Ethics and Information Technology (2008) 10:233–242
DOI 10.1007/s10676-008-9177-3

Search engines and the public use of reason

Dag Elgesem
Department of Information Science and Media Studies, University of Bergen, P.O. Box 7802, 5020 Bergen, Norway
E-mail: dag.elgesem@infomedia.uib.no

Abstract. How should the policies of search engines and other information intermediaries be ethically evaluated? It is argued that Kant's principles for the public use of reason are useful starting points for the formulation of criteria for such an evaluation. The suggestion is, furthermore, that a search engine can be seen to provide a *testimony* to the user concerning what information that is most relevant to her query. This suggestion is used as the basis for the development of a broadly Kantian account of a rational searcher. It is argued that the search engine companies are morally required to publish their information policies and act in accordance with them but given the threat of search engine spam they can have no obligation to publish the details of their algorithms.

Key words: bias, information policy, Kant, reason, search, search engines, testimony

Introduction

Search engines have become the central entry points for access to information on the web. *Google* and the other commercial search engines are thus increasingly powerful players in the new information ecology. My focus here is on the power of search engines in relation to their users. There is an imbalance in power because the criteria that the search engine uses to rank its results are not transparent to the user. The question I want to discuss in this paper is which moral obligation – if any – this power puts on the search engine companies. It might be suggested that the search engines have no moral obligations to their users at all because they are private companies, offering a service free of charge, and that the user can switch to a different search engine with no cost. But this cannot be right, I think. Millions of people use search engines every day to search for information.[1] Such tools have become indispensable to the work of journalists, for example.[2] Search engines have moral obligations to their users because users do in fact trust them as providers of access to the content on the web.[3] What is not clear, however, is the scope of these moral obligations.

Consider the question of paid hits in search engines. *Google* and most other search engines maintain a clear separation between so called organic results and paid hits, i.e. ads. On the *Google* result page, for example, the organic hits are listed in the middle of the page and the paid hits in a column on the right side. Sometimes there are ads also on the top of the page. This is a helpful way to divide up the results from the user's perspective. But I don't think *Google* or the other search engine companies are morally required to separate their results in this way. In the late 1990s there the company *GoTo* launched a search engine based completely on only paid hits: the best ranks on the result page were awarded to the

[1] According to a recent survey of American Internet users more than 50% of those online use search engines on a daily basis and almost 90% use search engine from time to time. Deborah Fallows. Search Engine Users. Internet Searchers are Confident, Satisfied and Trusting – but They are also Unaware and Naïve. PEW/INTERNET Report, 2005, p. 1.

[2] See Vinzenz Wyss and Keel, Guio. Google as Trojanisches Pferd? Konsequenzen der Internt-Recherche von Journalisten für die journalistische Qualität. In Die Macht von Suchmaschinen. Edited by Marcell Machill and Marcus Beiler. Herbert von Halem Verlag, 2007, pp. 143–162.

[3] See Deborah Fallows. Search Engine Users. Internet Searchers are Confident, Satisfied and Trusting – but They are also Unaware and Naïve. PEW/INTERNET Report, 2005.

highest bid in an auction.[4] *GoTo* failed to attract many users and must be considered a failure as a search service. This is not surprising because a search engine based on paid hits cannot be trusted to return the results that are the most relevant to the user's query. However, I cannot see that there is anything *morally* wrong with *GoTo*'s strategy as long as the users were informed about the basis for the ranking.

It was a great model for selling ads on the web, however, and *GoTo* was thus later acquired by *Google* who now uses their auction model to sell the ads that appear on the right hand side of *Google*'s first page of results. But at the same time *Google* learned from the failure of *GoTo* that users do not want a search engine where the organic results and the paid hits are mixed. Hence, *Google* chose as their strategy to have a clear separation between these different kinds of results. The design of *Google*'s pages is based on this separation and this transparency is probably one part of the reason for *Google*'s success. But in such a situation, where *Google* tell their million of users that there are no paid hits mixed up with the organic results, and the users have to trust this when using the search engine, it *would* be morally wrong of *Google* to allow paid hits among the organic results anyway. There is no reason to believe that this actually happens because such a practice would go against *Google*'s own business model. But my point is that it would also be a deceptive practice and hence morally problematic.

It has been suggested that *Google* and the other the big search engine companies are morally required to do more than this. In particular, some authors have argued that the search engines ought to publish the details of their algorithms and their criteria for the ranking of the results.[5] This suggestion raises a number of interesting problems pertaining to the ethical evaluation of information services. The problem is that if all the details of the algorithms for the indexing and ranking of results were made public, this would dramatically increase the problem of spam, generated by web masters trying to manipulate the ranking to get their sites higher on the list of hits. So, even if the publication of the algorithms would make the basic logic of the search engines more transparent, the result would be that they end up being heavily biased. Hence, users would end up with less useful services.

