



## UvA-DARE (Digital Academic Repository)

### Search engine freedom: on the implications of the right to freedom of expression for the legal governance of Web search engines

van Hoboken, J.V.J.

**Publication date**  
2012

[Link to publication](#)

#### **Citation for published version (APA):**

van Hoboken, J. V. J. (2012). *Search engine freedom: on the implications of the right to freedom of expression for the legal governance of Web search engines*. [Thesis, fully internal, Universiteit van Amsterdam].

#### **General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

#### **Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

## **Chapter 10: Search engine quality: pluralism, bias, transparency and user surveillance**

## 10.1 Introduction

It is generally agreed that the introduction of Google in the end of the 1990s had a significant positive impact on search engine quality. The leap in quality was mostly attributed to its better selection and ranking algorithms, discussed in Chapter 3.<sup>853</sup> Apart from its successful implementation of the PageRank algorithm,<sup>854</sup> which improved search result rankings significantly, Google's success from the perspective of search engine quality was also related to its extremely user-friendly and clean search interface. Google was not caught up in distracting portal integration strategies and in the first years it did not incorporate any kind of advertising, let alone the types of distracting banner advertising that were commonly found on existing search engines at that time.

But in general, what does it mean to say that a search engine is of high quality? The scientific literature on the subject has addressed a variety of subjects in this context. Decades ago, the information retrieval literature developed the standard quality measure for information retrieval technical systems in terms of recall and precision and such measures as well as the actual ranking models have been steadily improved.<sup>855</sup> Information scientists developed sophisticated theories about user-computer interaction and expectation of users which have helped the industry to adapt their services to the users' needs. From the perspective of commercial search engine providers, search engine quality will most often be defined as end-user satisfaction, which can be used to substantiate the often used more technological term 'relevance'.<sup>856</sup> Notably, in both cases search engine quality is defined from the perspective of end-users and not from the perspective of information providers or advertisers.

Broadly speaking, in this chapter the question about search engine quality will be seen as the question about the way in which fundamental notions relating to the quality of the public information environment are affected by the way in which search engines value, select, rank and prioritize information and communication streams. Obviously, the notion of quality will depend on the perceived purposes and role of search engines in general: quality will have to be seen in relation to that role and those purposes. Instead of discussing the notions of search engine quality from the field of information retrieval,<sup>857</sup> which is primarily focused on developing scientific models and tools to conceptualize and build better search engines, this chapter will focus on search engine quality from a regulatory and policy perspective. As search engines have become essential for the fulfillment of fundamental societal demands of general public interest, such as the free flow of and access to information and ideas, or consumer freedom in the market (not the specific focus of this study), it becomes essential to discuss how well the current offering of search engines is actually fulfilling these demands.

This chapter will build on the conclusions about the role of search engines in the public networked information environment, conceptualized in more depth in the preceding chapters, and will focus on a number of selected regulatory notions and issues that implicate search engine quality in particular. The

---

<sup>853</sup> See Section 2.2.2.

<sup>854</sup> Brin & Page 1998. See also Langville & Meyer 2006.

<sup>855</sup> See Baeza-Yates & Ribeiro-Neto 2011.

<sup>856</sup> See Van Couvering 2010, Chapter 6.

<sup>857</sup> See e.g. Lewandowski 2007.

discussion will be limited to a discussion of three main regulatory issues relating to search engine quality and freedom of expression. First, the question will be addressed whether and how the regulatory notions of diversity and pluralism in the public information environment could be used as normative signposts in the context of search engine governance. Diversity and pluralism are amongst the most fundamental normative starting points for the regulation of the public information environment, which leads to the question how these starting points can inform the regulatory debate in the context of search. To what extent do search engines impact on diversity and pluralism of the public networked information environment? On the basis of existing evidence, what would be the main concerns if addressing search engines from this perspective and what is needed if legislators or regulatory agencies were to move forward from this perspective, which is after all a fundamental concern for them on the basis of Article 10 ECHR?

Second, the regulatory debate related to the lack of transparency of the selection and ranking of search results, the possibility of various forms of (possibly hidden) bias, and more concretely the legal and policy issues relating to the advertisement business model of search engines and its impact on the quality of organic search results will be discussed. The question will be addressed to what extent the advertisement based business model entailing the sales and production of sponsored search results could entail incentives on search engine providers that would negatively impact on search engine quality. More specifically, the way in which the law and specific regulatory agencies have responded to concerns about search engine monetization through advertising with labeling obligations will be addressed as well as the merits of this response in view of the quality of search engines as mediators of the public information environment.

The chapter will conclude with an analysis of the legal issues relating to user data processing and its impact on the selection and ranking of search results from the perspective of the right to freedom of expression. The amount and nature of the user data processed by search engine providers, the increased personalization of search engine output as well as the lack of transparency about these practices pose some difficult regulatory problems. In the final section, the question will be addressed how data protection regulation and the right to privacy can be seen as a prerequisite for intellectual freedom of search engines users as well as the way in which the current legal regime incorporates this concern. Moreover, end-users may be served by extensive user data processing and profiling, since it would lead to higher search quality through ever better tailored search results or advertisements. At the same time, for users it may be hard to assess the actual trade-off that is taking place in the absence of proper information about the end-user modeling and user data processing that is taking place and the impact it has on search engine output. Consequently, the question will be addressed whether the way search engines impact on information consumption as a result of user data processing and personalization leads to undesirable outcomes from the perspective of the ideals underlying the right to freedom of expression more generally. These issues will be addressed in the final part of this chapter. In particular, it will explore the way in which data protection law, through its various rights and obligations relating to the accountability for the processing of personal data could contribute to the establishment of informational autonomy of end-users in their relation with search engine providers.

## 10.2 Search engine quality: pluralism, diversity and bias

### 10.2.1 The impact of search engines on pluralism and diversity: a short overview

Considering the opportunities offered by current networked communications technologies and services, to take issue with diversity and pluralism in the networked information environment may need some clarification.<sup>858</sup> The World Wide Web and networked communications more generally have both diminished and altered the power of traditional media and knowledge institutions to control the public information environment significantly. Because of the World Wide Web and shifting societal information practices, citizens no longer need traditional mass media to be able to make their ideas available for a broad audience. And because of the Web and search engines in particular, it may no longer matter a great deal whether or not a library selects a publication for their collection.

This much celebrated ‘disintermediation’ or even ‘democratizing’ effect of networked communications, however, typically stressed the potential of the new environment in comparison with some of the drawbacks of the old environment in terms of control over public information flows. This effect was a blessing from the perspective of pluralism and diversity, at least if these notions were understood to relate to entry barriers to information access, production and distribution. It is in this vein that a recent Communication of the European Commission on ‘A Digital Agenda for Europe’ notes that:

*“The Internet is also a driver of greater pluralism in the media, giving both access to a wider range of sources and points of view as well as the means for individuals – who might otherwise be denied the opportunity – to express themselves fully and openly.”<sup>859</sup>*

But as the networked information environment matures, it has started to produce new bottlenecks, mediating institutions and apparent biases. On the one hand, the ease with which information and ideas can now be published may have solved some of the issues relating to pluralism and diversity in the context of content production and dissemination. On the other hand, the abundance of information and ideas online implies sharply increased competition for audiences, leading to what is often denoted as an ‘attention economy’.<sup>860</sup>

In particular, the findability of information and ideas, as determined by the economical, institutional and increasingly complex technological infrastructure for the opening up of the public networked information environment, has sharply increased in importance. In the networked information society, it is simply not enough to publish one’s views to effectively participate in online debate. Winners and losers, from the perspective of effective dissemination of information and ideas, are partly determined by successful representation in search engines and related selection intermediary services. In particular, the impact of search engines and the algorithms they deploy for the prioritization of the publicity of certain sources of information over others have become an important issue in the debate about the effective dissemination of information and ideas, both from the perspective of information providers as well as end-users.

---

<sup>858</sup> On the meaning of pluralism and diversity, see section 10.2.2.

<sup>859</sup> European Commission 2010c.

<sup>860</sup> See Goldhaber 1997.

Central to the critique of the assumption of the positive impact of the Web on diversity and the equalizing impact on information access and production, is the following observation. In the public networked information environment, it is not decisive whether information and ideas are accessible or retrievable, but whether they are visible and likely to be encountered by Internet users.<sup>861</sup> This observation leads Hindman and others to specifically consider the impact of dominant search engine Google on pluralism and diversity. On the basis of an analysis of the structure of the Web, they claimed that

*“Rather than “democratizing” the dissemination of information, the prospect of googlearchy suggests that citizens may continue to get their political information from only a few sources, even on the apparently limitless information vistas of cyberspace.”*<sup>862</sup>

This led them to conclude that there is a need for a careful consideration of the impact of dominant search engines on the visibility and relative accessibility of information and ideas.

Some authors have addressed the actual impact of the selection and ranking mechanisms of various search engine services on the visibility of online information for end-users. An early example is the paper by Lawrence and Giles, who warned that *“not only are the engines indexing a biased sample of the web, but new search techniques are further biasing the accessibility of information on the web.”*<sup>863</sup> However, since it is arguably the role of search engines to prioritize certain sources of information over others, this argument does not persuade to the extent that it is based on the assumption that search engines could or should actually facilitate equal accessibility of sources of information in search engines.

The question is what normative standard should be used to evaluate search engine selection and ranking choices and their impact on the dissemination of and access to information and ideas. As Introna and Nissenbaum and others observed, dominant search engines may be expected to adopt ranking technology which emphasizes popularity over objective quality standards.<sup>864</sup> If taken to the extreme, this could lead to a situation in which information and ideas, worthy of effective representation and worthy of being found, could be doomed to relative obscurity. While the analysis of linking to pages already reflected an equation of relevance and popularity, the apparent focus of search engine providers on end-user satisfaction has made the label ‘popularity engines’ for dominant search services like Google even more appropriate.

And not surprisingly, considerable attention has been paid to the dominance of Google in the search engine market, and the possible influence this might have on access to information and ideas and the formation of opinion by end-users. Google’s market share - in terms of search engine users - is well above 90% in many European countries, including The Netherlands and Germany.<sup>865</sup> Although this might make some worried about the impact of one commercial enterprise on the dissemination of

---

<sup>861</sup> See Karppinen 2009, pp. 175-161. See also Hargittai 2000; Hargittai 2007a; Hindman et al 2003; Hindman 2009.

<sup>862</sup> Hindman et al 2003 p. 4

<sup>863</sup> See e.g. Lawrence & Giles 1999, pp. 107-109.

<sup>864</sup> Introna & Nissenbaum 2000 pp. 169-185. See also Finkelstein 2008, pp. 104-120; Marres & De Vries 2002.

<sup>865</sup> See Table 2.1, Section 2.2.1.

information and ideas online,<sup>866</sup> it is a largely an open question how the dominance of Google should be qualified in terms of the values of diversity and pluralism in the networked information environment.<sup>867</sup> This issue will be further addressed in the next section.

While a strict focus on end-user satisfaction may already favor information popularity over information quality,<sup>868</sup> the ways in which search engines monetize traffic may present additional issues from the perspective of search engine quality. In short, due to their monetization strategies commercial search engine providers can be expected to be at least as interested in providing value for their advertisers as in providing high quality search results for their end-users. End-user data processing and modeling in particular can be expected to be structured to provide value to potential advertisers, and the same can be expected of the optimization of search result pages, in particular the distribution and balance between organic and sponsored search results.<sup>869</sup> More generally, the commercial nature of dominant search engines might imply that existing services are more interested in optimizing the satisfaction of specific information needs of end-users, in particular the information needs of a commercial nature.<sup>870</sup> The amount and satisfaction of search engine users that are looking to buy a product or purchase a service will have a significant positive impact on the ability of search engines to make money, while at the same time having a possible negative impact on the satisfaction of other information needs.

Apart from the critiques of dominant commercial search engines in terms of their undoing of the promised equalizing potential of the Web, and the impact of the commercial business model on search result quality, one can also find more positive accounts of the impact of search engines on diversity.<sup>871</sup> One argument for search engines having a positive impact on diversity and pluralism would point to their impact on the visibility of information and ideas in otherwise heavily censored countries. In countries where certain minorities are not free to publish their views openly, or are not offered the same opportunities to speak openly in government controlled or heavily censored media outlets, the Internet in combination with search engine services has quite a dramatic impact on access to information and ideas that would otherwise have remained inaccessible. The facilitative role of search engines to access information and ideas about from or about such minority views led to fast responses by public authorities in countries like China, which pushes search engines to de-list certain websites, for instance relating to the Falung Gong movement. As was discussed in the last chapter, search engines have responded to certain government demands to filter out information but at the same time they have refused to cater too much to the demands of repressive regimes to filter out information.

---

<sup>866</sup> See e.g. European Commission 2007a, p. 16.

<sup>867</sup> For an in-depth analysis of the power of Google as well as a discussion of the different concepts of power used in search engine scholarship, such as the gatekeeper concept, see Röhle 2010.

<sup>868</sup> While popularity can be a sign of quality, it is taken for granted here that information quality and popularity are different properties that do not necessarily align. For instance, a false news report may be quite popular.

<sup>869</sup> See e.g. Röhle 2006. See also Yang & Ghose, 2010, pp., 602-623; Muthukrishnan 2009, pp. 1–12.

<sup>870</sup> See Section 3.3.2, 3.3.3

<sup>871</sup> See Fortunato et al 2006. See also Lev-On 2008, pp. 135-149.

Another such positive account, by Yochai Benkler, concludes that Google's specific decentralized mechanism for assigning relevance actually promotes diversity.<sup>872</sup> Benkler takes Google's PageRank algorithm, which de facto allows all web publishers to express their views on the relevance of other sources of online information through the act of hyperlinking, as an example of the positive effects of peer-production in the networked information environment on individual autonomy.<sup>873</sup> He observed, with a small empirical overview of search results for the search [Barbie], that Google's rankings seemed to treat critical information and ideas relatively favorably and concludes that Google may positively impact information diversity for end-users. Benkler conceptualizes diversity as a measure of the range of different of information and ideas end-users are presented with when searching the Web. In particular, he sees diversity as a measure of the extent to which end-users will be capable of making a meaningful critical assessment of the value of the sources of information they retrieve.<sup>874</sup> Notably, this concept goes further than a measure of mere accessibility of information and takes into account the observation discussed earlier that what ultimately matters is what sources of information individuals are likely to encounter.

Benkler's claim about Google's rankings may be somewhat anecdotal and his empirical findings may be outdated. Most importantly, however, it points in the same direction as identified by Hindman, Nissenbaum or others: The need for a careful study of the impact of selection and ranking practices of (dominant) search engines as mediators between information and ideas and Internet users, and a debate about the extent to which search engines do and should contribute to overarching fundamental societal interests related to the dissemination of and access to information and ideas, pluralism and diversity in particular.

It is important to note that due to the interactive nature of search engines, search engine quality and access to information and ideas for Internet users more generally depends a great deal on end-users. Above, the observation was discussed that diversity and pluralism in the public networked information environment depend on findability and actual visibility of information and ideas for end-users. This points to the need for empirical as well as normative evaluations of the mechanisms behind search engine selection and ranking mechanisms. At the same time, there is an equal need to understand actual end-user behavior. Clearly, the findability of information and ideas in search engines greatly depends on end-users' behavior and sophistication. The importance of end-user skills and education is positive for those that have them and negative for those do not. An experienced end-user remains quite powerful and in control of the search process, even when a search service prioritizes certain sources of information over others in ways that do not directly serve the interests of end-users. This means on the one hand that end-user skills and education are crucial for the overall quality of the search process. On the other hand, this means that some end-users do not have the appropriate skills to navigate the Web with the use of search engines effectively.<sup>875</sup> Some research has already been done about the use of search engines and access to information and ideas by different groups in society. Of particular interest

---

<sup>872</sup> Benkler 2006, pp. 285-294.

<sup>873</sup> On the conditions for quality to emerge in peer produced information environments, see Duguid 2006.

<sup>874</sup> Benkler 2006, pp. 285-294. For more theoretical background, see Benkler 2001.

