MedieKultur | Journal of media and communication research | ISSN 1901-9726

Article – Theme section

Beyond the bubble

Three empirical reasons to re-conceptualize online visibility

Anders Koed Madsen

MedieKultur 2016, 59, 6-27

Published by SMID | Society of Media researchers In Denmark | www.smid.dk

The online version of this text can be found open access at www.mediekultur.dk

Google is a powerful player in deciding how the world is represented to information-seeking citizens in a digitized knowledge-society. Eli Pariser has been influential in arguing that the company's algorithm leaves its users trapped in a biased 'filter bubble' in which information about the world is tailored to their preferences by algorithms. This paper proposes a shift in focus away from the metaphor of the 'bubble' when we try to understand how 'real world representations' are shaped by the dynamics of online visibility. Instead of a mono-causal focus on the algorithm, it suggests focusing on the distributed set of selection mechanisms that enable web users to navigate a world of 'big data'. The paper suggests a conceptual move from 'bubbles' to 'visions' to understand online visibility. It motivates this suggestion through three empirical analyses of the selection mechanisms involved in making the issue of synthetic biology visible to users of British Google from February 2011 – February 2012. The paper uses the findings of these analyses as a basis on which to suggest theoretical, empirical and practical implications for future studies on the impact of the digital on 'real world representation'.

Keywords

Filter bubble, web-vision, synthetic biology, Google, socio-technical systems

Introduction

Google is a powerful player in organizing the representations of the world that information-seeking people find in the digitized knowledge society (Battelle, 2006). While such organizing capacities were previously tied to professions such as librarians and journalists, it is increasingly the case that citizens and decision-makers trust Google's search engine to direct them to relevant sources of knowledge (Vaidhyanathan, 2011). This reconfiguration of selective powers has sparked a need for understanding and explicating the way Google shapes the 'real world representations' that guide the attention of its users.

This calls for the development of a theoretical framework to enhance our understandings of the dynamics of online visibility in a situation in which Google's 'big data' techniques increasingly shape the way engaged citizens see the world. Such a framework must provide two things. First, it must establish a foundation from which to understand the kind of selection mechanisms that are influential in making specific websites visible to Google's users. Second, it must find a way to conceptualize the role that these mechanisms play in contemporary practices of explorative inquiry.

Eli Pariser's (2011) concept of 'filter bubbles' is one of the most influential attempts to establish such a framework. It claims that Google's search engine gives rise to a 'bubble' within which its users find information that is tailored to their taste by a proprietary algorithm. Other frameworks, however, have suggested a need for a theoretical focus that goes beyond the algorithm and the SERP to explain online visibility. One suggestion along this line of reasoning is to see Google as giving rise to 'web-visions' that perform the world on the basis of a distributed chain of socio-technical selection mechanisms (Madsen, 2012).

This paper provides three explorative empirical analyses of the way the British version of Google represented the issue of synthetic biology from February 2011 to February 2012. The findings of these analyses motivate the future use of the concept of 'visions' for theorizing about the dynamics of online visibility. In short, they show that the representations that Google provides of this issue cannot be adequately understood by limiting empirical attention to the algorithm. The analyses illustrate how the attempt to understand Google's search results as a 'bubble' risks overstating the importance of the algorithm as a selection mechanism in the contemporary media landscape.

Instead, the paper suggests we think of Google's users as people who obtain a 'web-vision' of a specific issue. This vision is defined as the environment of information opened up by the network of websites that can be explored by following hyperlinks in the search results. This is a way of generating sensitivity towards the influence of non-algorithmic selection mechanisms.

The paper illustrates the importance of working with such a broad empirical operationalization through three empirical claims. First, it shows that there is a difference in the geographical origins of the websites that are visible in the SERP and the broader vision. Second, it indicates that the reason for this difference is that the latter is more sensitive to American events related to synthetic biology than the former. Third, it shows that these

effects of Americanization are partly due to the fact that British Google users share semantic with the Americans.

In combination, the three explorative analyses serve to open a different trajectory of research than the one set by the concept of the filter bubble. First, they provide a foundation for seeing the mechanisms involved in 'real world representation' as distributed beyond the algorithm. They show that the decision to work with a broader empirical unit of analysis than the SERP allows one to see online visibility as something that is shaped by a complex entanglement of human and non-human actors. Second, they provide a foundation for a different conceptualization of Google's role in performing the world citizens in an information society in which 'big data' techniques are gaining importance. Instead of seeing the search interface as something that conceals the world through algorithmic programming, the analyses support seeing it as something that makes the world visible in specific ways, depending on the characteristics of the distributed selection chains.

Clarifying the difference between 'filter bubbles' and 'web-visions'

The media studies literature is filled with relevant studies of how information is ordered on the web. Studies on information loss and instability in search engines (Bar-Ilan, 1999), the politics of search (Introna & Nissenbaum, 2000), the power-law distribution of Internet traffic (Hindman et al., 2003) and the bias of search (Thelwall and Vaughan, 2004) are just a few examples. However, this paper focuses on the concept of 'the filter bubble' because the bubble metaphor has been enormously influential in framing discussions about search and democracy during the last five years. Consequently, the SERP and the power of the algorithm have been given undue focus in discussions about representation and online visibility. The motivation of the paper is to give empirical reasons for challenging this tendency.

More specifically, the aim of the paper is to use longitudinal empirical data to question two important ontological assumptions that underpin Pariser's metaphor of the bubble. The first is that the user, the algorithm, and reality are distinct entities. The second is that the algorithm is the mediating device between the other two. Pariser is explicitly arguing that Google's algorithm should be seen as a technology situated between the user and reality in the same way as a camera lens is situated between the photographer and his motif. Accordingly, it is conceptualized as a filter that can represent the world in a more or less distorted way. Google's interface is argued to be a place where "the [users] end and the technology begin" (Pariser, 2011, p. 13). This is why the SERP is taken to be the relevant empirical object of study if one wants to study online visibility on the basis of the filter bubble theory. The SERP is seen as the place where the logic of algorithmic representation gets its most immediate empirical manifestation.

The filter bubble theory also carries with it a normative assumption that will be problematized throughout the paper. This assumption is that the democratic value of a bubble should be accessed on the basis of whether or not it hides important parts of the web from

the user. Following other scholars in Internet research (Sunstein, 2006; Gerhards, 2010; Schneider & Foot, 2005; van Os, Jankowski & Vergeer, 2007), Pariser indicates that the algorithmically constructed bubble runs the risk of creating an 'echo chamber' that conflicts with the dream of a common civic sphere in which people engage with information that falls outside their comfort zone (Pariser, 2011). The extent to which the bubble is representative of the sphere is seen as the crucial normative question. In other words, Pariser wants Google's representations of the world to be representative of the web sphere as a whole (Schneider & Foot, 2005).

The concept of web-visions has certain similarities to the concept of filter bubbles. Most importantly, it is born out of an interest in the way web users are exposed to information about issues of their interest. In fact, a web-vision can be defined as the environment of information that is open to exploration by a web user interested in specific issues. However, it offers a different way of understanding the creation of such environments of online visibility than the theory of the filter bubble. In short, it is an approach to online visibility that parts with all three assumptions listed above.

Most importantly, a web-vision is taken to be a result of a set of selection mechanisms that is distributed beyond the algorithm (Madsen, 2012). The concept is inspired by recent work (Rogers, 2013; Marres, 2012; Mager, 2012) that clashes with the idea of thinking about the SERP as a privileged empirical site where the influence of the users ends and the technology begins to shape their attention. More specifically, the concept of web-visions builds on Richard Rogers' suggestion of looking at the SERP as an 'ordering device' (Rogers, 2013) that functions as a starting point from which the user can follow the medium and repurpose its logic to see the world in a specific way. In other words, the SERP is taken to be a device that opens up explorative inquiry on the part of the user, who can use its outlinks as one instance in his or her process of inquiry.