The general algorithm of some search engines, notably *Google*'s PageRank, is of course well known. But *Google* uses in addition a large number of additional parameters in the indexing and ranking of web pages, which are not known in detail and which are probably changing.[6] Hence, the question is whether *all of* the details ought to be made public.

Witten et al.[7] sum up the dilemma nicely in their discussion of the search engines' use of secrecy to fight spam.

> Users have a legitimate reason to know the recipes that govern the emergence of certain web pages from underneath a mountain of documents. What are you missing if you inspect the first three, or the first 20 pages of the ranking? What if the ranking policy changes? The view of the web that the [search engines] present is arbitrary and changes mysteriously at unpredictable times. We are trapped in a dilemma: users want to know how their information is selected – indeed, they have a right to – but if this information were made public, spam would increase without bound, with disastrous consequences. And, of course, publishing the details of the ranking policy would create far more interest among spammers than among most ordinary users.[8]

The problem concerning the scope and limits of the moral obligations of search engines shows that a more general framework for the ethical evaluation of the practices of the search engine companies is called for. My account has three elements. First, I present and discuss Kant's ideals for the public use of reason which require of social arrangements that they shall not stand in the way of the free use of human reasoning. These principles are highly relevant, I suggest, to the ethical evaluation of search engine policies. Second, I suggest that the function of search engines is to literally provide a *testimony* to the user about what information is available and relevant to her query. The task of the searcher is to evaluate the testimony provided by the search engine using his knowledge in accordance with certain standards of rationality. This forms the basis, thirdly, for an account of what it is for a user to search rationally. By way of conclusion I draw on the Kantian principles for the public use of reason and argue that the

[4] See David A. Vise. The Google Story. Pan Books. 2005, p. 87.

[5] Wolfgang Schultz, Thorsten Held and Arne Laudien. Suchmaschienen als Gatekeeper in der öffentliche Kommunikation. Vistas. 2005.

[6] Tara Calishan and Rael Dornfest. Google Hacks. O'Reilly. 2003, pages 287 ff.

[7] Ian Witten Marco Gori and Teresa Numerico. Web Dragons. Inside the Myths of Search Engine Technology. Morgan Kaufman. 2007.

[8] Ian Witten, Marco Gori and Teresa Numerico. Web Dragons, p. 172.

search engine companies are morally required to make it possible for users to act as rational searchers to as large an extent as possible. The search engine companies are not morally required to publish the details of their algorithms, however, but neither are they morally required not to publish them.

Kant on the public use of reason

Kant is concerned with the use of reason in the public sphere. Because the web has become an important part of this public sphere informational intermediaries like search engines also affect the conditions for the public use of reason. First, these intermediaries provide access to much of the information on the web that serves as the basis for public discussion. Second, they are public services both in the sense that they are available to all and that they are used by everyone to navigate and access information on the web. They have thus become indispensable parts of the architecture of the public sphere. Third, the search engines use the social structure encoded in the web as the basis for their assessment of relevance of information. Via the search engines, the opinions and information generated by other people serve as input to our own reasoning. It is thus clear that search engines play important roles in the shaping of the public sphere and that they affect the conditions for the public use of reason. Kant, in his work on political philosophy articulates a number of principles for the public use of reason. I will argue that these principles are also relevant to the evaluation of search engines as intermediaries in the digitalized public sphere.

Following Onora O'Neill's interpretation,[9] Kant can be seen to have three principles for the public use of reason. The first O'Neill calls the maxim of "self preservation of reason". This principle is closely related to the ideas expressed by the second formulation of the categorical imperative, and is concerned with the conditions for the free exercise of human reason. Human reason should be able to operate according to its own rational principles, and be as free as possible from external constraints. In the essay, "What is Orientation in Thinking?",[10] Kant formulates this independence of thinking by requiring that one "ask oneself, whenever one is urged to accept something, whether one finds it possible to transform the reason for accepting it, or the rule which follows from what is accepted, into a universal principle governing the use of reason".[11] This principle can be used as a basis for criticism of arrangements that undermines the conditions for the free use of rationality.

The second principle is a "maxim of thinking from a universal standpoint",[12] i.e. an obligation to take the point of view of others into your reasoning. The ideal is to take a universal perspective on the subject matter by taking the point of view of other persons into account.[13]

The third principle O'Neill calls the "maxim of consistent thinking", which is a requirement of consistency in practical thinking. For example, we should not make plans with conflicting goals or without the means to realize our goals.