<sup>875</sup> See Section 8.6.2.

in this context is the work of Eszter Hargittai. In a series of articles, she presents empirical evidence of search engine use, warning against assumptions about the positive state of access to information and ideas online, in particular for the less well-off in the networked information society.<sup>876</sup>

A final critique of the way in which search engines impact on access to information and ideas is focused on their increased personalization of search results and the negative impact this may have on pluralism and diversity. Slowly but steadily, dominant search engines have moved towards increasingly personalized search results, both for organic as well as for sponsored listings. This personalization of the search process - and information services online more generally - has led to another critique relating to pluralism and diversity in the networked information environment, namely the related critiques of the 'daily me' and the 'filter bubble'. The 'daily me' refers to the idea of personalized media, made possible by digital information technologies which better and more directly reflect the actual interests of the Internet user than traditional media outlets.<sup>877</sup> Such personalization of media, Cass Sunstein warns, could lead to so-called information cocoons. Internet users would only access information and ideas they are already interested in and would not be confronted with other topics or perspectives. This could, amongst other effects, damage an important prerequisite for democratic debate: a shared platform to communicate about societal issues.<sup>878</sup> It could even lead to a situation in which extreme positions are amplified further, for instance, if personalization would respond to, and amplify anti-social, sexist or even racist character traits of end-users. Sunstein has consistently called attention to these possible drawbacks of a heavily personalized information environment and argued for the value of independent selection by shared societal institutions such as the press and libraries.<sup>879</sup> It must be noted that neither Sunstein, nor Eli Pariser, whose more recent warning for 'filter bubbles' is discussed below, provide much empirical backing for their claims that existing personalization developments are actually having the effects on democracy they warn for. As a result, their arguments can be best understood as a contribution to the discussion about the normative ideals that should inform selection practices by mediating institutions in the public information environment.

Sunstein's warning against the downsides of personalization has been expanded by others, such as Eli Pariser's in his critique of the 'filter bubble'.<sup>880</sup> Pariser adds an element to the discussion by specifically addressing the way in which a range of Web based services have started to personalize their services based on various types of end-user data processing and modeling which are not transparent to users.<sup>881</sup> He finds this is the case for selection intermediaries, search engines, and social networks in particular, which are precisely the services which have a significant impact on actual information consumption. Pariser concludes that end-users have come to experience the effects of a 'daily me' type of personalization in a range of online services. In addition, he notes that end-users may not be in control or even aware of the type of personalization that is taking place. This is despite the fact that such

---

<sup>876</sup> See e.g. Hargittai 2007b. See also Halavais 2009.

<sup>877</sup> The possibility and benefits of personalized media were explored by Bill Gates in the 1990s. See Gates 1995.

<sup>878</sup> See also Karppinen 2009, p. 161-165.

<sup>879</sup> See. Sunstein 2002, pp. 44-50. See also Sunstein 2007.

<sup>880</sup> See Pariser 2011. See also Morozov 2010; Pasquale 2010; Zuckerman 2008.

<sup>881</sup> See also Section 10.3 and 10.4.

personalized selecting, filtering and re-ranking has started to have a major impact on the process of finding different sources of information and ideas online.

The short overview of the various arguments related to the impact of search engine ranking and selection practices on pluralism and diversity clarifies a number of things. Many of the warnings of search engine bias may be better understood as a critique of overly positive assumptions about the democratizing effect of public networked communications. These assumptions may indeed be unjustified or overstated when looking at the actual exposure to information and ideas online. At the same time, however, it is questionable whether the normative framework for search engine quality should include the standard that search engines should facilitate access to information and ideas 'equally'. From the perspective of information quality and the role of search engines in the information environment, search engines should prioritize certain information and ideas over others. At the same time, as was argued at more length in Chapter 8, search engines should be allowed to make their selection and ranking decisions in relative freedom, due to the editorial nature of the ranking and selection aspects of the search medium, which should be protected by the right to freedom of expression of search engine providers.

While it is important from the perspective of the right to freedom of expression to recognize that certain forms of search engine bias may be legitimate and not the proper subject of legal or regulatory interference, it is equally important to arrive at a proper understanding of the way in which search engines should be evaluated from the perspective of pluralism and diversity. The discussions that could eventually lead to such a framework are in still in their infancy. The next section will discuss some of the initial regulatory developments and explore the possibility of getting to such a framework for the assessment of pluralism and diversity in the context of search engines. In particular it will take into consideration the positive obligation on the State to guarantee pluralism under Article 10 ECHR, on the one hand, and freedom of expression as a negative right protecting against government interference on the other hand.

### **10.2.2 Pluralism and diversity in the context of search: legal and regulatory background**

The related legal and regulatory concepts of pluralism and diversity in the context of the media are broad and contested. If one follows a recent independent EU study on media pluralism, which also focuses on new types of media and dissemination of content in general, pluralism can be understood to implicate

*“the diversity of media supply, use and distribution, in relation to 1) ownership and control, 2) media types and genres, 3) political viewpoints, 4) cultural expressions and 5) local and regional interests.”<sup>882</sup>*

As this specific study also notes, pluralism and diversity are contested concepts. It points out that there are generally two underlying but conflicting views on pluralism and diversity. One view focuses on freedom of choice in a marketplace of information and ideas. The other view places more emphasis on

---

<sup>882</sup> See Valcke et al 2009.

diversity in a shared platform for public discourse, focusing on the public interest that society is served with various political views and cultural values.<sup>883</sup> These two underlying views of pluralism and diversity also inform and reflect different regulatory approaches towards the media.<sup>884</sup> In the former view, which is dominant in the United States, pluralism is a measure of free access and choice in the market, and would encourage media policies that focus on facilitating competition. In the latter view the public interest with regard to pluralism and diversity suggests more active involvement with and regulation of the media environment to promote access of a broad range of information and ideas.

While the concept of media pluralism is strongly related to concerns over media concentration, media concentration is not the only concern. Pluralism and diversity can be internal and external in nature. From an internal point of view, the demand of pluralism would require that a wide range of social, political and cultural values, opinions, information and interests can find their expression within one media organization. From an external point of view, pluralism would require that this process is facilitated through a number of media organizations, each expressing a particular point of view.<sup>885</sup>

Notably, the legal and regulatory dimension of media pluralism is linked to the positive obligation of the State to promote it under Article 10 ECHR. As was noted in Chapter 5, in the second half of the 20th Century it became accepted that the state does have a role in preventing too much concentration in the press and media in general. In the interest of pluralism mass media, including the print media, are usually not only subject to general competition law, as any other commercial undertaking, but also to special media concentration and cross-ownership rules and policies.<sup>886</sup> Concentration of media outlets in the hands of a few would undermine pluralism, of which, according to the European Court of Human Rights, the State is the ultimate guarantor.<sup>887</sup> Therefore, a press and media policy aimed at preserving the conditions necessary for a pluralist media environment is a reflection of the right to freedom of expression and not an interference with the media's constitutionally protected freedom. In the broadcasting context, in which pluralism has had particularly strong regulatory history, special measures to promote pluralism in this traditionally heavily regulated environment were considered appropriate. These measures include broadcasting licensing criteria, must-carry obligations and rules about the types of programming that should be made available to the public.<sup>888</sup>

In the European context pluralism is, in other words, a regulatory concept which informs, explains and legitimizes the role of government with regard to the media. Because of the broadness of both concepts, pluralism and freedom of expression might be seen as different sides of the same coin. In this European context, in which active media policies to promote pluralism still play an important role pluralism also reflects the idea that government has an active role to play in guaranteeing certain qualitative aspects of the public information environment related to the diversity of viewpoints finding their expression through the media. As was noted in previous chapters, the State's obligation to

---

<sup>883</sup> See Valcke et al 2009, pp. 5-6.

<sup>884</sup> See also discussion of regulatory frameworks for public libraries, Section 7.3.1 and 7.3.2.

<sup>885</sup> See e.g. European Commission 1992.

<sup>886</sup> See Section 5.2.

<sup>887</sup> See Section 5.2.

<sup>888</sup> See Valcke 2004.

promote pluralism under Article 10 ECHR is too broad and vague to be used to effectuate specific rights and or obligations in specific contexts that may require government involvement to promote pluralism.<sup>889</sup> Member States have considerable leeway to make choices in the way in which they fulfill this obligation to promote pluralism.

The Council of Europe has played an important role in shaping the contours of the legal and regulatory concept of pluralism and the way in which Member States could or should fulfill their positive obligation to promote it. Apart from the case law of the ECtHR, several Recommendations of the Council of Europe Committee of Ministers pay attention to pluralism or are specifically dedicated to it, in particular focusing on print and broadcasting media.<sup>890</sup> In its most recent recommendations on media pluralism, the Committee of Ministers has started to develop the notion of media pluralism in relation to the digital networked information environment. The recommendation on freedom of expression and Internet filters contains safeguards in relation to the undue interference with the free dissemination of and access to information and ideas over the Internet.<sup>891</sup> The Committee of Ministers' *Recommendation (2007)2* states that Member States should guarantee fair access to electronic communication networks for content providers.<sup>892</sup> Moreover, the same recommendation on media pluralism contains a broad reference to promoting pluralism and diversity in the context of new media. It states that:

*“Member States should encourage the development of other media capable of making a contribution to pluralism and diversity and providing a space for dialogue. These media could, for example, take the form of community, local, minority or social media. The content of such media can be created mainly, but not exclusively, by and for certain groups in society, can provide a response to their specific needs or demands, and can serve as a factor of social cohesion and integration. The means of distribution, which may include digital technologies, should be adapted to the habits and needs of the public for whom these media are intended.”<sup>893</sup>*

The Council of Europe may issue additional recommendations on pluralism and diversity in the context of the public networked information environment, considering the emergence of a range of new types of services, search engines and other selection intermediaries in particular, which impact on the actual dissemination of and access to the variety of information and ideas online encountered by Internet users. At this point the Council of Europe does not offer any specific guidance on the issue of search

---

<sup>889</sup> See Section 4.4.1 and 5.5.1.

<sup>890</sup> See CoE, Recommendation (2007)2 of the Committee of Ministers to Member States on media pluralism and diversity of media content, 31 January 2007; See also CoE, Recommendation (99) 1 of the Committee of Ministers to Member States on measures to promote media Pluralism, 19 January 1999.

<sup>891</sup> For a discussion, see section 6.5.4.

<sup>892</sup> CoE, Recommendation (2007)2 of the Committee of Ministers to Member States on media pluralism and diversity of media content, 31 January 2007.

<sup>893</sup> CoE, Recommendation (2007)2 of the Committee of Ministers to Member States on media pluralism and diversity of media content, 31 January 2007.

engines from the perspective of pluralism and diversity. However, future recommendations may specifically address the role of search engines from this perspective.<sup>894</sup>

In the context of the European Union, it was already noted in Chapter 5 that pluralism is one of the single subject matters related to the press in which European institutions have taken an interest. Following the developments of convergence and the digital transition of (news) publishing, this interest has extended to pluralism in the electronic media environment. This interest has translated into a 2007 European Commission working paper and the independent report on media pluralism indicators mentioned above. Both the European Commission's working paper as well as the independent study note the apparent relevance of search engines when analyzing media pluralism and diversity.<sup>895</sup>

The EC's working paper shortly discusses and dismisses some warnings that search engines could be a problem from the perspective of pluralism. It does mention a concern that search engines could be manipulating users to visit specific sources of information, in particular those of their advertisers. This would obstruct users from navigating the Web freely. In reply to these concerns, the working paper asserts, without references, that *"there are commercial incentives for offering an objective search facility. The business model for search engines is based on the provision of clearly separated advertising links, derived from contextual analysis of a particular search."*<sup>896</sup> Second, the working paper expresses its optimism with regard to the availability of and competition between different search engines from the perspective of pluralism on the Internet and notes that *"there are no fundamental technical limitations on the number of search engines that the Internet could support."*<sup>897</sup> Moreover, the paper touches on the aspect of competition between various selection and ranking systems, as applied by different search engines, when it notes, optimistically, that *"there will be new search algorithms in [the] future, able to search multimedia content, a topic supported by EU Research projects. Through sophisticated indexing, cataloguing and filtering algorithms, these enable users to access wider news content than they would otherwise have access to."*<sup>898</sup>

The independent EU study on pluralism indicators does recognize that new media of all types play a role from the perspective of pluralism. With reference to a discussion of technological and economic developments in the context of networked electronic communications, including the emergence of search engines, it notes that traditional indicators of pluralism may have become obsolete.<sup>899</sup> At the same time, it points out that these developments may open up new avenues to ensure or promote pluralism. With regard to search engines, specific reference is made to the importance of user education about the functioning of search engines, the question of the existence of regulatory safeguards for the editorial independence of search engines, the possibility of including indicators on search engine

---

<sup>894</sup> Since 2010, the Council of Europe's Committee of Experts on New Media (MC-NM) has been working on a Draft Recommendation on the protection of human rights with regard to search engines. The draft recommendation was opened for public consultation in early 2011.

<sup>895</sup> See European Commission 2007a, pp. 16-17; Valcke et al 2009.

<sup>896</sup> European Commission 2007a.

<sup>897</sup> European Commission 2007a.

<sup>898</sup> European Commission 2007a.

<sup>899</sup> Valcke et al 2009.

ownership and the role of search engines in providing access to local content.<sup>900</sup> The extent to which media pluralism policies should lead to an in-depth consideration of the specific impact of search engines on pluralism, and in what ways, remains somewhat unclear however.

It goes beyond the scope of this research to develop an actual framework for the assessment of pluralism and diversity in the context of search engines. In view of the various early perspectives on pluralism, diversity and search engines, as well as the early regulatory responses discussed above, it is however possible to discern at least two dominant issues relating to pluralism and diversity, which deserve special attention. These are the question about the impact of the ownership structure of dominant search engines and the current market consolidation on the one hand, and the question about the diversity of references end-users are exposed to when using specific search engines on the other hand.

First, the dominance of Google in the search engine market quite intuitively raises the question about the impact of the search engine market structure on pluralism and diversity. The question of to what extent there is actual evidence of a negative impact will be evaluated below as well as the rather difficult question about the way in which concentration in the search engine market should be scrutinized from the perspective of pluralism and diversity. Second, there seems to be general agreement amongst those that have considered search engines from the perspective of pluralism that an evaluation of pluralism in the networked information environment requires an assessment of the relative accessibility of content. This agreement points to a central focus on the services which affect the visibility and findability of content and actual search engine user behavior. Although pluralism remains an issue in the context of the production of information and ideas, pluralism in the networked information environment, characterized by abundance, cannot be separated from the question whether or not it will become accessible and visible for Internet end-users.

### **10.2.3 Market concentration**

The market for general search is dominated by a small number of firms.<sup>901</sup> If we look at the European and American markets, the large majority of search requests is currently handled by two firms, Google and Microsoft. Yahoo!, once market leader with its famous directory, remains a competitor in the field of search traffic monetization, but it no longer produces its own search results but uses those of Bing instead.

There are a variety of reasons that can be given for this market consolidation, a full discussion of which goes beyond the scope of this research. Amongst the more plausible ones are (1) the current scale of the World Wide Web and the amount of users (billions of online destinations, billions of users and queries), (2) the knowledge and human capital that is needed to design and operate a general purpose Web search engine of competitive quality (dominant firms guard such knowledge as trade secrets and heavily compete for scarce engineering talent), (3) the existence of patents in search technology and intellectual property rights restricting competition, (4) the general satisfaction of users with the current

---

<sup>900</sup> Valcke et al 2009.

<sup>901</sup> For a discussion of market developments and figures on market share, see Section 2.2.3 and 2.2.4.

supply (market leader Google may simply offer the highest quality search results), (5) the lack of price competition for search engine users and the subsidization of the distribution of organic and paid search results through third-party arrangements by dominant search providers (meta-search engines, portals, third party sites and browsers, mobile operating systems), (6) the need for an extensive user base and large amounts of user data to improve search engine selection and ranking and innovative features such as search suggest, (7) the dependencies in the two-sided market for advertisers and end-users (advertisers may choose the platform with the most users) (8) the effective integration of search services into and with a variety of other information services provided by the same company that provide a soft 'lock-in' for users.