This way of thinking about online visibility provides a focus on the influence of non-algorithmic selection mechanisms that are difficult to detect by looking at the SERP. It entails thinking about algorithms as one among many selection mechanisms that guide the attention of a Google user and suggests a way of thinking about the SERP as one among many empirical sites where the power of selection is manifested. However, this does not mean that the algorithm and the SERP are irrelevant elements in establishing the web-vision of a Google-user. The SERP is the first guide to the world that meets a person who explores the world through Google, and it is arguably influenced by the company's algorithms. However, the analyses in this paper will illustrate that the choice of taking the SERP as the empirical site in which the way Google guides one's attention is made visible comes with a risk of over-emphasizing the importance of algorithms in shaping the 'real world representations' that become visible to the end users.

The concept of web-visions thereby suggests a shift in theoretical focus that can help overcome some of these shortcomings. This shift is motivated by the fact that interaction with a web-based information environment will most often involve a combination of

searching and browsing. As argued by Ferrara (2008), people will often select a search result that is closest to the topic they have in mind and then follow the links on that page to find their target information. In that way, searching and browsing come to function as a single behavior.

British queries for synthetic biology as the empirical case

The research design of this paper tests the empirical relevance of working with a theoretical distinction between filter bubbles and web-visions and highlights the different consequences of thinking about online visibility through these concepts. More specifically, the paper aims at providing a first explorative attempt at answering the following two questions. Are there relevant differences between the information environment made visible on the SERP and the information available in the network of websites that can be browsed by following hyperlinks on the websites in the SERP? If yes, what are the selection mechanisms that make the SERP differ from the information encountered when browsing?

These questions can only be answered by looking at a specific issue. The rest of the paper will, therefore, present analyses of the way the British version of Google (google.co.uk) guided the attention of its users on the issue of synthetic biology from February 2011 – February 2012. Synthetic biology is the latest attempt to utilize human engineering to optimize evolutionary processes in biological organisms and make them perform desired functions (The Presidential Commission for the Study of Bioethical Issues, 2010). As a scientific practice, synthetic biology has been met with hopes and fears since its introduction. When the data for the study were collected, synthetic biology was, accordingly, an ill-defined and controversial practice that had yet to present hard scientific results. This made it a good case for testing how Google (and the media ecology around the search engine) drew boundaries around this unsettled issue and made it visible to the end user.

It was decided to use the British version of Google on a computer in London to test the effects that Google has on online visibility around the issue. The fact that the UK has a specific tradition of publicly discussing emerging biotechnologies made it probable that there would be enough online material on the issue to conduct a proper study. The combination of the British location and the issue of synthetic biology made a good case for answering the questions set out above. More specifically, if the longitudinal analysis over the course of a year did not succeed in proving a difference between the information obtained from the SERP and the broader hyperlinked website environment, there would be no need to work with an empirical distinction between bubbles and visions. One would, then, be able to draw conclusions about the type of information available in the latter by attending to the former. Accordingly, it is the empirical relevance of this distinction that the paper sets out to explore.

Operationalizing bubbles and visions as comparative units of analysis

The research question posed above can only be answered on the basis of a clear operationalization of the two empirical objects of analysis – the bubble and the vision.

It has already been mentioned that the favored operationalization of the filter bubble is the SERP. In the analyses below, the bubble is, therefore, operationalized as the top twenty URLs returned in response to a query for "synthetic biology". Furthermore, it was decided to depersonalize the search results by adding "&pws=0"[1] to the search URL. This was done because the specific question concerning the difference between information encountered on the SERPs and the information environment opened by the network of hyperlinked sites was more easily answered by comparing depersonalized versions of the two empirical objects.

The web-vision has previously been defined as the environment of information opened up by the network of websites that is opened by the hyperlinks in websites suggested by the SERP. It is the network of information that a user could potentially browse in his or her inquiry about a specific issue. The web-vision was, therefore, operationalized as a network in which websites are depicted as dots and links are depicted as lines. This was done through Issue Crawler,[2] which is a server-side software that captures outlinks from a group of preselected 'seed sites'. In order to depict a web-vision, it was set to trace outlinks from all twenty URLs in the SERP. Since this can be done in various ways, it was necessary to decide on a few pragmatic criteria to ensure that the resulting network had a size large enough to include the kind of environment that a user would browse but small enough to be subjected to qualitative analysis.

The first choice was to set the crawler at a depth of two, which means that the websites linked by the sites linked to the URLs in the SERP marked the empirical boundary of the web-vision. This choice was based on the assumption that most people find relevant knowledge within a two-click distance from the SERP. The next choice was to clean the pool of websites returned from Issue Crawler. This was done to ensure that the websites in the network were relevant in relation to the issue of synthetic biology. This choice was based on the fact that people will browse the websites that are closest to the topic they have in mind (Ferrara, 2008).

This cleaning was done in three ways. First, websites were automatically excluded if they received less than two inlinks from the other sites. This was done by using the so-called co-link analysis in Issue Crawler. Second, Google Scraper was used to exclude sites that had never mentioned synthetic biology. The motivation for using these metrics was, in line with the argument above, to keep the web-vision close to the relevant topic of synthetic biology. Both of these forms of cleaning were automated.

The third step in cleaning the web-vision was done manually. It was decided to remove all websites whose visibility was the result of irrelevant links, such as common and generic links to licenses of Creative Commons. The author qualitatively evaluated every link. This means that it was possible to keep a website such as Creative Commons in the visualiza-

tion if its inlinks, for instance, had to do with relevant issues such as "open source biology". This manual cleaning made web-visions less likely to differ from SERPs on grounds that had nothing to do with the issue of synthetic biology.

The web-visions were finally given a visual form by exporting information about the remaining pages and their links into UCI net,[3] which is a software that makes it possible to draw a network that positions nodes with many shared links close to each other (Borgatti, Everett & Freeman, 2002). Whereas the visual manifestation of the filter bubble is the ordinarily-ranked lists of URLs in the SERP, the web-vision is, accordingly, depicted as a network of interlinked websites. This network is taken to represent the environment of websites that a user could potentially browse when exploring the issue of synthetic biology from the SERP. Pages with high 'betweenness centrality' were interpreted as the most accessible for such a user because these pages would have many browsing paths leading to them.

After having decided on the operationalization of the SERP and the web-vision, the next step was to define the parameters by which these two empirical objects were to be compared. Since the analysis has an exploratory aim, it was deliberately decided to keep the parameters simple and somewhat under-theorized. Three parameters were chosen. One concerned the *newness of visible websites*, which refers to the percentage of new URLs that the SERP and the web-vision make visible to the user every second month. The position of the SERP and the web-vision on this parameter should be thought of as a continuum. A complete change of visible websites from one month to the next would indicate extreme fluidity in the visible content, whereas zero new websites would indicate extreme inertia.[4] In order to compare this temporal element in the SERP and the web-vision, it was decided to collect data every second month from February 2011 to February 2012 and to analyze the change in fluidity and inertia from month to month.

The other two parameters by which the SERP and the web-vision are compared do not have the same temporal aspect as that of newness. The second parameter concerns the *professional identity of visible websites*, and the coding on this parameter was done on the basis of the self-presentation of the websites in relation to their professional status. The typology that was used to code websites manually in the SERPs and the web-visions were the following; 'news/magazines', 'public engagement/social science/ethics', 'natural science', 'commercial organizations', 'governmental agencies', 'funding agencies', 'science blogs' and 'other'. The third parameter concerns the *geography of visible websites*, which is used as a reference to their country of origin. The analytical categories used to guide the coding of this parameter were the following: 'USA', 'UK', 'global', 'other Europe', and 'other world'. The category of 'global' was taken to include websites that did not have a specific geographical home but, rather, presented themselves as outlets for a geographically dispersed set of actors.