The principles will provide the basis for criticism of policies that puts constraints on the use of human reason that cannot be justified to those concerned. The principles are thus important to the ethical evaluation of information intermediaries. In the discussion of the ethical evaluation of search engines below I will argue that the role of search engines is to provide the searcher with a testimony about what information that is relevant to his search. The criteria for the ethical evaluation will be based on the Kantian ideals for the public use of reason, discussed above. But can these Kantian principles be applied to testimonies of intermediaries?

Kant on testimony

Kant no doubt thinks that testimonies can form the basis of genuine knowledge. He distinguishes between *belief* and *knowledge* by claiming that belief is something that is held to be true, but which is considered by the subject himself to need further corroboration from other sources:

> All *belief* is a conviction of truth which is subjectively adequate but *consciously* regarded as objectively inadequate; it is therefore treated as the opposite of *knowledge*.[14]

[9] Onora O'Neill, The Public Use of Reason. In Onora O'Neill, editor, *Constructions of Reason. Explorations of Kant's Practical Philosophy*. Cambridge University Press, Cambridge, 1989.

[10] Reprinted in Kant. *Political Writings*. Edited by Hans Reiss. Cambridge: Cambridge University Press, 1971/1991a.

[11] Immanuel Kant. What is Orientation in Thinking? In Kant. *Political Writings*. Edited by Hans Reiss, p. 249.

[12] Onora O'Neill, The Public Use of Reason. In Onora O'Neill, editor, *Constructions of Reason. Explorations of Kant's Practical Philosophy*, p. 46.

[13] See Reiss' Postscript, in Kant. *Political Writings*, p. 255.

[14] Immanuel Kant. What is Orientation in Thinking? In Kant. *Political Writings*. Edited by Hans Reiss, p. 244.

Furthermore, Kant also distinguishes *belief* from *opinion*, where *opinion* is an epistemic claim that is based on what the subject thinks is insufficient evidence. But also this can become knowledge if it is supported by information from other sources:

> On the other hand, if something is considered true on objective grounds which are nevertheless consciously regarded as inadequate, and is therefore no more than an *opinion*, this *opinion* can nevertheless eventually become *knowledge* if it is gradually corroborated by further grounds of the same kinds.[15]

The point, now, with regard to our discussion is that knowledge is something that develops on the basis of new information. This new information can come in the form of *testimonies*. This is the case, Kant explains, for example with knowledge of historical facts

> The historical belief that a certain event has taken place (for example, the death of a great man, as reported in various letters) *can become knowledge* if the event in question (the funeral, will, etc.) is reported by local authorities. [….]
>
> It is therefore perfectly consistent that something should be considered historically true purely on the strength of testimonies. [….] as in the belief that there is a city called Rome and the fact that someone who has never been there could nevertheless be able to say '*I know*' and not just '*I believe* that Rome exists'.[16]

We see here that Kant uses testimonies from others as perfectly good reasons that can lead to *knowledge*, and not just to *beliefs* or *opinions*. Hence, the maxims for the use of public reason must also apply to testimonies. Of course, testimonies can also give rise to belief and opinions. But the point here is that the maxim for the self preservation of reason as the ideal of thinking for oneself is perfectly consistent, in Kant's view, with using the testimony of others as input to our own reasoning and thus serves as objective grounds of knowledge.

Testimony and the principles of the independent use of reason

It is important to note, however, that the principles for the public use of reason cannot be taken to be algorithms that will always result in true beliefs. To see this, consider the interesting phenomenon of *information cascades*. This phenomenon occurs because we take the judgments and behaviours of others as input to our own reasoning. For example, when we are in a foreign town and are looking for a good restaurant, we will usually prefer a restaurant with a good number of customers to an empty one. It is a reasonable guess that the popularity of the restaurant is an indication of the quality of the food. As this example illustrates, we regularly take the information carried by the choices of others as input to our own reasoning. This is often a rational thing to do, as we saw also that Kant held. After all, most of the things we think we know are based on evidence provided by other people.

But sometimes false beliefs can spread epidemically through such processes. In such cases we can have an informational cascade:

> [A]n *informational cascade* emerges when the information that is implicit in other people's actions is so dominating that a rational successor copies *unconditionally*, i.e. ignores his own information.[17]

In such a cascade, the problem occurs then, when the information provided by the judgment of others becomes so strong that it outweighs any other sources of information. It is important to see that even though a cascade can sometimes involve the spreading of false information, it is not an irrational process. There is a formalized theory of the phenomenon, which shows how cascades evolve under certain conditions, in communication between rational agents. We will not go into this formal account here. For our purposes it is sufficient to illustrate the development of an informational cascade with an example.