Clearly, a highly concentrated market for general purpose search engines implies that the impact of such dominant services on the visibility of content for Internet users becomes stronger. The de-indexing of content by a dominant search service will have a particularly strong impact, since the removed source will not be visible at all for its users. These issues relating to access have been discussed in Chapter 9. Whereas de-indexing has a clear impact on the findability of content for search engine users, it is much harder to assess the impact of the selection and ranking mechanisms of particular search services and the related implications of market consolidation for pluralism and diversity.

Even though the dominance of Google may intuitively suggest otherwise, it is wrong to rush to the conclusion that Google's dominant position in terms of user market share is detrimental to pluralism and diversity. In fact, there seem to be a number of arguments which warrant skepticism regarding this claim. First of all, it is possible that Google respects pluralism and diversity much better than any of its competitors.<sup>902</sup> Second, it is also possible that existing commercial general purpose search engines are converging in terms of the visibility of different kinds of content, implying that it does not matter – from the perspective of pluralism and diversity – which of them users choose. Third, and maybe most importantly, search services are interactive. Their selection and ranking algorithms do not carve the findability of sources of information in stone. Their output depends heavily on user input and perceived user expectations. The ranking and selection of organic results are strongly influenced by the online environment as a whole and by webmasters in particular. This even raises the question, whether one general purpose search engine service for the Web could be sufficient, from the perspective of diversity and pluralism of the public networked information environment.

Interestingly, in its review of the agreement between Yahoo! and Microsoft in 2009, an agreement that lead to a reduction of different sources of organic search result listings, the European Commission explicitly considered the possible negative impact on users by a loss of variety.<sup>903</sup> The agreement meant that Yahoo would no longer produce its own index and search results but would use the organic search results of Microsoft's search service, Bing, instead. Although both companies had a very low market shares in the EU relative to Google and the combined market shares of Yahoo and Microsoft were well

---

<sup>902</sup>For example, Benkler argued that Google, in comparison with other search engines, had an actual positive impact on diversity. See Section 10.2.2. Although Benkler might be overly optimistic, the reverse claim might be overly pessimistic.

<sup>903</sup> European Commission, Case No. Comp/M.5727 – Microsoft/Yahoo! Search Business, 18 February 2010, §202.

below the safe harbor in the horizontal merger guidelines,<sup>904</sup> the possible impact on search quality, however, seems to have been the reason the European Commission investigated further:

*“given that competition in this market takes place in terms of quality and innovation, and the entry barriers are high, the Commission undertook a thorough analysis of the possible effects of the transaction on the search market.”<sup>905</sup>*

The European Commission investigated whether the agreement would impact the incentive to innovate, whether lessened competition could incentivize to lower the quality of organic results (to increase click through rates for advertisements), and whether users would be harmed by a loss of variety. It concluded that none of these effects risked occurring. What is more, the European Commission concluded that *“it is possible that due to the transaction some benefits will materialise due to larger scale of the merged entity”*.<sup>906</sup>

Considering the high entry barriers, the current market structure is unlikely to change much in the near future. And with its market share of 90% or more for Internet searches which are being reported, Google is obviously important from the perspective of findability of information and ideas online. As mentioned above, this raises the question about the impact of Google on pluralism and diversity, and on the way in which pluralism can be supported internally. This question will be addressed in the next section.

However, it is important to note that Google still exists in a broader context of other information services and electronic publishers that have an impact on the relative accessibility of information and ideas for end-users. In the general purpose Web search market in the strict sense there is still competition in European and the United States, most notably from Microsoft. In some local markets, such as in the Czech Republic, Russia, and China there are strong local market players and leaders.<sup>907</sup> In addition, there are numerous vertical search engines which positively contribute to the findability of specific types of information. Furthermore, there are publicly funded search engine services, which may have little or negligible market share but can contribute to quality and diversity of the online search environment. There is a variety of other dominant information services, including Facebook and Twitter, which provides different types of selection and recommendation mechanisms for users to access content in the networked information environment. Through the use of hyperlinking, electronic publishing in general also contributes to the findability of information and ideas on the Web. And finally, traditional media formats such as print and broadcasting have to be taken into account as well.

To sum up, the intuitive claim that consolidation in the market for general purpose search engines is detrimental to pluralism and diversity would need more empirical backing to be taken seriously. This is especially the case from a historical perspective on access to publishing opportunities and access to information and ideas. Search engines must be seen in their proper context: they are important for the

---

<sup>904</sup> Id, §130.

<sup>905</sup> id, §118.

<sup>906</sup> Id, § 226.

<sup>907</sup> See Section 2.2.4.

relative accessibility of information and ideas but exist in a broader environment that includes a variety of services that also affect accessibility, such as publishers, directories, social networks and micro-blogging services, as well as offline media. Moreover, the search engines addressed in this study operate with respect to the World Wide Web. It is quite likely that the democratizing, equalizing and disintermediation effects of the Web for access to audiences on the one hand and access to information and ideas on the other hand have been enthusiastically overstated. Still it is arguably the most open and diverse platform for the dissemination of information and ideas in human history.<sup>908</sup>

#### **10.2.4 Diversity in the context of particular search services**

As was discussed in the last section, an important question is how to assess specific search engines, market leader Google in particular, from the perspective of pluralism and diversity. Google handles the large majority of global Internet search queries and close to all of them in many European countries. Like other issues related to search engine governance, the question of diversity of search results in the context of a specific search engine can be addressed from the perspective of all three stakeholders: the end-user, the information providers and the search engine itself.

From the information provider's perspective, the question of diversity is one about their level of representation in search results. Can they be found in a specific search engine at all, and if so, how easily? For end-users the question of diversity is first of all an issue of being able to find information through search engines. But more specifically and considering one of the conclusions in Section 10.2.1 there is the question of to what extent users are being confronted with a diverse set of search results, in terms of language, geographic location, source and viewpoint and other characteristics, and to what extent users have an impact on the answer to this question. As research has shown, and search engines have adapted to the reality, that most users will not look beyond the first page of search results, this question is of special relevance with regard to the first set of search results. From the search engine's perspective the question about pluralism and diversity may be best framed in terms of the set of values that ultimately determines the design and thus functioning of its service. In addition, one wonders to what extent search engines have incentives to promote or diminish diversity and whether they actually do.

As was discussed in more detail in the preceding chapters information providers depend on search engines to find a way to an audience, but have a range of options to optimize their exposure to users online. Seemingly, the simplest way to optimize one's ranking in search engines is to actually provide information that directly satisfies the information needs of users.<sup>909</sup> Wikipedia is an excellent example in this regard. For informational queries it provides an easily accessible first source of encyclopedic information for Internet users. In fact, this most likely explains the prominent ranking of Wikipedia articles in Google's search results.

However, it is naïve to think that the optimization of search results would be limited to the creation of useful content for users. In fact, maybe the biggest obstacle for information providers to gain a

---

<sup>908</sup> See e.g. Rogers 2004; Benkler 2006.

<sup>909</sup> See e.g. Cutts 2006.

competitive ranking is the effort and money it takes to actually optimize their representation in search results effectively. Search engine optimization (SEO) is a thriving industry and involves a variety of widely accepted and more controversial practices. Since SEO helps to establish winners and losers in search engine results, it is conceivable that search engine rankings will - to some extent - reflect the economic means of underlying information providers as well as their willingness to engage in less acceptable optimization practices. This state of affairs could be considered problematic from the perspective of pluralism and diversity. In addition, since not only sponsored results but also organic results are important platforms for commercial communications, the question about the impact of marketing on diversity of search results is of particular importance. The issue of the impact of advertising on search engine quality will be addressed in more detail in Section 10.3.

While it may already be considered problematic that search engine rankings could partly reflect publishers' financial means to optimize relevant search results, the possibility for information providers to optimize their rankings also leads to the spamming and manipulation of search engine results with wholly irrelevant sources of online information. There are various examples that show the feasibility of manipulating access to information for end-users through search engine optimization campaigns. Most famous are the so-called Google bombs that result in amusing or politically motivated search results for certain queries. The query [miserable failure] would for some time return the official website of former president G.W. Bush as a first result and the Dutch query [raar kaspel] (weird haircut) would return the website of former Dutch prime minister Balkenende.

The Google bomb technique was based on the use of the anchor texts of incoming hyperlinks to websites, meaning that a website would rank well for a particular search term if incoming hyperlinks included that term. Google has eventually found a specific algorithmic solution for these highly visible instances of external influence on search engine results.<sup>910</sup> Notably, the reasons that Google gave for ultimately fixing these Google bombs was that it wanted to correct the misperception that Google actually endorsed the opinion implied by the prominence of these search results, as well as that Google had hard-coded these specific results themselves into their search result pages.<sup>911</sup>

Google bombing, however, is only the tip of the iceberg and the algorithmic solution to defuse them does not prevent information providers from influencing search result selection and ranking more generally. Consider for instance the services that are offered to natural persons or businesses to push away unfavorable search results. As it may be quite hard in practice and legally difficult to have certain search results removed that negatively reflect on someone's reputation, a typical strategy pursued to minimize exposure to negative publicity is to optimize the ranking of more favorable results and author additional favorable content to fill up the first page of Google's search results. Google reputation management strategies include the strategic creation of websites, blogs and social network profiles in ways that have a high chance of ending up on the top of the list of search results for relevant queries. Remarkably, considering the possible negative impact of the overall quality of search results, Google itself has endorsed this strategy, giving the following advice on the company website:

---

<sup>910</sup> See Sullivan 2007.

<sup>911</sup> See Moulton & Carattini (Google) 2007.

*“If you can't get the content removed from the original site, you probably won't be able to completely remove it from Google's search results, either. Instead, you can try to reduce its visibility in the search results by proactively publishing useful, positive information about yourself or your business. If you can get stuff that you want people to see to outperform the stuff you don't want them to see, you'll be able to reduce the amount of harm that that negative or embarrassing content can do to your reputation.”<sup>912</sup>*

Hence, Google seems happy to grant information providers considerable power to influence rankings and seems unwilling to entertain the possibility that its algorithms will be unable to guarantee that negative or controversial information and ideas, which must be considered truthful or valuable from the perspective of end-users, could be effectively hidden from the public.

From the perspective of the ideals underlying the right to freedom of expression, the issue can be best framed as a question about the robustness of the search medium.<sup>913</sup> Clearly, the use of billions of third party signals and the lack of traditional editorial control, in terms of what enters their indexes and how well it ranks in search results, poses a risk to the ability of search engines to resist undue interference with their rankings. All major search engines are fighting these ongoing attempts to manipulate their search results, since they can easily deteriorate the value of their service for end-users. However, paradoxically, search engines also depend on these attempts to influence their rankings, which may partly explain Google's endorsement of the practice of reputation management discussed above. To a considerable extent the signals that are provided by information providers in the form of content, hyperlinks or otherwise, are necessary to make the service work in the first place.

Optimistically speaking, search engines like Google will be able to balance the different third party signals with each other and to develop sophisticated technology and organizational strategies to combat undue interference with their service. In practice, however, search engines' dependence on third party signals for the ranking of search results will predictably cause the service to produce some search results that no human editor would have selected as a relevant result for the query entered. And quite possibly, some third parties will be able to diminish the diversity of viewpoints that are present in the first set of search results, thereby negatively affecting exposure diversity for end-users. Apart from search engine optimization of organic results, another way this may happen is through the crowding out of organic results through advertising on search result pages. The optimization of organic results in combination with bidding on sponsored search results can effectively push other sources of information from the users view.<sup>914</sup>

For search engine providers, the first question to ask is how much they actually value diversity in the design of the service and how much it is a priority for them in comparison with other values such as

---

<sup>912</sup> See Moskwa (Google) 2009.

<sup>913</sup> In the sense of the word robust in the U.S. Supreme Court's conclusion in *Sullivan* that the First Amendment requires that public debate should be “*uninhibited, robust and wide-open*”. See *New York Times v. Sullivan*, 376 U.S. 254.

<sup>914</sup> A good example of these practices from the United States is the reputation management strategy of the company Lifestyle Lift to obscure legitimate consumer complaints in search results by creating and optimizing a variety of new websites themselves. For a discussion, see Levy 2008.

popularity or user satisfaction. Some of the relevant criteria with regard to diversity mentioned above, such as geographic location and language can be rather easily dealt with by existing technology. Hence, search engines could actively promote diversity by incorporating these aspects into their ranking algorithms. They could, for instance, aim to always include local results when available and more generally maximize the diversity of results in terms of their geographic location. In fact, Google currently offers its users the choice to specifically focus on local results and Google's offering of automatic translation tools have expanded the informational horizon of Internet users quite drastically.

For more content related aspects of diversity, such as particular viewpoints on different societal issues, it may be harder for search engines to actively promote diversity directly and they may for various reasons be reluctant and unwilling to do so. Still, it is possible to imagine that search engines would try to discern viewpoints on particular subjects in their indexes, for example with regard to the greenhouse effect, but doing so is much harder than detecting geographic location.<sup>915</sup> It is easier for search engines to prioritize diversity, directly or indirectly, with regard to certain types of sources of information such as official government websites that provide relevant information on the one hand, or new phenomena like Wikipedia and user generated content such as Internet discussion forums on the other hand. Predictably, these types of sources of information will provide search engine users with different possible perspectives, even though the search engine would not have selected those different viewpoints directly as traditional editorial media would have done.

There is one important aspect of operating a search engine that gives search engine providers room and possibly an incentive to promote diversity of search results. Because of the size of many user queries, which often consist of only a word or two, search engines are constantly making informed guesses about what their users are actually looking for. Consider, as a simplified example, the query [New York], The user could be interested in many things, including reading about New York's history or culture, obtaining the lyrics of a song by Frank Sinatra, or booking a flight or hotel room for a holiday or business trip. These different types of information are all available on the Web, and the search engine's ranking algorithms will have a direct or indirect impact on the types of information that it presents to its users. It makes sense for search engines to discern these different types of information in their index and adapt their search result selection and ranking practices accordingly. Yahoo!, for instance, used to offer a special user interface that allowed users to re-rank search results on a scale between commercial and research.<sup>916</sup> Google has recently included a similar choice for users in its new advanced search tools. In the absence of other information or knowledge of user preferences, it would make sense for search engine providers to try to include all these possibilities in its search results. In other words, to the extent that it finds the means to do so, a dominant search engine like Google is likely to continue to be relatively inclusive and try to satisfy minority interests, because it could lose market share to competitors and possible new verticals otherwise. Therefore, dominant general purpose search engine

---

<sup>915</sup> On possible tools for the analysis of online political dynamics with regard to specific issues online, see Rogers 2004.

<sup>916</sup> This tool to re-rank results was called Yahoo! Mindset and was released by Yahoo! Research Labs in 2005. It is no longer available. See Yamamoto et al 2007.

providers can be expected to try to optimize the diversity of their search result pages to cater to as many different information needs as possible.<sup>917</sup>

At the same time, a commercial search engine's focus on optimizing advertising revenues may entail incentives to focus on some specific user needs rather than others. In particular, search engine providers like Google or Microsoft may be strongly biased towards consumer aspects of the end-user, since their revenues directly relate to the quality of the search process for searchers that are shopping. A strong focus on these information needs may be necessary in their competition for end-users. Otherwise, they could lose their users to destinations which are completely specialized in shopping, such as certain vertical search engines or e-commerce services like Amazon. Interestingly, Microsoft's CEO Steve Ballmer made clear in 2008 that their efforts in Web search would be mostly focused on search with 'high consumer intent'.<sup>918</sup> Microsoft halted their efforts to develop a book search and academic search service, and instead it started to focus on integrated commercial verticals such as for travel.

Major search engines have also developed other strategies to cater to different types of users. First of all, major search providers offer country and language specific search engines. Secondly, they typically offer a range of search services for special types of material or information, including image and video search, geographic search, news search, academic search, book search and blog search. And finally, the general Web search service has evolved into 'universal search', integrating these different types of material into one search result page. While the inclusion of some of these types of specific search results might be driven by a desire to foreclose competition,<sup>919</sup> these strategies could also impact on diversity in the context of search, both favorably and unfavorably.

When addressing the question about diversity of search results, it is also interesting to look at the ways in which search engines have dealt with legal but indecent, offensive or controversial sources of information. Such search results could alienate certain search engine users from specific search engines. They can also lead to societal pressure on search engine providers to change their search engine algorithms. Therefore, dominant search engines may have external incentives to prioritize non-controversial information.