The second and the third parameters, accordingly, served to supplement the analysis of fluidity and inertia with insights into the characteristics of the actors made visible by the SERP and the web-vision, respectively. It should be emphasized that the coding was done

in an inductive, ad hoc fashion, which means that the categories were developed in interaction with the data from the first month and, then, used for the rest of the year.

It is on the basis of this operationalization of bubbles and visions that the paper puts forward three empirical claims that point to a need for a new approach to the dynamics of online visibility.[5] In combination, they question the ontological assumptions underneath the concept of the 'bubble' and indicate the relevance of conceptualizing online visibility with less focus on the algorithm and the SERP.

1st empirical claim: 'Real world representations' differ between bubbles and visions

The core purpose of the research design is to assess the empirical relevance of working with a distinction between the concepts of 'bubbles' and 'visions'. The first methodological step in this assessment was to conduct a comparison of their scope (the number of web pages they contain) and fluidity (the percentage of new web pages compared to two months before) in the seven moments of data collection from February 2011 to February 2012. The results of this descriptive analysis are shown in Table 1.

Google UK	Feb 11	Apr	June	Aug	Oct	Dec	Feb 12
Scope – SERP	20	20	20	20	20	20	20
Fluidity – SERP	-	20%	15%	10%	20%	30%	35%
Scope – Web Vision	82	82	68	85	85	95	76
Fluidity – Web Vision	-	39%	34%	27%	36%	48%	42%

Table 1: Scope and fluidity in the SERPs and web visions produced by querying the British version of Google for "synthetic biology" from February 2011 to February 2012.

Due to their operationalization as the first twenty websites returned by Google in response to a query, it is not surprising that the SERPs remain constant in scope across the period. However, the web-visions they give rise to do change on this parameter. The web-vision of June 2011, for instance, makes sixty-eight websites visible, whereas the web-vision of December 2011 makes ninety-five sites visible. We know from the operationalization that the web-visions are built on the basis of the SERPs, and Table 1 shows that this change in scope is somewhat explained by the fact that each SERP exchanges between 10-35% of their websites every second month. When a SERP includes new websites compared to the previous month, it is not surprising that these new websites will also contain links to new sites that discuss the topic of synthetic biology. This will inevitably affect both the scope and fluidity of the web-visions that are operationalized as the sites neighboring the SERP.

Even though the fluidity of the web-visions is larger than the fluidity of the SERPs, it is quite clear that they are in what could be called a 'synchronized flux'. When fluidity is low in one, it is also low in the other. This synchronization might indicate that web-visions are

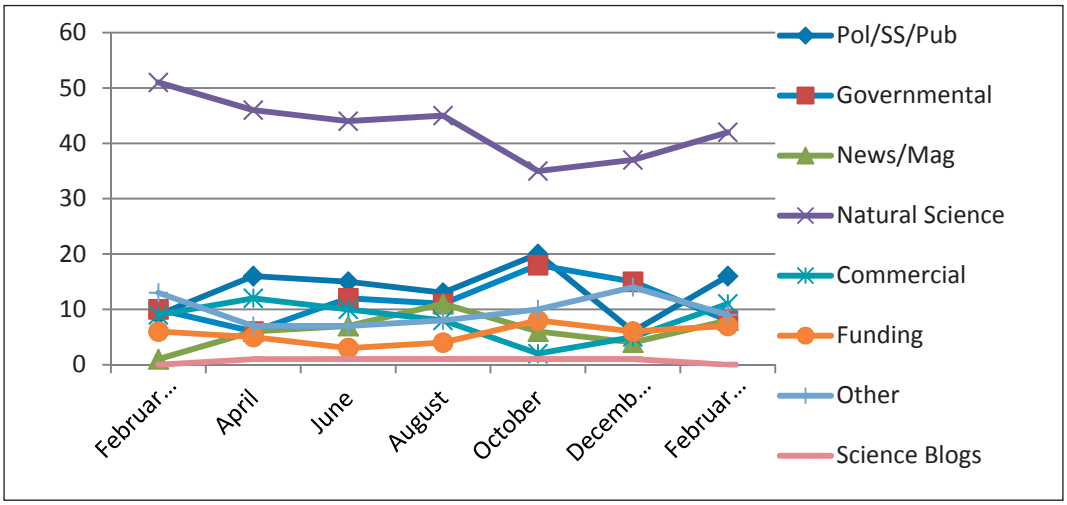

Figure 2: The percentage of specific types of professional websites in relation to the whole scope of the web vision in the period from February 2011 to February 2012 (e.g. 51pct. of the visible websites in the web vision of February 2011 identify themselves as belonging to the natural sciences)

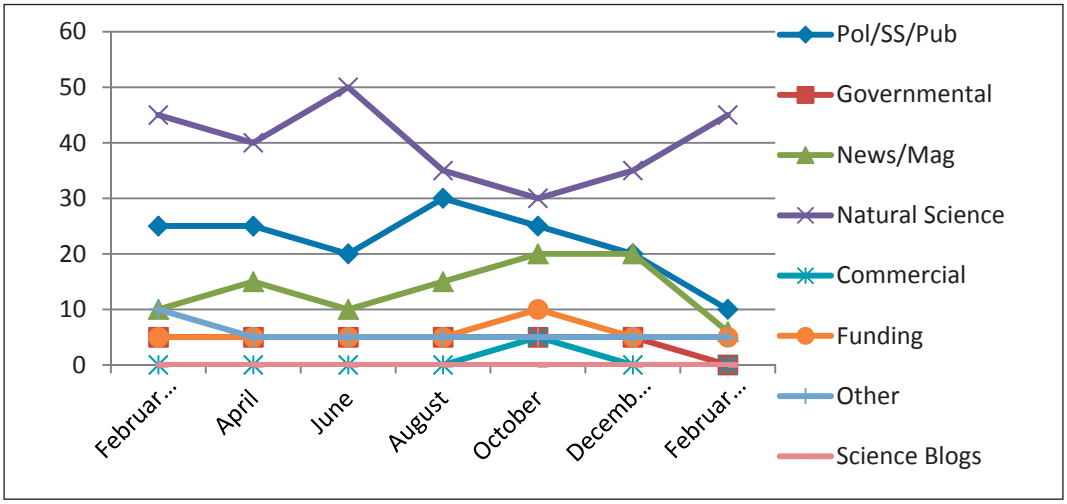

Figure 3: Figure 2: The percentage of specific types of professional websites in relation to the whole scope of the SERP in the period from February 2011 to February 2012

nothing but 'enlarged mirrors' of SERPs. If this is the case, it would be less relevant to conceptualize bubbles and visions as different empirical objects of analysis. One could, then, infer knowledge about the characteristics of the latter by looking at changes in the former. But Figures 2 to 5 question the hypothesis that the web-vision is nothing but an 'enlarged mirror' of the SERP. They show the results of descriptive analyses that illustrate how the

