is not the type of testimonial injustice that Fricker is concerned with.

In speaking about testimonial injustice, Fricker essentially calls our attention to the relationship between power and knowledge. However, the conditions for this individualistic type of injustice are dependent on a more pervasive structural type of epistemic injustice, which she calls hermeneutical injustice. This occurs when members of a socially powerless group lack the conceptual resources to make sense of certain distinctive social experiences (Fricker 2007: 1). For example, prior to the 1970s, victims of sexual harassment had difficulty making sense of their experiences and the harms that came with it. Without the words to articulate this even to themselves, they found it nearly impossible to then try to convey their experiences and harms to others. Importantly, the concept of "sexual harassment" had not yet been articulated in large part due to women's socially marginalized position (Fricker 2007: 15). It is with these two conceptions of epistemic injustice that Fricker calls attention to how social identity affects how one gains and expresses knowledge, and underscores the importance of evaluating an epistemic agent within her social entanglements rather than in an artificial abstract.

III Epistemic Responsibility and Entanglements

Because we must recognize the role that power plays when we discuss the sociality of knowledge, we must also speak about responsibility in our knowledge practices. Fricker develops two virtue-based correctives to address each of these forms of epistemic injustice. With regards to testimonial injustice, she develops the notion of *testimonial justice* whereby the influence of identity prejudice on the hearer's credibility judgment is detected and corrected for (Fricker 2007). *Hermeneutical justice* is the virtue whereby the hearer exercises a "reflexive critical sensitivity to any reduced intelligibility incurred by the speaker owing to a gap in collective hermeneutic resources" (Fricker 2007). In other words, the hearer must constantly monitor how the relation between his social identity and that of the speaker impacts the degree to which he finds intelligible what the speaker is saying and how she says it. Both virtues—testimonial and hermeneutic justice—ultimately guard against forms of identity prejudice, and the cultivation of these virtues renders the subject in which they dwell, epistemically responsible.[2]

Fricker's conclusions echo that of Barad's Agential Realism, the insights of which Simon references heavily in her own view of epistemic responsibility. Barad stresses the "fundamental inseparability of epistemological, ontological and ethical considerations" and that accountability and responsibility must be thought of in terms of what matters and what is excluded from mattering (Barad 2007). Importantly for Barad, epistemic practices are performative in that they are productive, and that different practices produce different phenomena. It is in this sense that

[2] Fricker originally calls this being epistemically virtuous, but for the sake of having a unified terminology I refer to it here as being epistemically responsible.

knowledge not only represents our reality, but actually shapes what is and what **will be** (Simon 2015). Ultimately, Barad comes to a similar conclusion as Fricker in stating that we have a "responsibility to intervene in the world's becoming, to contest and rework what matters and what is excluded from mattering" (Barad 2003). In other words, in claiming objective knowledge, we also take responsibility for what is left out, and this is always open to contestation.

Classic assessments of testimonial processes focus on communication between the hearer and speaker where both are abstract but ultimately human entities. Fricker's notion of hermeneutic (in)justice and Barad's epistemological-ontological-ethical framework are useful in that they open up the discussion as to how social structures bear on our epistemic responsibility as hearers, bringing to the forefront the importance of identity and situating both speaker and hearer. However, most of our knowledge practices today take place within systems comprised of both human and technological entities, or *socio-technical systems,* that are networked and deeply entangled. Developing a meaningful notion of epistemic responsibility is no easy task in a world where socio-technical systems have become commonplace and where similar concepts like accountability and agency have proven themselves difficult to define and even more difficult to locate. For instance, who do you blame if your wifi is not working? The internet service provider because they aren't providing the right connection? The company who produced it for making a bad product? The coders who designed it for introducing a bug? The user for not updating the firmware when they ought to? The answer is anything but a straightforward affair.

Before being able to assume or even attribute responsibility, we must first have a clear conception of accountability. In analyzing the difficulty of accountability in a computerized

world, Nissenbaum identifies the “problem of many hands”. When there is a multitude of people who participate, or are in some way involved in producing a computer system, the problem of identifying who is accountable when something goes wrong (a philosophical problem in its own right) is aggravated. Nissenbaum notes a few of the ways in which computing is especially vulnerable to this problem of many hands.