For certain types of controversial information, nudity and pornography in particular, the reaction of search engines has been to prevent accidental access through the application of filtering, which can be put on or off by the user and is typically on by default. The references remain accessible, but only for those users who actively indicate they are willing to find them by changing their default search settings. Generally, however, a search engine like Google seems to reflect the relative abundance of controversial information on the Web quite well in their search results. In general, the comprehensiveness of their

---

<sup>917</sup> Similarly, Van Couvering 2010.

<sup>918</sup> See Sullivan 2008.

<sup>919</sup> See Section 3.3.1.

indexes seems to carry more weight than the possible negative impact of confronting their users with sources of information they strongly disagree with or find offensive.<sup>920</sup>

Illustrative is the discussion about Google's treatment of the query [jew].<sup>921</sup> Predictably, various flagrantly anti-Semitic information providers, such as 'Jew Watch News', have used the Web as a platform to communicate. Search engines have generally not decided to remove those sites from their index, and, on top of that, sites like *jewwatch.com* even received very prominent rankings for the query [jew]. Initially Google responded to complaints about the prominence of anti-Semitic websites in its search results by arguing that these results were simply based on its neutral ranking algorithms, such as PageRank; the algorithms were applied without regard to the actual content. Eventually Google decided to add a specific hard-coded notice (as a search result and for google.com also in the form of a self-sponsored link) to the search results for this particular query in its search service, explaining the prominent ranking of the website *jewwatch.com*. This statement now explains that the relative high ranking of anti-Semitic websites for the query [jew] can partly be attributed to the fact that the word 'Jew' itself is typically used in a pejorative sense on anti-Semitic websites. Non-anti-Semitic websites tend not to use the word 'Jew' at all, but 'Jewish person' instead.<sup>922</sup> In short, Google's algorithms are argued to simply reflect both the availability of such information online and the typical meaning of the query [jew]. Hence, one could argue that Google prefers to confront its users with the fact that such websites exist than to hide the controversy from its users' sight.<sup>923</sup> Notably, this state of affairs does depend on the country specific service and regulatory context. For the Germany-specific search service google.de all results for the website *jewwatch.com* have simply been removed. The removal is comparable to the removal of the *stormfront.org* website for users of google.de, discussed in detail in the Chapter 9.

Returning to the need for search engines to match a typically weak signal of a user's information need with a very large set of possibly relevant search results, one of the most important ways in which search engines currently deal with the ambiguity of search queries is to analyze historic user behavior and evaluate and improve their results accordingly. The resulting user profiling and personalization combined with a strong focus on user satisfaction could be considered worrying from the perspective of pluralism and diversity in the online information environment. First, this combination could undo the incentive to confront users with as many relevant perspectives on a query as possible to increase the possibility of matching the user's information need. Second and related, it could lead to the biased information bubbles of which some have warned with detrimental effects on the possibility of an informed and shared debate about matters of public concern.<sup>924</sup>

---

<sup>920</sup> Some entrepreneurs have identified this as a market opportunity and have developed search services that aim to prevent certain information from showing up. For a discussion of some of these services and their quality, see Hopkins 2010.

<sup>921</sup> See also Section 8.5.1.

<sup>922</sup> See Google, 'An explanation of our search results', <http://www.google.com/explanation.html>.

<sup>923</sup> Notably, it could be argued that the fact that search engines are directing users to hate speech is, in a different and possibly more fundamental way, detrimental to pluralism. This question will not be further addressed.

<sup>924</sup> See Section 7.4.4 and Section 10.2.1.

However, the empirical evidence that personalization of search results does in fact cause such harm to our societies is not yet very convincing.<sup>925</sup> It is possible that personalization of search results makes search engines somewhat less diverse. Still, even in case of strong personalization search engines may keep confronting users with different stories or viewpoints on the same topic and search engines are hardly the only place where users encounter information and ideas online. In addition, search engines may also develop algorithms that satisfy the value of serendipitous encounters for end-users. Due to the interactive nature of search engines, much of the impact that personalization has will depend on the behavior of actual users. Related to the user's interest, the lack of transparency and control over personalization may be considered more problematic from the perspective of the informational autonomy of users than the mere fact that personalization is taking place.

It is clear that search quality is strongly related to the quality of interaction with search engine services, which to a considerable degree has to be shaped by the users themselves. The choice of the specific service, the knowledge of different languages, and the ability to formulate and reformulate effective search queries and use advanced search options has considerable impact on the quality of the search experience. A search engine's choices may have an impact on the ease with which a user finds a variety of information and ideas, but it will hardly help low-educated users who have no experience with search engine querying and know no alternative services to navigate the Web effectively. Hence, for end-users the quality of their search experience – for instance in terms of the diversity of views they are presented with - will depend on their knowledge and user sophistication. A dedicated, skilled search engine user will be able to reformulate queries to find different points of view and sources of information. Clearly, many users will not take the time or will lack the skill to do so.<sup>926</sup>

The considerations above may be enough reason to conclude that currently there is not enough evidence that specific search engines, Google in particular, pose significant risks from the perspective of information diversity and pluralism. Actually, there are a range of arguments pointing in the opposite direction, since Google does help end-users to access and navigate the abundant variety of information and ideas that are available online. However, considering the centrality of search media in the networked information environment, whether or not this is actually the case should be the subject of more thorough empirical research. The structural pressure on diversity by information providers which is inherent in the design of major search engines deserves special attention, as well as the ways in which end-users interact with search services and how the search experience depends on user knowledge, skills and education. The increased personalization of interaction between users and search engines does warrant special attention. While personalization may also have a positive impact on search engine quality, the way in which personalization is structured and the lack of transparency and control in this context is problematic from the perspective of end-users. The final section of this chapter will discuss the way in which data protection and privacy could be instrumental in dealing with this negative impact on informational autonomy of end-users.

---

<sup>925</sup> Currently, the most comprehensive case is made by Eli Pariser 2011. For positive impact of personalization on the problem of bias more generally, see Goldman 2006.

<sup>926</sup> See e.g. Hargittai 2007b.

Finally, it would help if search engines were more explicit about their underlying criteria for evaluating their search result pages. Does Google value pluralism and diversity, and if so, in what way?<sup>927</sup> Moreover, how can we hold a search engine like Google accountable for acting in line with their stated preferences about the ranking and selection of their search results? Arguably, the position of Google as market leader comes with an increased responsibility to clarify its views on this issue, which is from a broader perspective of fundamental importance for our society and the functioning of our democracies. This leads into the subject of the next section, namely the way in which more transparency about ranking and selection could enhance search engine quality and promote the interests of end-users.

### 10.3 Search engine quality, transparency and marketing

As was stated in the beginning of this chapter, search engine quality is addressed in this chapter as related to the way in which search engines value, select, rank and prioritize information and communication streams. From a legal and regulatory perspective, the relative opacity of industrial search engine ranking technology can be problematic and is typically seen as one of the central issues in the debate about search engine accountability, in particularly with respect to Google. What is actually at work when Google selects and ranks its search results? What goals does it pursue and which strategies does it deploy? What hidden biases, deliberate or accidental, impact on the relative accessibility of information and ideas for end-users and the possibility to reach an audience for information providers? Without some level of transparency about what is actually taking place, it is hard to arrive at any judgment about the way search engines rank and select search results.

Generally speaking, the actual selection and ranking practices of major search engines is best described as a grey box. Some may even call it a black box because search engines like Google do not disclose (all) the algorithms that are used, but this is too simplistic. Most providers do make various credible statements about the technology and software that is used to provide the search service. Google, for instance, explains that it now uses more than 200 different signals to determine the proper ranking of a search result and some of those signals are explicitly confirmed.<sup>928</sup> In addition, computer science provides general and specific knowledge about the ways in which general web search engines are or could be designed and operated.<sup>929</sup> Furthermore, part of the research and product development that goes on at industrial search engines such as Google, Microsoft or Yahoo!, is openly shared with the scientific community.<sup>930</sup> Finally, the SEO and SEM industry harbors a lot of knowledge about the selection and ranking practices of search engines based on continuous practical experimentation as part of their services to their customers.

---

<sup>927</sup> Since 2008, Google has published a range of articles specifically addressing search quality on *The Official Google Blog*. These articles are directed at a general audience and together provide quite some information about search quality issues in general and a good picture of Google's views on the issues involved. See e.g. Manber (Google) 2008; Singhal (Google) 2008; Singhal (Google) 2008b; Gomes (Google) 2008a; Gomes (Google) 2008b; Huffman (Google) 2008; Dembo (Google) 2008; Baker (Google) 2010; Aula & Rodden (Google) 2009; Cutts (Google) 2011; Champaneria & Yang (Google) 2011.

<sup>928</sup> One of them is PageRank, although the PageRank algorithms may no longer be used in the form it was described by Google's founders in their 1998 scientific paper, Brin & Page 1998.

<sup>929</sup> For a comprehensive overview of the state of the art in information retrieval from an computer science and engineering perspective, see e.g. Baeza-Yates & Ribeiro-Neto 2011. Baeza-Yates is working for Yahoo! Research.

<sup>930</sup> For a discussion, see e.g. Tunkelang 2010.

This section will address two of the regulatory issues related to the lack of transparency about selection and ranking of search results that have arisen in the last decade and discuss them from the perspective of search engine quality and freedom of expression. First, the typical business model of commercial search engines involves the sales of sponsored search results. This raises a number of issues relating to the transparency about search engines' motives to produce certain search results and the proper level of disclosure of the mechanisms behind search engine result page composition. Section 10.3.1 will review and critically reflect on the regulatory debate about search advertising and discuss some of the more fundamental structural complexities of transparency of search engine advertising. More specifically, the analogy between search media and traditional commercial mass media that underlies the current separation and labeling of organic and sponsored results will be critically discussed, drawing on an analysis of the dynamics of search media on the one hand and the discussion about the press in Chapter 5 on the other hand. In particular, some of the more structural limitations relating to the possibility of freedom from advertiser and third party influence for commercial search engine providers will be discussed, limitations which result from their relative weak distributed editorial process and their status of meta-media.

Second and related, the impact of the business model of dominant search engines on search engine quality has not been fully understood as well as the strategies search engines pursue to optimize their income more generally. Search engines have been accused of unlawfully manipulating the ranking of certain websites on various legal grounds, such as unfair competition. Some notable examples of this were already addressed in Chapter 8, and there are various other similar examples of such accusations, resulting, amongst other legal developments, in antitrust investigations into Google search in the EU as well as in the United States.<sup>931</sup> On a fundamental level, these accusations raise the question of to what extent search engines should or could be forced to disclose their business strategies and specific decisions about ranking and selection of search results and to what extent they should be allowed to freely determine the composition of their search result pages. These issues will be discussed in Section 10.3.2.

### **10.3.1. Search engine marketing: the labeling of sponsored results**

Search engines have become one of the most attractive marketing platforms in the networked information environment. Search engines are amongst the most popular and heavily used online services. They process billions and billions of search queries and each query presents the service with an opportunity to sell targeted user attention to the highest, or more precisely the most profitable bidder.<sup>932</sup> As discussed in previous chapters, all dominant general purpose search engines now monetize their services through the sales of sponsored search results through complex auctioning systems. The organic results are separated from these specially labeled 'sponsored results', which offer advertisers the opportunity to market their products, services or points of view more prominently to users. This

---

<sup>931</sup> See Dakanalis and Van Rooijen 2011. See also Sterling 2010, Consumer Watchdog's Inside Google 2010; Euractiv 2010, Catan & Efrati 2011; Singhal (Google) 2011; Heise Online 2011; Pollock 2010.

<sup>932</sup> When optimizing their income, advertising platforms like Google AdWords do not sell to the highest bidder. Google's revenue through AdWords is not only related to the willingness to pay a certain price per click but also related to the actual click through rates of specific sponsored results. See e.g. Varian 2009 and cited references.

practice is also referred to as paid placement. In addition, some search services offer paid inclusion, meaning that they guarantee certain service levels with regard to the crawling and inclusion of a website in the search engine's index in return for a fee. The case of meta-search engines or the integration of vertical search engines into general search engines could be seen as a special case of paid inclusion.<sup>933</sup>

Both paid placement and inclusion have led to regulatory debate about the appropriateness and lawfulness of these monetization strategies, as end-users may be misled about the composition of search result pages. More generally, the monetization of search engine traffic through paid placement has led to an ethical debate in the search engine industry about these practices.<sup>934</sup>

In 2002 the United States Federal Trade Commission reacted to a consumer organization complaint about deceptive practices by search engines in terms of paid placement and inclusion by outlining its views on the permissibility of these practices. In a recommendation to the industry, the FTC argued that with regard to paid placement:

*“search engines should clearly and conspicuously disclose that certain Web sites or URLs have paid for higher placement in the display of search results. [...] The failure to disclose paid placement adequately within search results deviates from the established deception principle of clearly distinguishing editorial content from advertising content. The purpose of such a demarcation is to advise consumers as to when they are being solicited, as opposed to being impartially informed.”*<sup>935</sup>

Most search engine providers had already started to act in line with this guidance offered by the FTC and this recommendation quickly has become an industry standard practice, including in Europe.<sup>936</sup> Hence, it is worth taking a closer look at the reasoning and premises behind this standard of clearly demarcating sponsored results from organic results. In particular, from a regulatory perspective it is worth taking a closer look at the implicit underlying analogy with commercial editorial media and the value of demarcating editorial content from advertising in that context.

In Europe the Directive on Electronic Commerce contains a specific obligation for an information society service to be transparent about commercial communications. Notably, search engines do, in principal, fall under the scope of this directive, as they are information society services.<sup>937</sup> Article 6 contains a specific transparency obligation with regard to commercial communications in information society services. This provision reads as follows:

Article 6 - Information to be provided

---

<sup>933</sup> See also Section 2.2.

<sup>934</sup> For a discussion of Google's early struggle with the question about ethical advertising strategies, see Douglas Edwards 2011, Kindle edition, location 1253-1305.

<sup>935</sup> Federal Trade Commission 2002.

<sup>936</sup> In Germany, for instance, see e.g. W. Schulz et al 2005b, pp. 1419-1433.

<sup>937</sup> See Article 2 sub (a) of Directive 2000/31/EC juncto Article 1 (2) of Directive 98/34/EC as amended by Directive 98/48/EC. See also ECJ 23 March 2010, Joined Cases C-236/08 to C-238/08 (Google Adwords), at 110.

In addition to other information requirements established by Community law, Member States shall ensure that commercial communications which are part of, or constitute, an information society service comply at least with the following conditions:

(a) the commercial communication shall be clearly identifiable as such;

(b) the natural or legal person on whose behalf the commercial communication is made shall be clearly identifiable [..]<sup>938</sup>

When applied to search media this provision seems to call, like the recommendation of the FTC, for a clear demarcation of sponsored and organic results. The ECD entails a broad, technology-neutral definition of commercial communications in Article 2 sub (f) as *“any form of communication designed to promote, directly or indirectly, the goods, services or image of a company, organization or person pursuing a commercial, industrial or craft activity or exercising a regulated profession.”*<sup>939</sup> Article 2 sub (f) ECD contains two exceptions, both of which are relevant in the context of search. The first exception is for *“information allowing direct access to the activity of the company, organization or person, in particular a domain name or an electronic-mail address.”*<sup>940</sup> As long as search results can be seen as purely navigational, this exception could apply. In reality, however, both organic as well as sponsored search results tend to contain promotional information and offers for specific goods or services which would make these results fall outside of the exception. The second exception is for *“communications relating to the goods, services or image of the company, organization or person compiled in an independent manner, particularly when this is without financial consideration.”*<sup>941</sup>

Hence, organic results in search media are probably not commercial communications affected by the obligations in Article 6 ECD; as long as providers are acting independently or simply do not receive compensation for their presentation. In other words, this interpretation of Article 6 and the definition of commercial communications of the ECD lead to a similar legal obligation in the European context with regard to the demarcation of organic and sponsored search results which that the FTC recommends. Sponsored search results are commercial communications under European law and need to be clearly identifiable. In addition, the entity behind the communication should be clearly identifiable.<sup>942</sup>

Upon a closer look, however, it is questionable whether the stated goals of the FTC’s recommendation can be achieved with the labeling of sponsored search results that has become industry practice and is expected from a regulatory perspective. The difference between commercial and other types of communications under European law may not so easily overlap with the difference between organic and sponsored search results.