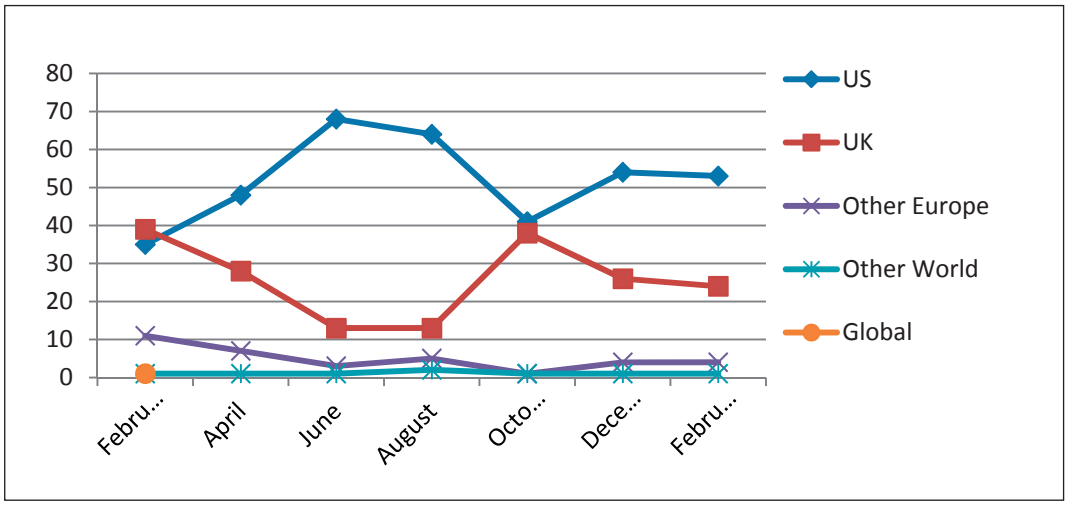

Figure 4: The percentage of websites with specific geographical origins in relation to the whole scope of the web vision in the period from February 2011 to February 2012 (e.g. 39 pct. of the visible websites in the web vision of February 2011 were British)

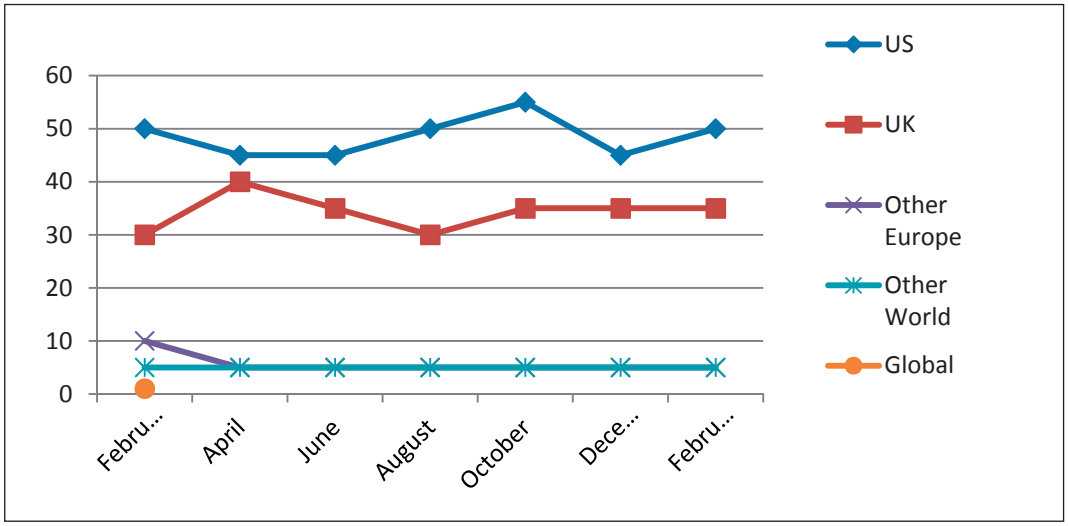

Figure 5: The percentage of websites with specific geographical origins in relation to the whole scope of the web vision in the period from February 2011 to February 2012

seemingly 'synchronized flux' in table 1 hides important differences between the SERP and the web-vision when it comes to their geographical composition.

If we start by looking at the composition of the SERP, it is clear from Figure 3 and Figure 5 that its fluidity has a radical influence neither on the type of professions it makes visible nor on its geographical composition in the period of data collection running from February 2011 to February 2012. Despite the fact that the SERP does change a substantive

number of its websites throughout the year, it remains dominated by American websites that identify themselves as belonging to the natural sciences in all seven moments of data collection. The interesting question, then, is whether this 'momentum' in the SERP is mirrored in the web-vision as was the case with the fluidity parameter above.

When we look at the web-vision that is built from the SERP, it is clear from Figure 2 that it mirrors a stable dominance of natural scientists – even in the months in which up to 35 percent of its websites were changed. However, Figure 4 indicates that the web-vision differs from the SERP by having a much more unstable geographical composition throughout the period of data collection. For instance, the web-visions of February 2011 and October 2011 give equal visibility to American and British websites, whereas this visibility is radically altered in June 2011 and August 2011 when American sites dominate. This difference between the SERP and the web-vision (Figures 5 and 4, respectively) proves that the latter cannot be reduced to an 'enlarged mirror' of the former when it comes to the parameter of geography. The web-vision must, accordingly, be treated as a distinct empirical entity that has its own dynamic and is influenced by distinct selection mechanisms. This proves the empirical relevance of working with a theoretical distinction between bubbles and visions as well as working with a broader empirical object of analysis than the SERP when theorizing about the kind of 'real world representations' that Google users encounter.

2nd empirical claim: Event-driven linking behavior is a selection mechanism that creates difference

The first empirical claim of the paper was that there is a difference between the SERP and the web-vision in relation to their geographical composition across time. This difference means that there must exist selection mechanisms that make the latter different from the former. Whereas the SERP can be seen as a direct outcome of the way Google's algorithm processes digital traces, this is not the case with the web-vision, which represents the websites that a user encounters when exploring neighboring sites to the SERP. Therefore, there must be non-algorithmic selection mechanisms at play that make the web-visions guide attention to American websites in some months and not in others (in a period in which the geographical composition of the SERP remains stable).

However, the details of these mechanisms cannot be identified by looking at the tables and figures shown above. The second empirical contribution of the paper, therefore, is to explore these mechanisms by taking a closer look at the structural composition of link patterns between websites in the specific web-visions that were identified as having a different geographical composition than the SERPs from which they were built. More specifically, this was done by coloring nodes in the network visualizations of the web-visions based on their geographical origin. Figures 6 to 9, accordingly, illustrate the network of British (blue) and American (red) websites that were visible in the web-visions of February 2011, June 2011, October 2011 and February 2012. Comparing the web-visions of these months gives a