> First, most software systems in use are produced in institutional settings; this could be a private corporation or a government agency. Second, computer systems are usually constructed out of modules where each module may be the work of a team of individuals... When systems grow in this way, sometimes reaching huge and complex proportions, there may be no single individual who grasps the whole system… Fourth, performance in a wide array of mundane and specialized computer-controlled machines – from rocket ships to refrigerators – depends on the symbiotic relationship of machine with computer system. When things go wrong... it may be unclear whether the fault lies with the machine or with the computer system. (Nissenbaum 1997)

Indeed, computer systems are the product of a network of agents each only partly conscious of how their decisions affect the final product. It is easy to imagine how unintended consequences can emerge from this network, given the limited level of participation and foresight of each of the involved agents.

In order to address the reality in which most of our epistemic practices take place, Simon develops her own account of epistemic responsibility that takes into consideration non-human epistemic agents. She borrows the concept of *mindless morality* to distinguish between being responsible and being accountable. According to this concept, "something qualifies as an agent if it shows interactivity, autonomy and adaptability... [and] while agency and accountability do not require intentionality, responsibility does" (Floridi & Sanders 2004). The distinction allows us to view non-human entities, e.g. software, as accountable agents, while still holding humans as the only agents capable of being responsible for the outputs of such socio-technical systems. This is how we can view a machine as being accountable for something that goes wrong, while attributing responsibility solely to its human designer.

Though the concept was originally developed as a solution to the problem of attributing *moral* responsibility in a distributed environment, combined with the insights from feminist theory about the relationship between power and knowledge, it can also address the problem of *epistemic* responsibility in information communication technologies. Simon develops an account of what it means to be epistemically responsible at the individual and collective level. At the individual level, assuming epistemic responsibility today first means acknowledging our current socio-technical entanglements, accepting to be challenged for what we claim to know, providing proof for our claims, updating our beliefs in the face of new compelling evidence, and being sensitive to power asymmetries within entangled socio-technical environments (Simon 2015). However, assuming responsibility at the individual level presupposes proper governance at the collective level, which supports this type of conduct. In order to know responsibly, we need to be able to access and understand the systems which we rely on for so much of our beliefs about the

world. If we cannot understand how these systems operate, we cannot be said to be on guard against epistemic injustice in all its forms. Our lack of access and understanding about how ranking algorithms work is problematic for assuming epistemic responsibility. We cannot provide proof for our claims to knowledge if we do not understand the algorithms that presented them to us in the first place. We cannot be critical of potential power asymmetries if we do not understand how the algorithm arrived at its search results. In the next section I discuss a few of the ways in which epistemic injustice can be exacerbated by computer systems and the use of algorithms.

IV Biases and the Google Search Engine

Friedman & Nissenbaum outline three types of biases most relevant to computer systems, two of which I will discuss for my analysis of the Google Search Engine. First, a pre-existing bias is when a computer system embodies biases that exist prior to the creation of the system. These biases can stem from an individual or an institution and can enter the system through an explicit conscious effort or despite the best intentions of the designer. For example, in determining an applicant's credit risk, an automated system that advises on loan applications, may negatively weight applicants who live in low-income or high-crime neighborhoods, a practice known as redlining. To the extent the program embeds the biases of clients or designers who seek to avoid certain applicants on the basis of group stereotypes, the bias of the automated system is preexisting (Friedman & Nissenbaum 1996).

Second, technical biases emerge from the resolution of issues in technical design and can be thought of as the natural limitations of computer tools. When attempting to quantify the qualitative, or in trying to "make human constructs amenable to computers," inevitably many factors are left out (Friedman & Nissenbaum 1996). Think of, for example, the limitation a page imposes on any list that is the result of a search. A first page can only exhibit so many results before other results are pushed back to the second or third page. While results on subsequent pages may be just as relevant, their position makes them less visible to the searcher and therefore less likely to be selected. In this case, the algorithm chosen to rank flight options become critically important, bestowing a significant advantage to the top result. This kind of bias is a result of the technical design, and does not necessarily carry a moral charge. Nevertheless, technical limitations impact our epistemic practices as certain results will inevitably receive preferential treatment (so to speak).