---

<sup>938</sup> 2000/31/EC.

<sup>939</sup> Directive 2000/31/EC.

<sup>940</sup> Directive 2000/31/EC.

<sup>941</sup> Directive 2000/31/EC.

<sup>942</sup> Notably, this last obligation was taken into account in the ECJ’s judgment on Google AdWords, in which the Court concluded that a company is not allowed, under trademark law, to use trademarks of others as keywords, if Internet users would be confused about that company’s identity. See ECJ 23 March 2010, Joined Cases C-236/08 to C-238/08 (Google Adwords), at 99.

Turning back to the FTC recommendation, which unlike Article 6 ECD explicitly focuses on search engines and advertising transparency, it is important to note that it is based on search engine user expectations. Search engines could, because of the inclusion or placement of advertisements in search results without disclosing these practices to their users, deceive these users about the reason behind the inclusion and ranking of certain results. At the time of the FTC complaint reports claimed that a majority of search engines users were unaware of paid placement in search engines.<sup>943</sup> The solution to require more transparency in regard to these practices seems logical. However, for such transparency to make sense in terms of possible deception of users about the value of search results, one first has to ask what justified expectations end-users may have about the selection and ranking of search results in the first place.

In the recommendation, the FTC argued, that because search engines

*“historically displayed search results based on relevancy to the search query, as determined by algorithms or other objective criteria, the staff believes that consumers may reasonably expect that the search results displayed by individual search engines are ranked in accordance with this standard industry practice - that is, based on a set of impartial factors.”<sup>944</sup>*

There are a number of obvious and less obvious problems with this reasoning. First, labeling of different types of search results presupposes a level of knowledge of the functioning of search engines in general, which is arguably absent. Second, the user expectations that the ranking and selection of organic search results is objective, impartial or independent of third party influence may not be justified. Third, the value of current labeling practices itself breaks down due to the functional similarities between sponsored and organic results for end-users as well as for information providers.

As discussed before, search engine users typically do not have a good idea how search engines actually select and rank their results. Search engines are and have been secretive about the precise particularities of their selection and ranking practices. And the end-users mentioned above who are unaware of advertising practices in search engines can be expected to understand even less about the selection and ranking of organic search results more generally. It requires expert knowledge to understand even a little of the actual functioning of search engines and the reasons why certain search results may have been selected and ranked highly in response to a particular query.<sup>945</sup>

Moreover, the way in which search engines select and rank online destinations in their organic search results leaves room for the deception the demarcation is aiming to prevent. It might simply be the case

---

<sup>943</sup> See Princeton Survey Research Associates 2002. A later study from the Pew Internet & American Life project confirmed these results; 62% of searchers in the U.S. market were found to be unaware of the distinction between paid and unpaid results. See Fallows 2005, p. 17.

<sup>944</sup> Federal Trade Commission 2002.

<sup>945</sup> For instance, even relative experts are still confused about the question if Google sells access to higher placement in organic results. In a recent higher educational guide on using search engines, including Google, for research, the writer states: “As is known, you end up higher in Google’s search results if you pay for it – that is how the search engines makes its money. For instance when one types: books Willem Bilderdijk (number of results 129.000), gets two references to bol.com on the top (and these are no advertisements (Translation by the author).” See Sanders 2011, p. 13.

that end-users cannot justifiably expect the selection and ranking of search results to be based on an independent, objective and impartial valuation by search engine providers, but will have to make their own valuation instead. Put differently, search engines may rely too much on third party signals and do not – and probably cannot on the scale at which they operate – exercise the amount of editorial control which could be considered the prerequisite of a claim to objectivity, independence and impartiality. Some of these choices with regard to organic results could entail the prioritization of commercial communications. And more generally, providers might themselves deploy ranking algorithms that are not based on ‘scientific’ criteria to objectively assess the ‘relevancy’ of certain sources of information in respect to specific queries, but rather on their own undisclosed commercial or ideological motivations and editorial choices with regard to the selection and ranking of search engine results more generally.<sup>946</sup>

Hence, information providers have too much control over the ranking and selection of organic search results for the demarcation to make sense for end-users. This may be best illustrated by the fact that SEO of organic results is considered an effective strategy to market products and services, and even a substitute for advertising in sponsored search results.<sup>947</sup> Therefore, the FTC’s statement that “*any Web sites [...] that have paid to be ranked higher than they would be ranked by relevancy, or other objective criteria, should be clearly labeled as such using terms conveying that the ranking is paid for*” is confused and arguably naïve in its trust of objectivity and impartiality of organic search results.<sup>948</sup> The application of the provisions from the Directive on Electronic Commerce on the search medium discussed above suffer from the same deficit, by substituting the requirement ‘*without financial consideration*’ for the requirement “*in an independent manner*.”<sup>949</sup>

Ironically, one of the reasons for the effectiveness of the optimization of organic search results might be that some end-users unjustifiably expect that these results are delivered free from commercial influences.<sup>950</sup> Similarly, end-users may undervalue sponsored search results as proper answers to their queries for the mere reason that they are labeled as sponsored links. However, the fact that end-users use search engines in the context of e-commerce implies that advertisements can be quite relevant or even more relevant than non-commercial communications from the perspective of end-users. In fact, the willingness to pay, measured through an advertisement auction, is a perfectly objective measure to rank advertisements and can be used as one of the signals to assess the relevance of an online destination for end-users.

Finally, there are some conceptual problems with the application of the traditional demarcation between editorial content and advertising in editorial media to the context of search media. On a functional level, search media – and other selection intermediaries like Twitter for instance – may be an example and perhaps even the main driver of convergence of information retrieval, recommendation and advertising systems, in a way that cannot be disentangled by the mere labeling of so-called

---

<sup>946</sup> See also Section 8.4.3.

<sup>947</sup> See Yang & Ghose 2010, p. 2. See also Arnold et al 2011.

<sup>948</sup> Federal Trade Commission 2002.

<sup>949</sup> See Article 2 sub (f) Directive 2000/31/EC.

<sup>950</sup> See Jansen & Spink 2009.

sponsored results.<sup>951</sup> Organic links and sponsored links are the same type of targeted reference: a link, a title and a short explanation, none of which tend to be written by the search engine provider.<sup>952</sup> Search media are not editorial with respect to their organic search results in the way that newspapers ideally are. Search media typically include everything in their index anyway and allow information providers to control what ends up being indexed or not. Advertisers do not participate in the sponsored results programs of major search engines to be included at all, as would typically be the case with advertising in editorial media. They participate to acquire relatively more prominence and subsequent attention of end-users. As a result, sponsored links will typically supplement organic results to the same information provider, and these two types of links to the same destination can be very similar and do not tend to provide different perspectives for the end-user. Paradoxically, the sponsored results are more heavily editorialized than the organic ones: they typically have to follow editorial guidelines for search advertisements, while, in principle, no such editorial guidelines exist for organic results.<sup>953</sup> The distinction between organic and sponsored links breaks down even more, because the ranking and selection of sponsored links, as is the case for organic results, heavily depends on the relevancy of the underlying content with respect to the search query entered by the user, as well as their success in terms of click-through rates, and not merely the willingness to pay by advertisers.<sup>954</sup>

Notably, in the context of the press the labeling of advertising is only a small part in a more general best practice of the treatment of advertising and special interests and the protection of the press' and journalistic independence and impartiality. Considering that the model of editorial media has helped to shape their treatment of paid inclusion programs, the search engine industry could adopt some of the related best practices in the field of editorial media. Search engines could for instance structure their internal operations in such a way to prevent direct profit considerations from having a negative impact on the quality of their search results.<sup>955</sup> They could adopt codes of conducts about the way in which they deal with the various incentives which arise from their advertisement business model. While some search engine providers, including Google, would claim that they are already acting in line with such recommendations, the problem is that they have not yet developed a framework that makes their statements verifiable for the general audience. In one way or another, more transparency and disclosure will be needed for the labeling of search results to become really meaningful. For instance, more meaningful self-regulatory practices would include some kind of independent audits to verify the truthfulness of stated practices.

To sum up, while the demarcation between sponsored links and organic results was meant to prevent end-users from being deceived about the mechanism behind the composition of search result pages, its

---

<sup>951</sup> On this new type of convergence, see Garcia-Molina et al 2011.

<sup>952</sup> See Section 3.2.

<sup>953</sup> One could consider the guidelines for permissible SEO of organic results to be editorial guidelines. Such guidelines include a ban on the sales of outgoing links, hidden links, and the development of shadow domains or doorway pages. See Google 2010b.

<sup>954</sup> Thus a lack of relevance leads to lower rankings (or even removal) of the sponsored link, but can be compensated in a willingness to pay more per click in the auction.

<sup>955</sup> See Van Couvering 2010. Van Couvering's interviews in the search engine industry show this to be considered an issue internally and to the extent that the different perspectives of information retrieval engineers and marketers lead to internal tension.

logic is too simplistic to capture what is really going on. In reality, end-users' expectations with regard to organic results cannot justifiably be in line with the FTC's premise. Instead, end-users should be or be made aware of the many ways in which organic search results are in fact shaped by different forms of outside pressure, including pressure of advertisers but also of other special interests. Search engines should complement the existing practice of labeling with additional internal policies and strategies to prevent the crowding out of objectively valuable, non-commercial references in their index. They should also make themselves publicly accountable for adhering to these policies. In the absence of such additional safeguards, search engine users may be better off, not having any expectation of impartiality, objectivity of any type of search result. In that case, they will have to simply rely on their own judgment about the value of the references that are provided to them.

### **10.3.2. Search engine advertising, quality and transparency**

While the labeling of sponsored search results at least clarifies that search media engage in paid placement strategies, the monetization of search traffic raises other issues with regard to the ranking and selection of search results and search engine quality more generally that have yet to be fully understood. For instance, an important question is what the impact is of the monetization of search engine traffic on the overall quality of search media and the willingness of search engines to innovate and improve the quality of their search results. Transparency about the strategies of existing search media in the context of ranking and selection to optimize their income is lacking. If anything, commercial search media can be expected to optimize their income through the optimization of clicks on sponsored search results. This could have adverse effects on the quality of search media seen as information retrieval systems from the perspective of end-users. At the very least, the business interests of commercial search engines to monetize as much traffic as possible implies that any statements of search media that present their selection and ranking strategies as merely directed at satisfying information needs of end-users should be critically assessed.

As discussed in the previous section, until now regulators such as the FTC cited above have been rather uncritical in following search engines' claims about independence, relevance and objectivity. Similarly, the ECJ's Advocate-General was happy to state that:

*“natural results are selected and ranked according to their relevance to the keywords. This is done through the automatic algorithms underlying the search engine program, which apply purely objective criteria.”*<sup>956</sup>

The uncritical adoption of these premises about the functioning and goals of commercial search engine providers is also the basis of the current labeling of sponsored search results, which, as argued in the previous section, falls short in providing real value for end-users. While the values of independence and impartiality in search engine ranking and selection practices could surely contribute to search engine quality for end-users, selection and ranking practices of dominant search media currently depend too heavily on third party signals to be characterized as such.

---

<sup>956</sup> ECJ, AG Opinion, Joined Cases C-236/08, C-237/08 and C-238/08 (*Google Adwords*).

On the one hand, the ranking and selection practices of major search media are a complicated, mostly opaque and automated set of decisions about the 'value' and 'relevance' of sources of information in view of a particular understanding of information needs of end-users. As argued in Chapter 8, the realization that selection and ranking involves a variety of subjective decisions about the relevance and value of information also offers search engine providers a freedom of expression defense to freely decide which choices to make in this context. As the Financial Times concludes, summarizing various critiques of the supposedly objective nature of search engine rankings:

*“By changing its mathematical formula to modify the results returned to a particular query, Google’s engineers are making judgments very similar to the editorial decisions made at a more traditional media organization [...].”<sup>957</sup>*

On the other hand, the judgments about the ranking and selection of search results may be primarily aimed to advance the business interest of the search medium. In fact, the advertisement based business model has some obvious implications for the incentives on search engine operators to make certain choices rather than others. Search engines are in the business of matching information needs of end-users and the offer and willingness to communicate of information providers. It is important to realize that this is exactly what advertisers have been doing all along, be it with one particular mindset.

Ultimately, the question about search engine quality is a question about the values and goals of search engine providers and the compatibility of those values with societal interests that go beyond their direct business interest. And since end-user attention is scarce, the question is, ultimately, what kind of end-user information needs a search engine decides to satisfy: those that it considers, on the basis of some idea about the scientific, societal or political value of information and ideas, most worthy of finding satisfaction or those that would be most profitable? It would be quite unconvincing to argue that these perspectives will always overlap. And again, the question is how search engines will deal with the fact that information providers' willingness to find the attention of end-users, whether expressed through advertising in search engines or through SEO, does not necessarily reflect their value as sources of information.

The negative impact of advertising on broadcasting quality is well-documented: commercial media can in general be expected to produce the level of quality of programming that is necessary to keep people watching.<sup>958</sup> This is used as one of the reasons for the subsidization of public broadcasting and the regulation of advertising in broadcasting. Will commercial search engines act similarly and produce the minimum level of quality of results that ensure that people will not switch to a competitor? Should regulators respond with public service search engines or search engine regulation? More specifically, can Google, as market leader, be expected, and allowed if it does so, to mainly focus on sponsored links prominence and quality? And in view of their business model, how important is it for major search engines to preserve or improve the quality of their organic listings, considering that clicks on organic links can be seen as missed opportunities for them to monetize attention? Notably, in the context of

---

<sup>957</sup> Waters 2010.

<sup>958</sup> See e.g. Baker 2002, Benkler 2006, pp. 176-212.

commercial search engines, the functional overlap between sponsored and organic search results further increases the pressure on the quality of organic search results. Both organic as well as sponsored search results are avenues in which information providers can influence their prominence.

Currently, it is still easier to ask these questions than to answer them properly and regulators have only quite recently started to address some of these issues more seriously, most notably in the competition law context and possible abuse of market power of industry leader Google. The European Commission addressed the possible impact of advertising on organic search quality in its review of the deal about organic search listings between Microsoft and Yahoo.<sup>959</sup> Generally, the Commission is quite optimistic about the current levels of innovation in terms of search engine quality:

*“Search engines spend enormous efforts in order to continuously improve the search algorithm and their ability to match users and advertisers. [...] The high degree of innovation in these markets is proven by the rate at which new innovation has been introduced in the past (for example introduction of the auction mechanism, quality rank, analytics, spell check, etc.) and by constant experimentations that search engines undertake daily on several fronts.”<sup>960</sup>*

Further on, the Commission explains what kind of possible market behavior could be expected and considered negative for the quality of search engines for end-users:

*“Theoretically, the rationale for possibly degrading the organic search stems from the trade off that search platforms appear to face between the incentive to provide relevant organic and paid results. The trade off arises because when a platform tries to attract more users through greater relevance on the organic search it runs the risk of losing revenues on the advertising side (i.e. less clicks on ads) due to users clicking predominantly on the organic side (especially if both types of clicks would bring the user to the same kind of information). [...]*

*[P]latforms might have an incentive to dedicate a smaller part of the result page to organic results in favour of search advertising links [] thereby providing proportionally more advertising links. Alternatively, the platforms may rank the sponsored and organic search results in a way that firms offering competing products to the sponsored links are ranked, from the user's perspective, on the organic side lower than optimally.”<sup>961</sup>*

In its actual analysis of the impact of the merger on the search engine market the Commission does not come to the conclusion that this behavior takes place.<sup>962</sup> Due to the nature of competition law, it merely draws the conclusion that the deal under review would not worsen the situation. It does so on the basis of four arguments, namely that (1) users would be responsive with respect to variations to the relevance provided and be able to switch to alternatives, that (2) there was very limited switching of end-users between the two parties under review, that (3) other search engines (Ask or local ones) and competing

---

<sup>959</sup> European Commission 2010a.

<sup>960</sup> European Commission 2010a.

<sup>961</sup> European Commission 2010a.