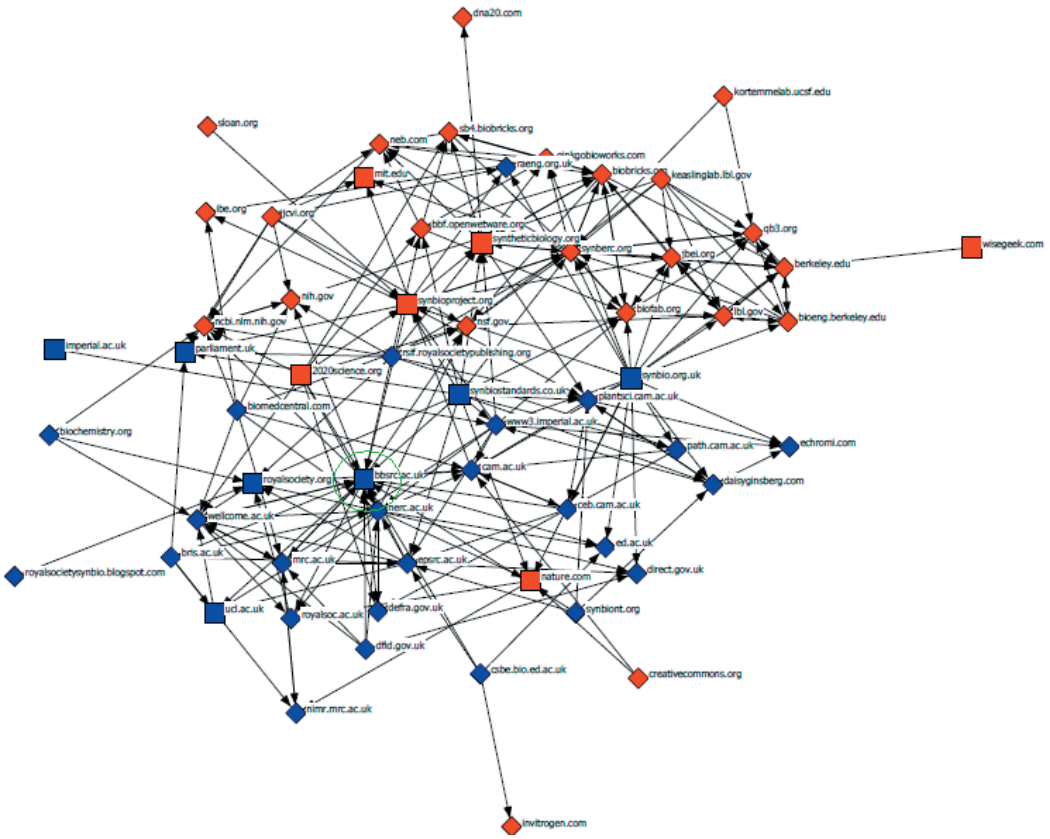

Figure 6: The web vision resulting from querying the British version of Google for "synthetic biology" in February 2011

good impression of the changes in their geographical composition throughout the year of data collection, and it indicates specific types of linking behavior that make the web-visions in June 2011 and February 2012 different from the SERPs from which they are built.

If we start by looking closer at the web-visions of February 2011 (Figure 6) and October 2011 (Figure 8), it is clear that they are both dominated by two distinct geographical clusters that distribute attention primarily to their own websites. One cluster contains American sites, and it is dominated by research centers and networks of American scientists working with synthetic biology. The other cluster contains British sites, and it is dominated by public institutions such as the BBSRC, which is a research council that funds bioscience research in the UK. The two clusters, accordingly, are generating quite different forms of online visibility around the issue of synthetic biology. The fact that they both have a strong visibility in the web-visions in February 2011 and October 2011 is not surprising if we look back at Figure 4. From this figure, it is clear that, in those months, British and American websites were equally visible and, thus, the web-vision and the SERP were somewhat similar.

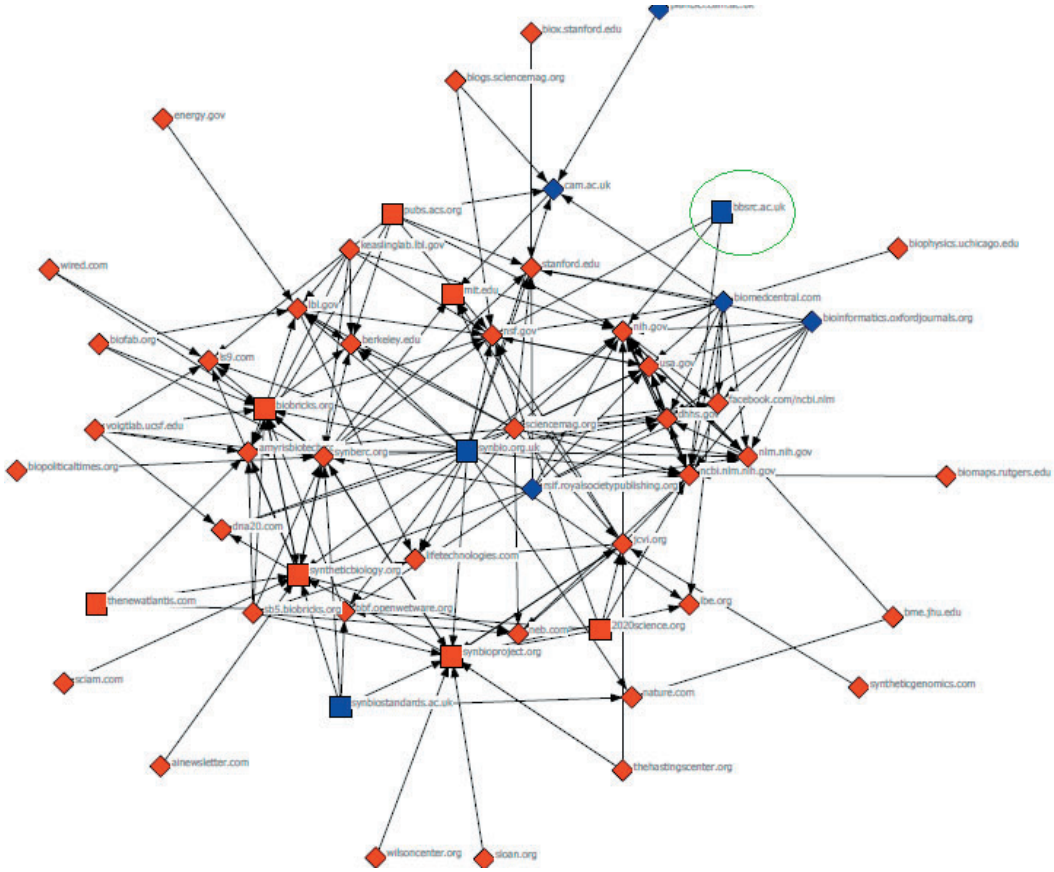

Figure 7: The web vision resulting from querying the British version of Google for "synthetic biology" in June 2011

However, whereas the American websites maintain a strong visibility in the web-visions across the whole year of data collection, it has already been indicated by Figure 4 that the British websites lose visibility in some months. At the same time, it has been shown that this drop in visibility cannot be traced back to geographical changes in the SERPs, whereas the number of British websites remains more or less constant. It was this finding that led to the conclusion that there must be selection mechanisms at work that make the British websites lose visibility in the web-vision without influencing their visibility in the SERP.

A closer look at the structural composition of the web-visions in June 2011 (Figure 7) and February 2012 (Figure 9) indicates that the linking behavior of specific American websites is such a mechanism. In a comparison of these two web-visions with the web-visions of February 2011 (Figure 6) and October 2011 (Figure 8), it is clear that June 2011 and February 2012 are the months when an otherwise strong British cluster loses its visibility. A manual comparison of the specific linking patterns that make up the British cluster in the web-visions of February 2011 and October 2011 with the ones that make it up in the web-visions of June 2011 and February 2012 makes it possible to trace the lack of visibility in the British cluster during the latter two months back to a significant change in the inlinks received by