The Google Search Engine in particular aggravates these biases and challenges our ability to assume epistemic responsibility. I show this by referring to two incidences of exclusion of "what matters", to use Barad's terminology. The first is a concrete example, and the second is a concern with how the pervasive use of the GSE can result in the homogeneity of evaluative notions that require contestation.

Safiya Noble has observed that when "professor style" is searched for in GSE images, the results retrieved are mostly images of Caucasian men in western-clothing. One can see a few women, and an even lower frequency of any people of color. It is unlikely that this was an intended effort of Google's programmers. This image of a professor originates in a bias that predates the creation of computers entirely and persists largely unnoticed. These results are a

reflection of our social biases. Though universally applied, it is still very much bias, presenting results that are the product of a dominant class, and though the algorithm can be said to be accountable for the results, it is ultimately the programmers and the institute that is Google that is responsible for this shirking of epistemic responsibility.

Perhaps even more worrisome is the potential that our over-reliance on Google's search engine may lead to the eventual homogeneity of concepts that should be continuously contested. Gallie came up with the term "essentially contested concepts" to describe concepts where disputes about the concept's essential meaning are central to the concept itself; instances where we do not know or cannot agree on what the ideal is (Gallie 1955). Gallie refers specifically to evaluative notions like art and justice. While essentially contested concepts provide the venue for a particular sort of adversarial discourse, Google's ranking algorithm seems to render that venue null by introducing a hierarchy of interpretations based on popularity. The top result will reflect the views of a dominant voice in the debate, someone or something who is seen as having epistemic authority on the subject, and results on subsequent pages are likely to reflect the views of those who have by some measure been judged less credible. However, if we commit ourselves to the virtue of hermeneutic justice, then we commit ourselves to the possibility that views of these evaluative notions that are judged as less credible may not be receiving their due credit.

Bibliography

Agre, P. E. (2002). Real-Time Politics: The Internet and the Political Process. The Information Society, 18(5), 311-331. doi:10.1080/01972240290075174

Barad, K. (2003). "Posthumanist Performativity: Toward an Understanding of How Matter Comes to Matter." Signs: Journal of Women in Culture and Society 28 (3): 801-831.

Barad, K. (2007). Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning. Durham, Duke University Press.

Barlow, J. (1996). A Declaration of the Independence of Cyberspace. [online] Electronic Frontier Foundation. Available at: https://www.eff.org/cyberspace-independence [Accessed 28 Feb. 2019].

Floridi, L. & Sanders, J. (2004). On the Morality of Artificial Agents. *Minds and Machines*. 14: 349-379. https://doi.org/10.1023/B:MIND.0000035461.63578.9d

Friedman, B. and H. Nissenbaum (1997). Bias in Computer Systems. Human Values and the Design of Computer Technology. B. Friedman. Cambridge, Cambridge University Press: 21-40.

Fricker, M. (2011). Epistemic injustice. Oxford: Oxford University Press.

Gallie, W. B. (1955). Essentially Contested Concepts. *Proceedings of the Aristotelian Society.* 56 (1):167 - 198.

Hendricks, V. and Vestergaard, M. (2019). Reality Lost.

Nissenbaum, H. (1997). Accountability in a Computerized Society. Human Values and the Design of Computer Technology. B. Friedman. Cambridge, Cambridge University Press: 41-64.

Simon J. (2015) Distributed Epistemic Responsibility in a Hyperconnected Era. In: Floridi L. (eds) The Onlife Manifesto. Springer, Cham

En.wikipedia.org. (2019). Power law. [online] Available at: https://en.wikipedia.org/wiki/Power_law [Accessed 28 Feb. 2019].