<sup>962</sup> Currently, there seems to be no hard evidence of such practices taking place in the major search engines operating in Europe and the U.S. The current complaints at the European Commission, when reviewed, could provide more information.

verticals could still be expected to exercise a constraint on the main competing platforms not to degrade their service, and that (4) the improved scale of organic traffic for Microsoft would allow it to actually improve the quality of its search results, thereby keeping up pressure on search quality for Google as well.<sup>963</sup>

From the perspective of competition law, these conclusions seem reasonable. Competition law can be expected to provide some constraints on search engine quality for end-users, but its role is necessarily limited for several reasons. First, from a market perspective it is difficult to address the impact of search engine market dynamics on end-users, since they (or rather, their attention and targeted navigation) are the product being sold instead of the paying customers. The Commission addresses this issue by using the economic theory of two-sided platforms. However, its analysis remains simplistic. In particular, it does not address the trade-offs between the participants on the two sides of the market. What could be considered harmful for end-users from the perspective of search engine quality, such as more and more space reserved for sponsored results,<sup>964</sup> can be expected to be beneficial for advertisers. Competition law does not answer which sets competing harms and benefits are preferable.<sup>965</sup> In addition, advertisers are competing with other information providers for attention. It is hard to imagine how competition law can address the trade-offs between the interests of advertisers and other sources of information worthy of representation in search results or the trade-offs between different types of users.

While this analysis of the European Commission did not directly target the business practices of Google, in 2010 both the European Commission and U.S. antitrust authorities started to investigate complaints about possible anticompetitive ranking and selection practices of Google.<sup>966</sup> Amongst other complaints and accusations of abuse of dominance, the investigations are focusing on the anti-competitive effects of vertical integration of general purpose search engines and the possible preferential treatment of Google's own services in Google's search results.

Considering the complexities of the underlying markets, the result of these investigations can be expected to be of major significance for the future of the search engine industry. In addition, the investigations go to the heart of the question about the freedom of (dominant) search engine providers to rank and select search results and the proper limitations on such freedom in view of possible harm for third parties. Consider for instance the claim that Google harms competition by degrading the ranking of search results to competing verticals search engines and directories. While it may be questionable whether this behavior takes place, and if it takes place, whether it should be considered an abuse of

---

<sup>963</sup> European Commission 2010a, § 219-223.

<sup>964</sup> The European Commission does qualify this as a possible harm for search engine users, but concludes that the deal between Yahoo! and Microsoft is unlikely to provide both parties with additional incentives in this direction, considering the dominance of Google and the lack of competition between the parties under review. It does not discuss this harm in relation to possible benefits for the advertisers.

<sup>965</sup> Notably, some have complained about Google's emphasis on non-commercial results in its organic listings, such as Wikipedia, which pushes commercial information providers to participate in Google Adwords. See also Edelman 2009. There are also rumours that Google would rather have websites that run AdSense, hence other than Wikipedia, rank more favourably but that it feels a similar change would harm its credibility. The source for this rumour adds that Google watching is like 'Kremlinology'.

<sup>966</sup> European Commission 2010d. See also Bartz, 2011. See also Mayer (Google) 2010.

market power, it also raises the question of to what extent the law could and should limit the freedom of search engines to decide that verticals and directories are simply inferior destinations from the perspective of search engine quality. In other words, the right to freedom of expression and the editorial freedom it arguably offers search engines with respect to the choices of which selection and ranking practices to engage in should be taken into account in these investigations by competition authorities.

Under the current legal framework and apart from the specific labeling obligation discussed in the previous section, transparency about ranking and selection is provided voluntarily and encouraged through self-regulation. There are no general legal disclosure requirements for search engines about their ranking and selection practices, which also typically enjoy protection as trade secrets. Some argue that it will be necessary to impose specific transparency obligations on major search engine providers. The most sophisticated proposals in this direction have proposed to install special oversight committees to prevent general disclosure to the public and competitors, while still providing the benefits from the perspective of accountability.<sup>967</sup>

While everything may seem to point in the direction of the value of more transparency about ranking and selection practices, it is important to note that there are some arguments against disclosure requirements for search engine ranking and selection practices. First of all, search engine providers would argue that disclosure might decrease search engine quality to the detriment of end-users, as it would allow interested parties to game their systems.<sup>968</sup> The production of search engine results takes place in a continuous competition for Internet traffic, search engine optimization and marketing. The current levels of disclosure and transparency are at least to some extent tailored to address the negative impact of this ongoing competition.

Second, and more fundamentally, the nature of the information that would have to be disclosed goes to the heart of the editorial aspect of the search medium. The observation that decisions about ranking and selection of search engines are editorial decisions implies that the law should not too easily impose restrictions on the freedom to decide how to rank and select search results. And one may argue that the protection of the search provider's freedom to select and rank search results based on a theory of editorial freedom should also involve a limitation on requirements to disclose the precise practices, since such a requirement could result in chilling effects and provide search engine providers with incentives not to engage in certain selection and ranking practices. As a comparison, one would not imagine similar obligations on traditional editorial media. No one would propose to legally and generally require publishers to factually disclose the actual decisions that are made in the editorial room. This kind of regulation would be considered an infringement of the freedom of the press. In sum, from the perspective of the right to freedom of expression it is preferable that transparency would be offered through self-regulation.

Another feasible approach would be the more systematic monitoring of the industry and the funding of independent research into the incentives search engines have with regard to the interplay between

---

<sup>967</sup> See e.g. Bracha & Pasquale, 2008, p. 1201.

<sup>968</sup> See e.g. Grimmelmann 2007; Van Couvering 2010.

quality of their organic search results and the optimization of income through sponsored search listings.<sup>969</sup> In addition, in the absence of sector-specific regulation of search engines competition law may be expected to continue to provide some insights into the selection and ranking practices of dominant search engine providers. In particular, the regular reviews of commercial agreements and the recent investigations into the complaints about the abuse of market power involving or relating to Google's search business by competition authorities will help to provide some transparency about the functioning of the search engine market, the possibilities and realization of abuse, and the motives behind the production of particular search result orderings.

#### **10.4 Search engine quality and user data: privacy, personalization and intellectual freedom**

This final section will address the regulatory issues relating to search engine quality and the processing of user data by search engines. High quality user data could be considered the holy grail of search engine providers (and most other commercial Web services for that matter). In his book on Google and the importance of search, John Battelle called Google's collection of user data 'the database of intentions'.<sup>970</sup> It goes well beyond the scope of this research to fully discuss the practices of user data processing by major web search providers like Google. In the next section, a short overview will be provided, after which some of the legal issues relating to the processing of user data by search engine providers will be discussed from the perspective of the right to freedom of expression. More specifically, Section 10.4.2 will address the way in which the observance of the right to privacy and data protection laws can be seen as a prerequisite to intellectual freedom of search engines users as well as the way in which the current legal regime incorporates this concern. Section 10.4.3 will discuss the way in which EU data protection law, as interpreted by the Article 29 Working Party, could in theory contribute to the informational autonomy of search engine users, if it were to be applied more rigorously on current data processing practices.

##### **10.4.1 Search engine user data processing: background**

Search engines rely on complex, sophisticated processing of massive amounts of user data in many aspects of their service. Many users may only be aware of providing one relevant data point to search engine providers, namely the search query. But search engine user input consists of a range of data, over significant amounts of time. Some of these data are also directly provided by the user such as search preferences, which are typically stored through the use of Web cookies, a file on the users' computer that helps to individualize end-users. In addition, data relating to a search will be registered and processed, such as the date and time a search query was entered, as well as subsequent user actions like a click on a search result. A lot of other data are provided by the user as a result of them interacting with the Web search service, such as their IP address, the operating system, the browser and its settings. Users are typically tracked over many sessions of use with the use of so-called cookies containing unique identifiers. Whereas the IP address might not mean much to the average Internet

---

<sup>969</sup> The incentives this produces is a complex question in search engine economics. See e.g. White 2008; Yang & Ghose 2010, p. 2. See also Arnold et al 2011; Chan et al 2011.

<sup>970</sup> Battelle2006. Since then Battelle has extended the database of intentions, for which major Web services compete, to include the social environment, the status of action, as well as the location of individuals. See Battelle 2010.

user, these data allow web services to say something about someone's geographic location, when combined with other data sources. Similar combinations can subsequently be made with other data sets, such as demographics connected to ZIP codes.<sup>971</sup>

User data is an essential ingredient for the monetization of search, and competition in the search engine and Web services market more generally is strongly focused on the ability to collect user data.<sup>972</sup> The advertisement based business model of search engines essentially turns around user data as every click by a user on a sponsored result is a micro payment from the advertiser to the service. These user clicks are registered and are optimized in view of the profitability of the service. And in view of the interests of paying customers, these clicks need to be legitimate and not fraudulent. In addition, the more user data search engines have, or perhaps better, the more search engines can infer with some degree of statistical relevance, the more targeted the advertisements that are displayed can be. The level of targeting increases the value of their product for advertisers, thereby increasing the profits of the search engine provider.<sup>973</sup>

User data collections also allow search engine providers to do research on user behavior.<sup>974</sup> Amongst other things, this research involves machine learning on large user data sets. The goal of such research could be to develop new statistical qualifiers to improve ranking and selection and sponsored result placement. User data can be analyzed in view of relevance and user satisfaction, particular patterns of user behavior can be predicted to be possibly fraudulent and search queries in general can be better understood. User data can also be used to innovate and develop new services, an often cited example of which is the 'suggestion' function that corrects spelling mistakes and suggests possible improvements of the search query.

Finally, user data can be used to personalize the service, in the sense that search engine providers will select and rank their search results on the basis of that user's history of interaction with the service or other data related to that specific user.<sup>975</sup> Any of the more simple data mentioned above, such as the IP address or language preference can be used to personalize results, but other and more sophisticated signals could be used also, such as historic preference for certain media formats, for instance video's or link navigation speed. Personalization can also be based on specific combinations of signals and rich data collections available from other services, such as email, social networks or video on demand. Since 2009 Google personalizes both its organic and sponsored results for all its users.<sup>976</sup> For authenticated users, it promotes websites that have been visited before, shows how many visits took place, and more generally seems to develop a kind of profile of interests which has an impact on user's search results.<sup>977</sup> As mentioned, Google personalizes search results on the basis of a variety of criteria, most of which are

---

<sup>971</sup> See e.g. Weber & Castillo 2010.

<sup>972</sup> See Bermejo 2007; Röhle 2007b. See more generally, Elmer 2004.

<sup>973</sup> See generally Turow 2006.

<sup>974</sup> See e.g. Baeza-Yates & Ribeiro-Neto 2011, pp. 185 et seq.

<sup>975</sup> See Pariser 2011; Feuz et al 2011; Zuckerman 2011. See also Stalder & Mayer 2009.

<sup>976</sup> See Sullivan 2009.

<sup>977</sup> See Feuz et al 2011.

unknown. It is clear that the possibilities to personalize further are only just being explored and can be expected to become the subject of intense competition between dominance Web services.

The regulatory debate about the processing of user data by search engines has been channeled through data protection and the right to privacy and confidentiality topics. Early commentators' worries about the amounts and sensitivity of data being registered by search engines were publicly confirmed in 2006 when AOL released a set of user data for research purposes.<sup>978</sup> The data set consisted of a large number of search histories of AOL users, in which the IP addresses had been replaced by a unique number. The release was generally considered a mistake because of the sensitivity of the data. The New York Times showed in their report 'A Face is Exposed for AOL Searcher No. 4417749' that the search history could be used to identify an actual Internet user.<sup>979</sup>

The regulatory debate intensified in the years after the release of the AOL search logs. Most notably, European data protection authorities started an investigation into the processing of user data from a European data protection law perspective, first addressing only Google and later addressing the Web search industry at large. This investigation was concluded in 2008 in the form of an official opinion by the Article 29 Data Protection Working Party on 'data protection issues related to search engines'.<sup>980</sup> In the United States the FTC looked into search engine user data processing in the context of its recommendations about data privacy safeguards (in the form of self-regulation) with regard to behavioral advertising.<sup>981</sup>

As mentioned above, the scrutiny of the processing of search log data processing has mostly taken place through the lens of data protection and privacy law and regulation. Consequently, the regulatory debate has placed particular emphasis on data privacy related issues, such as the anonymization of user data and the retention period of individual search queries in non-aggregate form, the question about third party access to search query data, and the legitimacy of the goals underlying the processing of so much user data by search engine providers in the first place. Apart from these issues, the question about the proper application of jurisdictional provisions of the EU Privacy Directive (95/46/EC) to operating Web search engines operating from outside the EU and the question of whether or not the user data in search logs should be qualified as 'personal data' under the Privacy Directive have been important in the legal debate about the proper application of EU law in this context.

Much less attention has been paid to the question of how individual user data processing has shaped and continues to shape access to information by Internet users. In line with the focus of this study on the right to freedom of expression in the context of search, this section will specifically address this perspective by addressing two of the issues at the interplay of data protection, search engine user

---

<sup>978</sup> On the importance of access to search logs for the research community and the negative impact of the public outcry about the release for information retrieval research, see Bar-Ilan 2007.

<sup>979</sup> See Barbaro & Zeller 2006. See also Lenssen 2006. The data release has had a mayor impact on search engine privacy policy discussions and has been used by legal scholars, computer scientists and artists. See e.g. Tene 2008; Cooper 2008. Dutch artists and documentary makers Lernert & Sander made the movie 'I Love Alaska' about someone's queries. See Lernert & Sander 2008.

<sup>980</sup> Article 29 Data Protection Working Party 2008.

<sup>981</sup> See Federal Trade Commission 2007. See also Federal Trade Commission 2009.

privacy and freedom of expression. First, in section 10.4.2, the question will be discussed how the current privacy and data protection framework for search engine user data processing takes into account the freedom of expression interests of users to seek and access information and ideas freely. Generally, the instrumental value of privacy for the right to freedom of expression is well accepted.<sup>982</sup> For instance, this value lies in the possible chilling effect of the absence of appropriate privacy safeguards on expressive conduct, including the search for and the access of information and ideas online.<sup>983</sup> In the context of networked communications, the free exercise of the right to freedom of expression takes place in a new environment and with the use of new types of services, such as Web search engines. When data processing practices by new intermediaries do not go hand in hand with appropriate data protection and privacy safeguards, they may not only impact the right to privacy and data protection but also profoundly impact the enjoyment of the right to freedom of expression.

In Section 10.4.3 the question will be addressed if and how European data protection law could promote accountability in the context of search from a user's perspective, in particular with regard to the impact that search engine user data has on the selection and ranking of organic as well as sponsored search results. Building on the conclusion of European data protection authorities that much of the processing of user data by search engine providers amounts to the processing of personal data - processing which therefore falls under the scope of EU data protection laws -, the question will be explored in what ways data protection law could contribute to transparency and accountability in the context of search engine ranking and selection practices. This question has grown in relevance due to the increased personalization of search results. It will be shown that, in theory, EU data protection law may indeed promote the informational autonomy of end-users in the context of search, but that until now data protection compliance and enforcement levels has been wanting in delivering this potential.

#### **10.4.2 Intellectual freedom and search engine user surveillance**

As was noted in Chapter 8, the search engine user's privacy can be seen as a condition for the fundamental right to search, access and receive information and ideas freely. Information seeking behavior could be seriously chilled if the main available options to find information online entail comprehensive surveillance and storage of end-user behavior without appropriate guarantees that such information is not be used to one's disadvantage.

Historically, libraries have had the most experience with the need to guarantee the privacy of their patrons and the confidentiality of their reading habits.<sup>984</sup> In Chapter 7 it was shown how public libraries consider user privacy a fundamental concern, since it is instrumental in preserving the intellectual freedom of their patrons.<sup>985</sup> In some countries and in several states in the United States, access to library records by government agencies is restricted in view of the interest of privacy of library patrons.

---

<sup>982</sup> See also Section 7.4.7 and Section 8.6.4.

<sup>983</sup> See Solove 2009a, p. 193.

<sup>984</sup> On comparing privacy in the context of libraries with the context of search engines, see Nissenbaum 2010, pp. 194-195.

<sup>985</sup> See cross ref, chapter 7. In particular, the question how and whether new digital library services such as Google Books respect the user's privacy to the same extent as used to be the ideal in the traditional library setting is an important question. For a discussion, see Zimmer 8 September 2009.