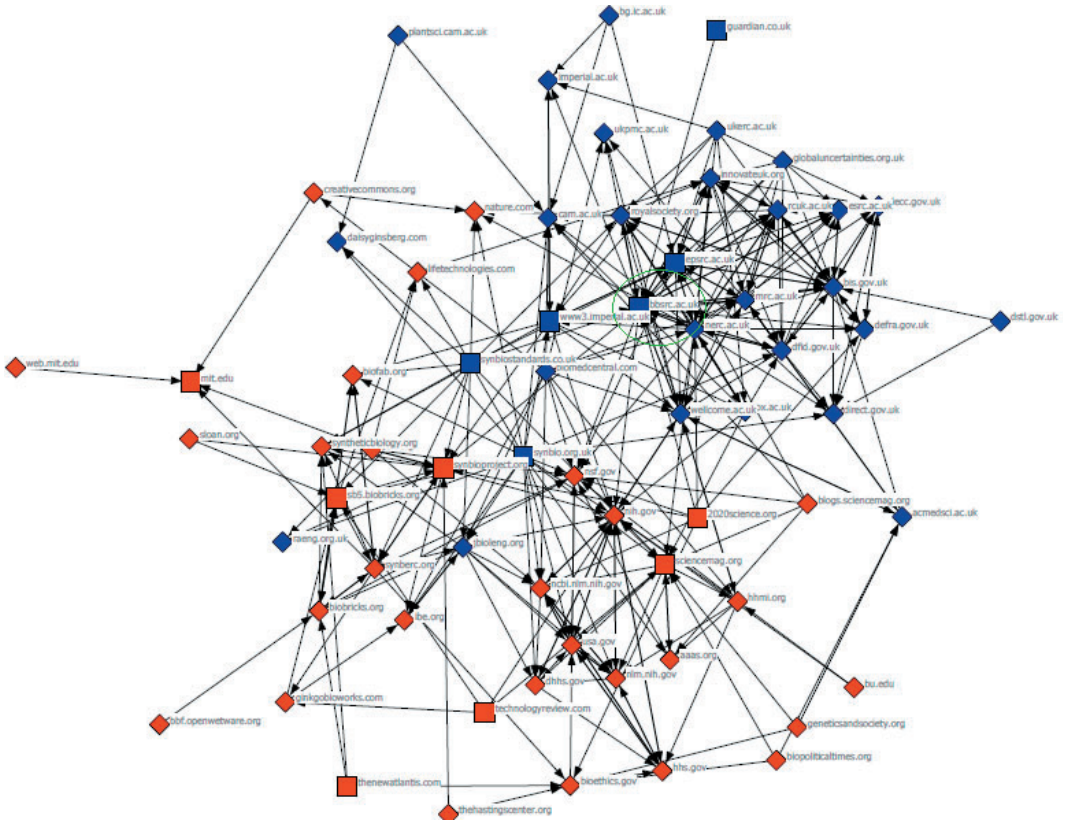

Figure 8: The web vision resulting from querying the British version of Google for "synthetic biology" in October 2011

the most central websites in this cluster. The details of the linking behavior were detected by browsing the backlog of Issue Crawler, which saves information about all the specific links made by each site in the network.

A telling example of this kind of change can be seen by comparing the patterns of links around the website of BBSRC (marked with a green circle) in Figures 6 to 9. BBSRC is central to the web-visions of February 2011 and October 2011 when it is interlinked with websites in both the British and American cluster. However, measures of centrality indicate that this position is lost in the web-visions of June 2011 and, to some extent, February 2012. When measuring its 'betweenness centrality', it comes in second of all the visible sites in both February 2011 and October 2011. This is very different from June 2011 when it ranks as number 43 and February 2012 when it ranks as number 15. A central difference between the structural position of the BBSRC in these months is that it receives links from American sites such as '2020 Science' and 'The Synthetic Biology Project' in February 2011 and October 2011, whereas it does not receive such links in June 2011 and February 2012.

By exploring the context around these linking behaviors, it becomes apparent that the American links help make BBSRC a central website in the web-visions of February 2011 and

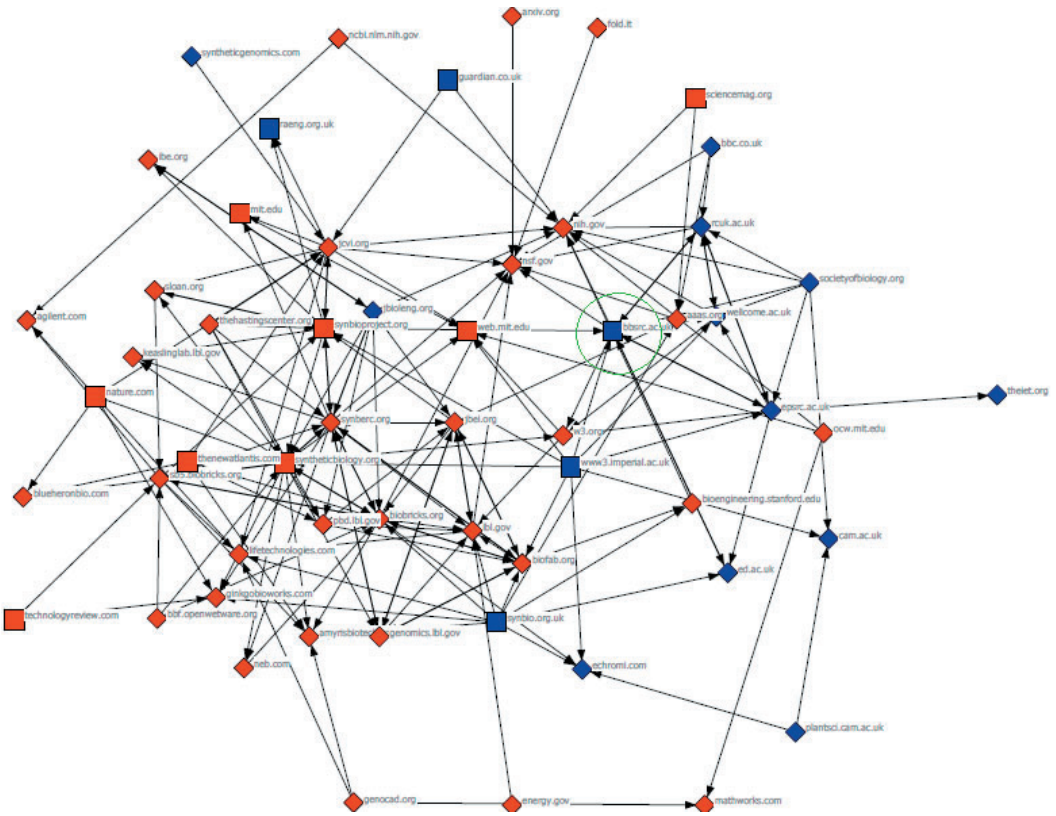

Figure 9: The web vision resulting from querying the British version of Google for "synthetic biology" in February 2012

October 2011. The reason the site receives so many inlinks in these months is that members of both the British and the American cluster mention a specific report on the social and ethical challenges of synthetic biology that BBSRC published in 2008. For instance, the American website '2020 Science' links to a specific section of the report that takes a stance on the issue of 'garage biology', which it perceives as relevant to the American debate. Accordingly, the network visualizations in Figures 6-9 make it possible to detect a correlation between an American interest in a specific report of BBSRC and the visibility of the British cluster to the British Google users.

Further inquiry into the geographic context of links in the web-visions from February 2011 to February 2012 indicates a second – but related – reason for the Americanization in those specific months: the launch of a specific report on synthetic biology that Barack Obama requested in 2010 (The Presidential Commission for the Study of Bioethical Issues, 2010). The influence of this report on the geographical composition of the web-visions can be illustrated by attending to specific differences between the web-vision in April 2011 (this visualization is also not printed because of space limits) and the web-vision in June 2011 (Figure 7). This is a period during which the Americanization of the web-visions are on the

rise and a closer look at the figures illustrates that this rise is correlated with a change in the geographical status of the websites in the web-visions that fall under the category of 'public engagement, social science or ethics'.

If we take a look back at Figure 2, we can see that the percentage of this type of website stays stable, but this quantitative similarity covers an increasing institutionalization and Americanization of the websites that gain visibility within this category. Daisy Ginsberg and James King, who are both UK-based designers who work on illustrating social issues around synthetic biology, disappear from the web-vision between April 2011 and June 2011. Throughout this period, their visibility is substituted by the visibility of North American NGOs such as The Hastings Center and The ETC Group. A manual analysis of the context of this linking behavior indicates that the change can be explained by the fact that central websites began linking to the Presidential Commission report in April 2011. Accordingly, a shift has taken place so that attention is guided towards institutions that are closely involved in the making of the report. The detection of this change is another example indicating that the visibility of British elements in the web-visions is fragile in the face of American events such as the launch of the Obama report.

These findings, of course, do not mean that event-driven linking behavior on the part of specific American websites can be isolated as *the* selection mechanism that influences the changing geographical composition of the web-vision of the British Google user. However, the perceived relevance of the BBSRC report and its focus on the issue of garage biology seems to be a necessary condition for the visibility of a broader British cluster. Similarly, it seems that the institutions mentioned as working with societal issues in the Obama report gain visibility at the expense of British websites. These findings are interesting because the changing geographical composition of the web-visions occurs in a period of time in which there is no change in the geographical composition of the SERPs from which they are built. For instance, the ranking of the BBSRC is stable in the SERPs throughout the months during which its structural position in the web-visions is changing dramatically. While this is far from being a bulletproof causal connection, it is, at least, an indication that specific types of event-driven linking behavior have effects on the geographical composition of the web-vision but do not have any effect on the SERP. It is an argument for the relevance of working with a theoretical distinction between the two objects of analysis.

3rd empirical claim: Query semantics influences the composition of both SERPs and web-visions

The third empirical contribution of the paper builds on the second claim in the sense that it tests whether the detected differences are strong enough to be influential if the user alters the semantics of the search. More specifically, it was decided to construct a parallel set of SERPs and web-visions in which the original query term, "synthetic biology", was substituted with its Danish translation, "syntetisk biologi". By holding constant every other

variable except the national semantics of the query, it was possible to detect the influence of this specific semantic variable on the composition of the SERPs and the web-visions and thereby to compare the strength of its effects with that of the linking behavior discussed above.

The effect of changing this semantic variable is quite dramatic. Changing the query to Danish has the consequence that both the SERP and the web-vision become dominated by European websites throughout the whole period of data collection (these visualizations are not shown in this paper because of limited space). The SERP never contains more than one or two British or American websites. Furthermore, the European websites it makes visible are, to a large extent, involved in 'public engagement, social science or ethics', which was not a dominant category of websites in the web-visions resulting from the British searches. This tendency is mirrored in the web-visions resulting from the Danish search, which are also dominated by websites from the same European tradition of technology assessment. This tradition dominates despite the fact that the search was carried out from a computer in London, and this domination stands in contrast to the dominance of American natural scientists in the SERPs and the web-visions resulting from the British search.

A relevant finding to take way from this experimentation with search language is that web-visions based on Danish semantics are more stable than web-visions based on British searches. Whereas we have just seen how the latter shifts between being dominated by American natural scientists in some months and having a more balanced distribution in others, it is quite evident that the Danish web-visions are dominated by specific European actors throughout the whole period of data collection. From this finding, it is possible to conclude that national semantics is a selection mechanism that erases the differences between the SERP and the web-vision that were just identified as the outcome of event-driven linking behavior in the British searches.

This suggests that linking behavior only produces a difference between SERPs and web-visions under specific conditions in specific situations. The fact that the British share semantics with Americans simply makes their web-visions of synthetic biology more fragile to the linking behavior of specific American websites. Again, this is not bulletproof causal argument, but it illustrates the complex and distributed arrangement of mechanisms influencing the 'real world representations' that a Google user encounters. It suggests a need for situated studies of the influence of different mechanisms and their influence on the SERP and the web-vision, respectively. It illustrates that the SERP is only a good indicator of the broader web-vision in some situations (for instance, the situation in which a user conducts a Danish search in the British version of Google) but not in others (for instance, the situation in which a user conducts a British search in the aftermath of an influential American event). In sum, it provides one more reason for abandoning mono-causal theories about algorithmic selection and broadening empirical interest to other sites than the SERP when devising a framework through which to examine the dynamics of online visibility.

Theoretical, empirical and practical implications of the three empirical claims

One theoretical implication to draw from the analyses in this paper is that any framework that attempts to make sense of Google's representations of the world must think about selection as something that is happening beyond the company algorithm. In other words, it is different from a concept such as the filter bubble, which has an implicit mono-causal orientation. A useful theoretical framework must provide concepts that make it possible to understand online visibility as an outcome of a complex entanglement of human and non-human actors. The concept of web-visions is one suggestion for a concept that can accommodate the influence of algorithms as well as other selection mechanisms such as event-driven linking behavior and trends in national query languages. Accordingly, it will motivate different takes on the way 'big data' techniques in digital society are having an impact on the 'real world representations' that its citizens encounter.

The concept of web-visions suggests a move away from seeing online visibility as something that can be controlled by a single organization with a specific technology. The concept is also supported by other recent writings on this topic. For instance, Noortje Marres (Marres, 2012) has forcefully argued that the chain of skills involved in the creation of online visibility—from data collection to analysis and visualization—is inevitably distributed across a range of socio-technical actors such as online platforms, web users, meta-data providers, algorithms and professional analysts. This call for a focus on the redistribution of information processing represents a quite different theoretical move than a focus on the way the capacities for such processing are being displaced to a private company with a powerful algorithm.

The proposal to focus on distributed selection mechanisms also comes with a specific empirical implication, which is the need to work with a broader unit of analysis than the SERP when it comes to understanding online visibility. The SERP is the favored empirical operationalization of the filter bubble, but the empirical analyses above have indicated that a focus on the SERP is only telling half the story about the creation of online visibility in the case of synthetic biology. The fact that web-visions have their own dynamics makes them a necessary, supplementary unit of analysis. They cannot be reduced to 'enlarged mirrors' of the SERPs because they perform the world in a distinct way. They are the outcome of selection chains in which different mechanisms play the central role. Working with the concept of web-visions entails studying Google's performance of online visibility in a way that allows us to pinpoint differences between SERPs and visions rather than collapsing them into a single bubble. Such a reduction is problematic because important details are lost, and there is a risk of exaggerating the effects of algorithms on the 'real world representations', which ends up guiding debates among citizens in contemporary digitized societies.

A second theoretical implication to draw from the analyses above concerns the role that Google does – and should – play in representing the world for these citizens. The

concept of the filter bubble carries with it the normative implication that Google and its algorithm should be evaluated on the extent to which it succeeds in providing an unbiased sample of the web sphere connected to the issue searched for. The distributed character of web-visions, however, makes it difficult to point to a single mechanism as either good or bad in this regard. Instead of seeing the search interface as something that conceals the world through algorithmic programming, the analyses in this paper support seeing it as part of a distributed chain of selection mechanisms that makes the world visible in specific ways, depending on the characteristics of the elements enrolled in this chain.

In other words, web-visions are seen as heuristic devices that should not necessarily be evaluated on the basis of whether they provide a representative sample of a predefined web sphere. Rather than discussing them on the basis of such *a priori* ideals, it is possible to ground the discussion about their value in knowledge about the distributed selection chain that produces them. More specifically, the proposed framework opens a possibility for evaluating web-visions on the basis case-study logics rather than logics of representation (Flyvbjerg, 2006). For example, we now know that web-visions resulting from a Danish search using 'synthetic biology' are 'less likely' to be fluid in the face of big American than web-visions emerging from a British search. This knowledge equips us to interpret an Americanization of the former as a stronger indication of changes in the geo-location of the issue than an Americanization of the latter. This makes it possible to use the Danish web-visions as a 'least likely case' that is geared toward make interesting anomalies visible in relation to questions about the geography of the specific debate about synthetic biology.