It must be noted that public libraries also process more and more user data in the context of electronic library products. User data can also be beneficial in view of the possibility to better serve the library user's needs through better targeted recommendations.

For other institutions such as the press, the processing of user data at the current scale is a new phenomenon. For some political publications subscriber records, if they did exist, were considered a sensitive matter. However, the scale in which electronic online media can and do process data on their readers or viewers is unprecedented. In addition, more and more electronic publishers allow end-users to make contributions as well. Typically, on these users additional data will be collected. The publishing sector at large has yet to show serious signs of reflecting on the possible need to balance the benefits of the processing of user data, such as personalization and targeted advertising, with the intellectual freedom of their users.

In the context of electronic communications service providers, such as Internet access providers, the individual's interests in communications privacy and confidentiality of the contents of communications are both protected by the right to respect for private life in Article 8 ECHR and safeguarded through specific legal norms at the EU and national level. The e-Privacy Directive at the EU level (2002/58/EC) provides that Member States have to ensure the confidentiality of electronic communications through legislation and restricts the permissibility of the processing of traffic and location data relating to communications. The e-Privacy Directive can, like the general EU Privacy Directive, be seen as an instance of the Member States acting under their positive obligation to protect communications privacy in horizontal relations.

In the context of information services such as search engines, there are no sector-specific rules for the protection of the privacy of users. General privacy and data protection laws, the EU Privacy Directive (95/46/EC) in particular, provide the legal framework for the processing of personal data of users by information society services, such as electronic publishers or search engine providers.

Arguably European data protection authorities have been somewhat successful in addressing the collection and processing of user data by search engines through an application of European data protection laws. The Article 29 Working Party's opinion on search engines and data protection clarifies that major search engine providers are subject to general data protection law and it outlines the way in which the EU Privacy Directive guarantees the fair processing of personal data in the context of Web search engines.<sup>986</sup> More specifically, the Article 29 Working Party concludes that search engine user data can typically be considered personal data and can only be processed for legitimate purposes and must be deleted or irreversibly anonymized afterwards.<sup>987</sup> In addition, search engine providers have to inform their users about the processing of their data and their rights to access, inspect, or correct their data.

---

<sup>986</sup> Article 29 Data Protection Working Party 2008.

<sup>987</sup> There has been extensive debate about the threshold for anonymization of search logs to be meaningful. See e.g. Soghoian 2008. More fundamentally, computer scientists have shown that it may be theoretically impossible to anonymize rich data set, since any information that distinguishes one person from the other can be used to re-identify data. See Shmatikov & Narayanan 2010, pp. 22-26.

Notably, the regulatory debate about search engines' obligations under European data protection law has not been settled. The industry has been somewhat reluctant to recognize the application of EU data protection law, arguing on the one hand that most of the data cannot be considered personal data and on the other hand consistently avoiding stating their policies in terms of European data protection rules. The possible resulting lack of compliance will be addressed in the next section.

From the perspective of the right to freedom of expression, an important downside of the European data protection framework as applied to the processing of user data by search engine providers, or other information services such as the electronic press, is that it does not contain provisions that specifically recognize the special nature of the data involved. As searching the Web has become a primary means for members of the information society to find and or access information and ideas, and the provision of the services that satisfy those needs involves the processing of massive amounts of user data relating to one's interests and intentions, this could be reason for reflection on the need to include specific legal norms relating to the processing of data relating to an individual's information seeking and accessing behavior.

The Article 29 Working Party does recognize the importance of search engines as first points of access to information online in its opinion on search engines. First, it notes that search logs contain a footprint of a user's interests, relations and intentions and explains that these logs may be used commercially or become the subject of requests by law enforcement and national security agencies. It then expresses its opinion that:

*"Search engines play a crucial role as a first point of contact to access information freely on the Internet. Such free access to information is essential to build one's personal opinion in our democracy. Therefore, Article 11 of the European Charter of Fundamental Rights is of special relevance because it provides that "information should be accessible without any surveillance by public authorities, as part of freedom of expression and information"<sup>988</sup>*

However, there are no specific rules in the EU Privacy Directive that acknowledge this fundamental concern related to the processing of user data by search engine providers. The way in which the Privacy Directive recognizes the special nature of certain categories of personal data is through the special regime for 'sensitive data'. Article 8 of the Privacy Directive contains a number of special categories of personal data, namely *"data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life."*<sup>989</sup> For the processing of this data stricter rules apply. The processing of these special categories of data is forbidden outside the circumstances listed in Article 8 of the Privacy Directive. If the processing of sensitive data is based on the data subject's consent, such consent needs to be explicit instead of unambiguous (Article 8 (2)(a) Privacy Directive).

---

<sup>988</sup> Article 29 Data Protection Working Party 2008, p.8.

<sup>989</sup> These categories of data are based on Article 6 of the Council of Europe, Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data Strasbourg, 28 January 1981.

While some may argue that search queries reveal sensitive types of information about an individual, the assumption that a direct link exists between a data subject's information access behavior and their proper characteristics as an individual is itself highly problematic. Notably, the Article 29 Working Party did not address this question in its opinion on search engines. It is true that search engine user data, the user's queries and subsequent navigation in particular, could be used to infer certain information about end-users, such as information about diseases, political viewpoints or religious beliefs. A user's search history may reveal an interest into a special heart disease, a certain political party or a particular religion. It may even be possible to statistically predict the ethnicity of the user on the basis of his or her search history and other data accessible to the search engine provider.<sup>990</sup> The possibility that search engines themselves do profile their users and personalize their offerings in these ways could be an additional argument for treating search user data that relates to these special categories as sensitive data themselves.

However, it is problematic to conclude that end-users who search for information about a disease, viewpoint or belief actually have that disease, viewpoint or belief. This may actually go to the heart of the value of privacy and intellectual freedom in the context of access to information and ideas. While the end-user uses the search engine to search for information and ideas, search engines are using the data users leave behind as an opportunity to predict more and more about the end-users' proper characteristics. Consequently, particular patterns in information seeking and accessing behavior lead to the attribution of certain characteristics to the individual. However, even if search engines process a user's data to predict that user's future interests, this does not necessarily mean that the search engine equals someone's interests with the person's proper characteristics. Search engines may actually not be interested at all in identifying their users. They may merely be interested in optimizing the chance of satisfying a user's query as well as monetizing it.<sup>991</sup> Still, this behavior by search engines could have chilling effect on the use of search engine services.<sup>992</sup> To conclude, while search engine user data may reveal information that relates to sensitive types of information, the special regime for the processing of sensitive data does not seem to fit to the processing of search engine user data.

At the same time, the fact that so much information can be statistically predicted on the basis of search engine user data is the reason that search engine providers work hard to obtain more and more user data and are actively exploring new ways to use and analyze them. Search engine providers use this data to improve the quality of the search service, to increase the granularity of their targeted advertising programs, or to develop new and search-related products, such as the search suggest functionality. In addition, there are numerous ways in which the analysis of search engine user data collections, on the scale that a search engine provider like Google has access to, could be used in contexts completely unrelated to access to information and ideas or the optimization of marketing.

---

<sup>990</sup> Recent research at Yahoo! shows that search engines may be able to predict (with some statistical accuracy) ethnic origin on the basis of query analysis. See Weber & Castillo 2010.

<sup>991</sup> See also Stalder & Mayer 2009, p. 112.

<sup>992</sup> Data protection, seen as the right to informational self-determination, can be more generally understood as protecting data subjects interests to develop themselves freely in a democratic and pluralist society. See BVerfGE 65,1 – Volkszählung. For a discussion, see Fischer-Hübner, Hoofnagle et al. 2011.

Google, for instance, has indirectly advocated the benefits of its extensive user surveillance and the research it enables through its Google Trends tools and Google Flu Trends in particular.<sup>993</sup> This service harvests aggregate search data to give an indication of geographic flu activity. It shows how search query analysis can be useful to predict various other events of general public interest and to further our understanding of the world more generally.<sup>994</sup> The following claim by Microsoft researcher Matthew Richardson summarizes the potential of search engine user data collections:

*“With the advent of Web search engines, a new source of data about people and the world has become available. Every time a person queries a search engine, he provides a small window into his life, his interests, and the world around him. Taken as a whole, across millions of users, these queries constitute a measurement of the world and humanity through time.”<sup>995</sup>*

This leads him to advocate against limitations on the retention period of search engine user logs, claiming that the potential to use these data would be severely reduced otherwise. Search engines have an obvious business interest in establishing more and more legitimate purposes for this data, thereby solidifying the legitimacy of the processing of large amounts of user data in the first place.

Most problematically from the perspective of intellectual freedom of end-users, are the signs that government agencies and search engine providers may be coming into agreement on the need for extensive processing of search engine user data for the purposes to investigate or prevent crime.<sup>996</sup> Undoubtedly, specific search engine user data may sometimes be helpful in the context of a criminal investigation. But as soon as the collection of user data by search engine providers becomes instrumental for the purposes of preventing and combating crime, intellectual freedom will be greatly sacrificed.

A similar sacrifice was made by the European legislature when it enacted the Data Retention Directive (2006/24/EC) in the context of electronic communications data. This directive mandates the retention of traffic and location data of publicly available electronic communications networks and services in view of their usefulness for law enforcement and national security agencies. More recently, proposals have been made in the European Parliament to extend data retention obligations to search engine providers in view of the combat of sexual harassment, pedophilia and child pornographic material.<sup>997</sup> In other words, the collection of user data by search engines would become instrumental, *a priori*, to the objectives of governmental agencies, in particular law enforcement and national security agencies.

Libraries in particular have been vocal opponents of attempts to access records of their patrons. Search engine providers have a mixed record. On the one hand, they have sometimes opposed access to their user data by government agencies. In 2006 Google scored a victory in a United States court over government attempt to gain access to individual search records.<sup>998</sup> Google often cites this legal victory

---

<sup>993</sup> See Google Flu Trends, <http://www.google.org/flutrends/about/how.html>. For a discussion see Carneiro & Mylonakis 2009.

<sup>994</sup> See Richardson 2008.

<sup>995</sup> Richardson 2008, p. 2.

<sup>996</sup> See generally, Birnhack & Elkin-Koren 2003.

<sup>997</sup> See McNamee 2010.

<sup>998</sup> See *Gonzales v. Google, Inc.*, 234 F.R.D. 674 (N.D. Cal 2006).

as an example of its effective efforts to protect its users' privacy, also noting it was the only one resisting the attempt to gain access.<sup>999</sup> In general, however, major search providers are responsive to lawful government request to access user records.<sup>1000</sup> Moreover, Google has repeatedly defended its retention of user data by claiming that they could be necessary data for law enforcement agencies to track down criminals.<sup>1001</sup> Google's CEO, Eric Smith, publicly stated that

*"If you have something that you don't want anyone to know, maybe you shouldn't be doing it in the first place."*<sup>1002</sup>

Interesting in this context are the court proceedings between Google and Viacom in the United States about Google's liability for copyright infringement on YouTube. In 2009 Viacom secured a court ruling which ordered Google to hand over the complete log related to its YouTube service, an enormous record of the viewing history of all visitors.<sup>1003</sup> The court dismissed Google's opposition, which was referring to its users' privacy and also specifically relied upon Video Privacy Protection Act (18 U.S.C. Section 2710), with the following statement:

*"[...] defendants cite no authority barring them from disclosing such information in civil discovery proceedings, and their privacy concerns are speculative. Defendants do not refute that the "login ID is an anonymous pseudonym that users create for themselves when they sign up with YouTube" which without more "cannot identify specific individuals" [...] and Google has elsewhere stated:*

*'We . . . are strong supporters of the idea that data protection laws should apply to any data that could identify you. The reality is though that in most cases, an IP address without additional information cannot.'*"<sup>1004</sup>

In other words, the court concluded there was no law that provided for the protection of user records against third party access, and in view of Google's own treatment of and view on user data the Judge dismissed Google defense in view of user privacy. Hence, the protection of privacy for end-users under United States law is dependent on the views and policies of the service provider, and the privacy policies of dominant players in the market do not (unambiguously) provide end-users with reasonable

---

<sup>999</sup> See Cutts 11 June 2007; For a discussion of the way in which Google talks about privacy see Hoofnagle 2009.

<sup>1000</sup> See e.g. Google, 14 March 2007.

<sup>1001</sup> See Soghoian, *An end to privacy theatre*, 2010, p. 194. More specifically, Soghoian refers to the interview by Robert Siegel with Eric Schmidt, Google's CEO, on National Public Radio, 2 October 2009, available at <http://www.npr.org/templates/story/story.php?storyId=113450803>. In this interview Schmidt states that: "[T]he reason we keep [search engine data] for any length of time is one, we actually need it to make our algorithms better but more importantly, there is a legitimate case of the government, or particularly the police function or so forth, wanting with a federal subpoena and so forth - being able to get access to that information."). Former Google employee Edwards describes how after 9/11, Google employees tried to help, voluntarily, identifying the terrorists on the basis of their logs. See Edwards 2011, pp. 233-235.

<sup>1002</sup> See Tate 2009.

<sup>1003</sup> For a discussion and further references, see Hoofnagle 2009.

<sup>1004</sup> See *Viacom International et al. v. Youtube et al.*, 07-CIV-2103 (S.D.N.Y. 1 July 2008).

expectations of privacy.<sup>1005</sup> Seen against this background, European data protection law may better serve the interests of end-users since data protection laws apply independent of the willingness of service providers to take their user's privacy seriously.

At the same time, the EU Privacy Directive itself does not recognize the special nature of the data which is being processed by search engine providers. It does not ensure the intellectual freedom of search engine users by preventing, or setting special standards on, access to this data by third parties once the data has been collected. Lawful access by third parties, including law enforcement and national security agencies, is a matter of the law of the Member States. Until now, no special restrictions have been adopted with regard to the processing of and access to search engine user data. Whereas Article 8 ECHR is an important legal safeguard, the scale of user data processing in the context of search may warrant specific norm setting at the legislative level to ensure the fundamental interest of end-users to search for and access information freely in the online environment.

#### **10.4.3 Accountability for user data processing**

The processing of user data has a real impact on the selection and ranking of search results in response to a query.<sup>1006</sup> Beyond queries and site navigation, one's geographic location and a range of other data are and can be used and interpreted by search engines to tailor the service to specific end-users. In practice there are more questions than answers about the way in which user data is actually being used by search engine providers as search engines tend to give only general information about the actual data processing that is taking place. Hence personalization presents end-users with a problem in terms of their informational autonomy.<sup>1007</sup> Search engines tailor their offering on the basis of some kind of picture of who their end-users are, without actually giving end-users the capacity to determine whether they agree with that picture and the impact on the recommendations of online sources of information it implies.

Considering that European data protection law does apply to the processing of user data, following the Article 29 Data Protection Working Party's opinion on search engines, it is worth looking at the question of to what extent data protection laws could help make search engines more accountable from a user's perspective for their impact on informational autonomy.<sup>1008</sup> This would fit with the nature of data protection law, which does little to prevent the processing of personal data and more often than not is a means to support other fundamental legal interests relating to the freedom of the individual in a technology and information driven society, interests that are different than the right to private life in the strict sense.<sup>1009</sup> More specifically, European data protection law contains a number of provisions relating to the transparency of the processing of personal data by controllers, in view of the possibility to exercise control over data processing by data subjects.

---

<sup>1005</sup> This is, from a legal perspective, not a very big surprise. See Solove2004, pp. 201-209.

<sup>1006</sup> For a discussion, see Section 10.4.1.

<sup>1007</sup> See Feuz et al 2011; Zuckerman 2011; Rotenberg 2007.

<sup>1008</sup> See also Rotenberg 2007.

<sup>1009</sup> See Gutwirth 2002.

According to Article 10 and 11 of the EU Privacy Directive, data controllers have a general obligation to provide information about the processing of personal data that is taking place, the recipients of this data, the purposes of processing, the categories of personal data that have not been obtained from the data subject itself, and the existence of the rights of data subjects. Correspondingly, data subjects have a right to gain access to their personal data (Article 12), “*without constraint at reasonable intervals and without excessive delay or expense.*”<sup>1010</sup> This right of access of data subjects, includes the right to receive

- confirmation as to whether or not data relating to him are being processed and information at least as to the purposes of the processing, the categories of data concerned, and the recipients or categories of recipients to whom the data are disclosed,
- communication to him in an intelligible form of the data undergoing processing and of any available information as to their source,
- knowledge of the logic involved in any automatic processing of data concerning him at least in the case of the automated decisions referred to in Article 15 (1)<sup>1011</sup>

Article 15 of the Data Protection Directive contains a strengthened transparency obligation in view of the fundamental interests of data subjects in relation to automated decision making on the basis of personal data. Article 15 (1) grants data subjects a right

not to be subject to a decision which produces legal effects concerning him or significantly affects him and which is based solely on automated processing of data intended to evaluate certain personal aspects relating to him, such as his performance at work, creditworthiness, reliability, conduct, etc.<sup>1012</sup>

Article 15 (2) contains an exception to this general rule for automated decisions that are

- (a) [...] taken in the course of the entering into or performance of a contract, provided the request for the entering into or the performance of the contract, lodged by the data subject, has been satisfied or that there are suitable measures to safeguard his legitimate interests, such as arrangements allowing him to put his point of view; or
- (b) is authorized by a law which also lays down measures to safeguard the data subject's legitimate interests.<sup>1013</sup>

Finally, consideration (41) of the EU Privacy Directive contains another reference to the right of access of the data subject and the right to know the logic involved in the automated processing of data concerning him, clarifying that its purpose is in particular that data subjects can “*control the accuracy of the data and the lawfulness of the processing.*”<sup>1014</sup> Importantly, the Directive here also notes that on the one hand, these rights of the data subject “*must not adversely affect trade secrets or intellectual*

---

<sup>1010</sup> Privacy Directive (95/46/EC).

<sup>1011</sup> Id.

<sup>1012</sup> Id.

<sup>1013</sup> Id.

<sup>1014</sup> Id.

*property and in particular the copyright protecting the software” and on the other hand stating that “these considerations must not, however, result in the data subject being refused all information.”*<sup>1015</sup>

If one looks at the processing of user data by Web search engines, in practice, the first question to arise is whether the user data that is being processed can be qualified as personal data. It is beyond the scope of this research to address this much debated issue in detail. We will simply rely on the Article 29 Working Party’s conclusion on this subject that search engines generally process a wide variety of user data which qualify as personal data, both in case of authenticated and non-authenticated users.<sup>1016</sup> Notably, this conclusion is not legally tested. It remains to be seen what the proper status of search engine user data is under the data protection directive. Clearly, this will depend on the precise practices of search engines with regard to user data.

It is important to note that the question whether search engine logs qualify as personal data cannot be reduced to the question about the status of IP addresses under the definition of personal data, which is currently one of the issues most prominently debated in this context. Even if IP addresses of end-users, as processed by information service providers, may not in and by themselves be regarded as personal data, search engines tend to process such a wide variety of user data in a form relating to specific individuals that the question about the identifiability of the underlying end-user does not depend upon the status of one specific type of data. More fundamentally, it should be noted that the question about identification of end-users in an online context may need to be addressed a little differently. Providers of free services do rely on a range of individual user data to single out individuals for specific treatment. In the online world the basic identifiers of the offline world, such as name and home address, are typically less meaningful, especially if the service is provided for free.<sup>1017</sup>

Thus when following the Article 29 Working Party, the transparency obligations and the data subject’s right to access and correction typically apply to the user data as processed by search engine providers. This has a number of interesting consequences for the accountability of search engines for their personalized search results, consequences which in practice have not been fully materialized. First, search engines may need to provide users with a lot more information about the data that is being processed and the purposes of such processing of user data. Second, end-users have a right to access all personal data relating to them as they are processed by search engines. Third, data protection law provides data protection authorities with various regulatory means to ensure public oversight as regards the veracity of the information as provided by search engine providers in their privacy policies.

Data controllers tend to comply with the obligation to inform data subjects about the processing of personal data through privacy policies. All major search engines have such privacy policies, including Google, the policy of which will be taken as an example to show a seeming lack of compliance with EU data protection rules.<sup>1018</sup> Notably, the following is meant as an illustration of the possible unfulfilled

---

<sup>1015</sup> Id.

<sup>1016</sup> See Article 29 Data Protection Working Party 2008, pp. 8-9, 27.

<sup>1017</sup> See Article 29 Data Protection Working Party 2007, p. 16.

<sup>1018</sup> Google’s privacy policy is a moving target. The discussion here is based on Google’s privacy policy of 3 October 2010. See Google 2010c.

potential of EU data protection laws in the context of search, rather than a in-depth discussion of the question of whether Google's data processing policy and practices are in compliance with EU law.

Google's privacy policy seems to systematically inform end-users of the ways in which different user data is being collected and processed. But at a closer look, many of the statements are quite vague and may not be in full compliance with the recommendations of the Article 29 Working Party. First, Google systematically refrains from providing information in a form which makes it legally accountable under European data protection law. In particular, it does not use the word 'personal data' once in any of its privacy policies, but instead uses the term 'personal information'. It defines 'personal information' as *"information that you provide to us which personally identifies you, such as your name, email address or billing information, or other data which can be reasonably linked to such information by Google."*<sup>1019</sup> This definition is clearly more restricted than the definition of personal data in European data protection law. In particular, it does not include data which is not provided by the end-user itself, or data which identifies an individual indirectly. By failing to clarify which of the information it processes must be considered personal data the value of Google's privacy policy from the perspective of European data protection law is limited from the start.

Implicitly, Google's choice to use its own definition of 'personal information' makes clear it is ready to defend the position that much of the user data that it processes in the context of its Web search service should not be considered personal data and hence do not implicate data protection obligations. In its main privacy policy it explicitly states that it offers *"a number of services that do not require you to register for an account or provide any personal information to us, such as Google Search."*<sup>1020</sup> In its response to the Article 29 Working Party addressing the opinion on search engines it notes on the one hand that *"Google has always taken the view that IP addresses should be regarded as confidential information that deserves a very high standard of protection"* but on the other hand it states that *"there is significant debate as to whether an IP address should be considered "personal data" for purposes of data protection obligations. Legal analysis of the potential status of IP addresses as personal data should be as rigorous as possible."*<sup>1021</sup>

Second, the information Google provides about the purposes of the processing of user data in the context of its search engine may be too general. One of those purposes, namely *"developing new services"* is overbroad from the perspective of Article 6(1)(b), which provides that new purposes must be compatible. Often the information provided could be much more specific, for instance the information about the purposes of processing IP addresses in view of geographic location of end-users. Google understates the granularity of geo-location information it processes, by stating that an *"IP address can often be used to identify the country from which a computer is connecting to the Internet."*<sup>1022</sup>

---

<sup>1019</sup> See Google Privacy FAQ.

<sup>1020</sup> See Google 2010c.

<sup>1021</sup> Fleischer (Google) 2008.

<sup>1022</sup> Compare the information (implying more granular targeting), for AdWords customers. See Google Adwords Help.

Google also generally informs end-users that it processes personal information in view of “*the display of customized content and advertising.*”<sup>1023</sup> Apart from the fact that much of the information that is mentioned in Google’s policy, such as logs files, links and cookie data, are not defined as personal information, it is unclear which information is used and how it is used to customize the content and advertising in Google’s services. Consider the statement about the processing of query log data, the historic search records of every single user of Google Search over lengthy periods of time. Google explains:

*“Logs data also helps us improve our search results. If we know that users are clicking on the #1 result, we know we’re probably doing something right, and if they’re hitting next page or reformulating their query, we’re probably doing something wrong.”*<sup>1024</sup>

Does this mean that query logs are used to develop an understanding of the interests of specific end-users? If so, what kind of understanding is Google trying to develop? Does Google remember the types of links its individual users follow in its universal search product, such as news, entertainment, video or blogs? Does Google try to develop an understanding of more specific topics of interests of individual end-users to tailor search results and advertising, such as its ‘interest based advertising’ product it developed for its advertising network for Web publishers? And how exactly does this have an impact on the selection and ranking of future search results?

Third, the right to access personal data as processed by Google or any other search engine provider may provide some answers to the pressing questions outlined above. In its privacy policy, Google seems to provide for the right to access personal data. However, its narrow definition of personal information means that for some of the data which would qualify as personal data under European data protection law this policy does not apply. Google does provide access to some of the user data it collects on individual users through its search history service. This service, however, is only available for end-users that use the search engine while logged into a personal account. For these authenticated users, Google offers an overview of some of the data that are stored in connection to an account and the various Google products that are being used.

All in all, European data protection law, at least as interpreted by the Article 29 Working Party in its opinion on search engines, clearly provides a framework of obligations and rights that could help to address some of the accountability deficits in the context of the search engine data processing and its impact on informational autonomy of end-users. Were European Data Protection Authorities to seriously enforce the law as they have interpreted it in their opinion on search engines, the result would be that end-users and the public at large would be much better informed about the way in which search engine operate and have an impact on the information we end up finding while using their services.

The recent episode in which German data protection authorities decided to use their legal powers to investigate personal data processing of Google in light of its collection of data for the controversial Street View service shows how effective an actual audit can be in ensuring compliance with the law.

---

<sup>1023</sup> See Google 2010c.

<sup>1024</sup> Google Privacy FAQ.

Through the audit it was concluded that Google had been collecting and storing information about and communications sent over personal wireless networks of European citizens.<sup>1025</sup> Subsequently, Google has been fined, has had to stop these practices and destroy the data collected. The search engine market may need to be addressed in a similar manner to ensure full compliance with data protection law. The public oversight as provided for in the data protection framework may also be a solution for the problem that full disclosure of data processing practices and purposes may infringe on the trade secrets and intellectual property rights of search engine providers or be harmful for end-users, in view of the possible manipulation of search results upon full public disclosure. Data protection authorities may present the findings pursuant to an audit of user data processing in the context of search engine operations to the general public without having to disclose specific trade secrets or infringe on the legitimate business interests of search engine providers in other ways.

### **10.5 Conclusion**

This chapter has addressed a number of regulatory issues relating to search engine quality from the perspective of freedom of expression. The first section considered search engines from the perspective of pluralism and diversity. After a short overview on early perspectives on diversity and pluralism in the online information environment and the role of search engines and a discussion of some of the early regulatory responses addressing diversity and pluralism and diversity, a general starting point and a number of concerns were identified and further discussed.

The starting point is that in an online information environment characterized by abundance an analysis of pluralism and diversity should take special account of search engines and selection intermediaries more generally, as they have a large impact on the information and ideas that individuals will encounter. Moreover and related, in this context characterized by abundance an analysis of diversity and pluralism should also emphasize on exposure to information and ideas instead of merely addressing what is available online.

A first possible concern is the consolidation of the search engine market, which could have negative effects on pluralism and diversity. Ultimately, the conclusion is drawn that there is not enough evidence of a negative impact from the current market structure and attention was drawn to a number of arguments that warrant skepticism with regard to claims of such an impact. Arguably, many of the critiques of search engines and their possible bias can be best understood as the debunking of overly optimistic assumptions about the equalizing, democratizing and disintermediating effects of the Internet and the Web. To the extent that such assumptions were simply unrealistic or untenable, such as the idea that search engines would facilitate access to information and ideas equally, these critiques are rather unsurprising, and can hardly serve as a starting point for a further analysis.

To be able to address the question about the impact of search engines on diversity and pluralism properly, much more research will need to be done on the impact of market competition on search engine quality. On the one hand, search engines may end up trying exactly the same approaches in their competition for users. On the other hand, a lack of competition could diminish incentives to innovate on

---

<sup>1025</sup> See Federal Commissioner for Data Protection 2004.

search engine quality for end-users. Other open questions that need to be addressed in this context are the specific impact of dominant players on diversity and pluralism, as well as the way in which other players in the public networked information environment more generally could alleviate possible concerns following from market concentration in the search engine market.

The analysis of the way in which individual market participants could have an impact on diversity and pluralism followed three perspectives: the end-users, the information providers and the search engine itself. The analysis shows that search engines have a number of incentives to promote diversity of their search results, since it increases the chance one or more of these results fulfill the information needs of their end-users. At the same time, the search engine business model implies that search engines may be particularly interested in optimizing their offering in view of particular information needs of end-users of a commercial nature, since this optimizes their attractiveness as a marketing platform for advertisers. This could lead to a reduction of diversity of search results that satisfy other interests of end-users. Personalization of search results could become problematic from the perspective of diversity and pluralism, but the question whether that is actually the case depends on the values and principles underlying such personalization and warrants further research. On a more practical level, search engines could be more forthcoming about the value they attribute to diversity and how this impacts on their decisions of how to select and rank search results for end-users.

When focusing on information providers, most problematic is the impact that certain information providers could have on the diversity of search results and the overall robustness of the search medium. Because of the reliance on third party signals, the opportunities for the optimization of search result rankings for information providers and the editorial model of search more generally, some information providers may be effectively pushing legitimate and valuable sources of information out of the end-users' view. Remarkably, this is to some extent a practice endorsed by market leader Google. Search results can be expected to be biased towards information providers with sufficient financial and or organizational means to participate in the ongoing competition for favorable rankings, which can be expected to have a negative impact on diversity and pluralism.

From the perspective of end-users it is clear that search quality to a considerable degree has to be shaped by the users themselves. Due to the interactive nature of search engine services, the quality of the search experience in terms of the diversity of information and ideas users are presented with will depend upon their knowledge and user sophistication and critical engagement. The choice of the specific service, the knowledge of different languages, and the ability to formulate and reformulate effective search queries and use advanced search options all have a considerable impact on the quality of the search experience. Hence, the control of end-users over the search process warrants special attention. It is clear from empirical research about search engine users that search engine quality depends on the skills the level of education and background of specific end-users.

Some of the questions relating to search engine quality and the search engines' advertisement based business model were discussed in more detail in the Section 10.3, which specifically focused on transparency obligations in regard to sponsored search results as well as the way in which optimization of revenue more generally could negatively impact on search engine quality. First, the discussion of the

obligation to delineate organic and sponsored listings showed that this obligation may be too simplistic to deliver real value to the quality of the search medium. This conclusion is based on the observation that search engines do not operate in manner that can provide the independency and information quality guarantees for organic listings that the obligation assumes. The underlying assumption that a parallel can be drawn to traditional editorial media simply breaks down when taking a closer look. Search engine users may be better off if they were informed about the various ways in which search engine results are being influenced through targeted campaigns and optimization strategies of specific information providers.

This points to the need for increased transparency about the way in which search engines operate, transparency which currently is only provided voluntarily. This voluntary nature of transparency in the context of search has some obvious drawbacks. Search engines like Google claim to engage in a number of best practices with regard to search engine quality, but a proper mechanism for verifying those claims is still lacking. On the other hand, transparency obligations could also be considered problematic. On a practical level, they could impact on the ability of search engines to resist manipulation. More fundamentally, strict transparency obligations could be at odds with the right to freedom of expression as applied to the choices of search engines to rank and select in relative freedom. Competition law can be expected to be one of the drivers for increased transparency as regards search engine selection and ranking practices by Google, as several complaints of monopolistic abuse are pending at the European Commission, at national competition authorities in Europe, and in the United States. These investigations may also, at some point, have to come to terms with the question about the editorial freedom of search engine providers to rank and select search results.

The final section focused on two regulatory issues related to the massive processing of user data by search engine providers from the perspective of freedom of expression, privacy and data protection. The analysis showed that concerns underlying the right to freedom of expression can be reason to reflect on the possibility to enact specific legal rules for the processing of user data about information accessing and seeking behavior. The current legal framework does not recognize the special role or status of these activities have in the information society. Currently, the processing of user data by search engine providers is covered by general data protection law, which in the view of European data protection authorities provides the legal framework through which search engines are made accountable for the processing of user data.

Data protection law could be of great value in view of the impact user data has on search engine operations. Increased personalization in combination with a lack of transparency about end-user modeling poses a real threat to informational autonomy of end-users. Data protection law contains a number of provisions which guarantee transparency about the processing of the user data that can be considered personal data. If data protection authorities are right in their interpretation of how European data protection law applies to search engines and will prove more successful in imposing this view on the market this framework could also be instrumental in restoring some balance between end-users and search engine providers.