The proposal to evaluate 'real world representations' on the basis of such internal benchmarks is different from evaluating them on the basis of external benchmarks connected to the idea of a web sphere that is to be fairly represented. It is suggested that we think about such representations as 'modes of seeing' with specific characteristics rather than 'modes of concealing' that can be traced back to a single mechanism.

Conclusion

The empirical analyses in this paper have provided a basis for making three tentative claims about the way the attention of British Google users was guided towards the issue of synthetic biology from February 2011 to February 2012. The first is that there is a difference between the way the SERP guides the attention of Google users and the way their attention is guided when they explore the broader web-vision to which the SERP serves as an entry point. The second is that this difference is partly a result of event-driven linking behavior on the part of specific American websites that can influence the composition of a web-vision but do not influence the SERP from which it is built. The third is that the effect of such linking behaviors can be undermined by stronger selection mechanisms such as the semantics of the query. This shows that the 'real world representations' emerging from a world of digitalized 'big data' should be seen as distributed and situated phenomena.

These empirical findings were used as a basis for drawing theoretical, empirical and practical implications for future attempts at developing a framework with which to understand and conceptualize the role that devices such as Google play in contemporary practices of explorative knowledge inquiry on the web. One theoretical implication is the need to go beyond seeing the algorithm as the main selection mechanism and, instead, broaden the analytical focus to the distributed chain of socio-technical mechanisms that, together, construct 'real world representations' in specific ways. An empirical implication of this theoretical stance is that there is a need to broaden the empirical focus beyond the SERP, which has been the main operationalization of the filter bubble. A second theoretical implication was an invitation to think about devices such as Google as providing specific 'modes of seeing' rather than devices that conceal something in a pre-given web sphere. This move opens the possibility for evaluating the outcome of such devices on the basis of case-study logics rather than logics of sampling and representation. It opens up different forms of interaction with information-processing devices such as gaming them to be 'least likely cases' to make a specific social dynamic visible in specific situations.

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Notes

1 This addition to the URL simply removed personalization issues related to, for instance, the history of the browser. This, however, does not have any effect on the influence of the location. See, for instance, http://www.briangardner.com/disable-google-personal-search/. The choice of comparing de-personalized versions of the SERP and the vision was made on the background of a pilot study proving that personalization has little effect on the searches returned from Google. While this finding is only indicative, it is consistent with other experiments that have recently shown personalization to have only minor effects on other types of queries as well (see, for instance, Weisberg, 2011).

2 The crawler can be found at https://www.issuecrawler.net/. It was originally produced to remediate existing methodological critiques of citation analysis, but it is here used in a slightly different way than originally intended. It is used to simulate the web-vision of a Google user. For more on this use, see Madsen, 2012.

3 The software can be found at www.analytictech.com/ucinet/

4 A note of clarification is needed here. The reference to 'new' sites in this sentence hints at the appearance of sites in a given month of data collection that were not visible in the previous months of data collection. This means that a 'new' site could, accordingly, be a site from 1996 and the concept of newness is, therefore, not related to the calendar. It is a reference to the degree of change from one point in time to another.

5 It should be noted that the decision to use the SERP as the starting point for building the vision could be used to raise the argument of circular reasoning. One reason for this concern might be that Google's PageRank uses the structure of hyperlinks to rank its results; and, therefore, it makes little sense to com-

pare the network visualizations of the web-visions back to the SERPs from which they are derived. This is a quite fundamental critique of the research design that needs a few comments.

First, there is no doubt that the construction of the SERP is the outcome of the algorithmic processing of hyperlinks and other digital traces. In fact, hyperlinks were the central trace in Brin & Page's (1998) original paper on the PageRank. The argument in this paper, therefore, is *not* that web-visions are influenced by linking patterns, whereas SERPs are not. The argument is, rather, that a user who follows the URLs from the SERP to explore an issue further will potentially be effected affected(?) by selection mechanisms that do not influence the SERP. It is the location of such mechanisms that illustrates why sticking to the SERP as the empirical operationalization of online visibility risks exaggerating the power of the algorithm. Second, the comparison between the SERP (operationalized as a ranked list of URLs) and the web-vision (operationalized as a hyperlink network) is warranted even though the latter is built with the former as the starting point. The influence of the SERP on the web-vision is a central part of the research design, and it is never assumed that they are independent. To the contrary, the web-vision is created by following the outlinks that a web user gets from the SERP. The web-vision thereby shows the possible browsing paths that a user can follow from this specific starting point. It is interesting that these paths will in some months provide the user with an information environment that mirrors the one he or she will get by reading the URLs in the SERP, while it will provide a quite different environment in other months. This finding is interesting *precisely because* the web-vision is built with the SERP as the starting point.

Bibliography

Bar-Ilan, J. (1999). Search engine results over time: A case study on search engine stability. *Cybermetrics*, *2(3)*, 1.

Battelle, J. (2006). *The Search - How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*. New York: Portfolio.

Borgatti, S.P., Everett, M.G., & Freeman, L.C. (2002). Ucinet for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.

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.

Ferrara, J. (2008). Search behavior patterns, Available at: http://boxesandarrows.com/search-behavior-patterns/.

Flyvbjerg, B. (2004). Five misunderstandings about case-study research. *Sosiologisk Tidsskrift*, *12(2)*, 117-42.

Gerhards, J., & Schaefer, M.S. (2010). Is the Internet a better public sphere? Comparing old and new media in the USA and Germany. *New Media & Society*, *12(1)*, 143-160.

Hindman, M., Tsioutsiouliklis, K., & Johnson, J.A. (2003, April). Googlearchy: How a few heavily-linked sites dominate politics on the web. In *Annual Meeting of the Midwest Political Science Association* (Vol. 4, pp. 1-33).

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

Koed Madsen, A. (2012). Web-visions as controversy-lenses. Interdisciplinary Science Reviews, 37(1), 51-68.

Mager, A. (2012). Algorithmic ideology: How capitalist society shapes search engines. *Information, Communication & Society*, *15*(5), 769-787.

Marres, N. (2012). The redistribution of methods: On intervention in digital social research, broadly conceived. *The Sociological Review*, *60(S1)*, 139-165.

Pariser, E. (2011). *The Filter Bubble – What the Internet Is Hiding From You*. London: Penguin Press.

Rogers, R. (2013). *Digital Methods*. Cambridge, MA: MIT Press.

Schneider, S.M., & Foot, K.A. (2005). Web sphere analysis: An approach to studying online action. In C. Hine (Ed.), *Virtual Methods - Issues in Social Research on the Internet*. Oxford: Berg Publishers, 157-171.

Sunstein, C.R. (2006). *Infotopia - How Many Minds Produce Knowledge.* New York: Oxford University Press.

The Presidential Commission for the Study of Bioethical Issues (2010). *New Directions: The Ethics of Synthetic Biology and Emerging Technologies.* Washington D.C.: The Presidential Commission for the Study of Bioethical Issues in the USA.

Vaidhyanathan, S. (2011). *The Googlization of Everything (and Why We Should Worry)*. USA: University of California Press.

Van Os, R., Jankowski, N.W., & Vergeer, M. (2007). Political communication about Europe on the Internet during the 2004 European Parliament election campaign in nine EU member states. *European Societies*, *9(5)*, 755-775.

Vaughan, L., & Thelwall, M. (2004). Search engine coverage bias: Evidence and possible causes. *Information Processing & Management*, *40(4)*, 693-707.

Anders Koed Madsen
Assistant professor, PhD
Department of Learning and Philosophy
Aalborg University
akma@learning.aau.dk