On November 29, 2006, the science editor of *The Times* wrote an article entitled "The fish with the most powerful jaws in history". It was a report on the recently excavated fossils of a marine dinosaur. The point of the article was that "the ancient sea monster, known as *Dunkleosteus terrelli,* could bring its fangs together with a force of almost 5,000 kg (11,000 lb), making it almost four times more powerful than *Tyrannosaurus rex*".[18] The article was based on a scientific paper published in the journal *Biology*

[15] Immanuel Kant. What is Orientation in Thinking? In Kant, *Political Writings*, p. 244.

[16] Immanuel Kant. What is Orientation in Thinking? In Kant, *Political Writings*, p. 244.

[17] Jack Hirschleifer. The Blind Leading the Blind: Social Influences, Fads, and Cascades. In Matiano Tommasi and Kathryn Ierulli, editors, *The Economics ofHuman Behavior*, p. 191. Cambridge University Press, Cambridge, 1995.

[18] Mark Henderson. The Fish With the Most Powerful Jaws in History. *The Times*, 29.11.2006. http://www.timesonline.co.uk/tol/news/world/article653453.ece. Retrieved July 16, 2007.

Letters and written by researchers from a well-known research institution.

Later on the same day, the Norwegian newspaper *Aftenposten* reproduced the story.[19] The article was based entirely on the article in *The Times*, and focused on the same point: a new creature has been discovered that had the most powerful bite found on any animal in history. The *Aftenposten* article mentioned the names of the researchers and their institution to give credibility to the report. I read the article in *Aftenposten* and told my son in the evening that a creature had been found that had stronger jaws than *T-Rex*.

It turned out, however, that the editor of *The Times* had made a mistake in the calculation of the power of the jaws of *Dunkleosteus terrelli*. He had apparently overlooked that the power of the creature's jaws had been given in *newtons*, not kilograms, in the scientific paper. Hence, the comparison with the measure of the power of *T-Rex'* jaws in kilograms was flawed. This was made clear by a Norwegian paleontologist who was interviewed about the case. In fact, the jaws of the *Dunkleosteus terrelli* had power comparable to that of a hyena.

But was I *irrational* when I told my son that a creature had been found that had more powerful jaws than *T-Rex*? It seems clear that I was not irrational. I had the information from a credible source that I knew had it from an even more credible source. Furthermore, the source of the information appeared to be good scientist with first-hand knowledge of the matter. And, indeed, my belief would be counted as rational on the theory of rational choice. Still, there was an informational cascade involved because the information generated by the chain of testimonies outweighed all other information I had.

This brings us back to Kant's maxims for the use of public reason. They are also normative principles that will not always lead to true beliefs. To see this, consider the principle of independent reason. It invites us to consider, "whenever one is urged to accept something, whether one finds it possible to transform the reason for accepting it, or the rule which follows from what is accepted, into a universal principle governing the use of reason".[20] As noted above, this principle does not preclude the rational use of testimonies from others as input to our own reasoning. Furthermore, it would not have stopped me from accepting the story about *Dunkleosteus terrelli*, because my reasons for accepting it make perfectly good sense as a universal principle. In a complex information society, with a highly developed division of intellectual labor, we have no option but rely on information from sources that are usually trustworthy.

At the same time, Kant's other two maxims for the use of public reason, i.e. the duty to "think from the standpoint of everybody else", and the principle of consistency in practical thinking, would not have prevented me from accepting as a fact the false information about the *Dunkleosteus terrelli*.

Kant's maxims for the public use of reason must therefore be taken to be *normative* principles and not recipes that will always secure true knowledge. Rather, they should be seen as standards against which we can evaluate information policies. These principles should be used, then, as bases for improving the conditions for the independent use of reason.

Other commentators have also suggested a normative interpretation of these principles. Reiss explains:

> If these principles are observed in political discussion, the public use of reason can provide a yardstick by which we can judge whether and to what extent political arrangements are moving towards a just political society. Moreover, it is the only way which citizens can properly debate questions in accordance with standards set by reason and determine whether a state is governed in accordance with (just) principles of politics.[21]

In the same way, I suggest, the principles can be used to ethically evaluate information policies. In the following I will argue that they can also be used as the basis for clarifying the moral obligations of search engine companies.

The testimonies of search engines and the rational searcher

In this section I will develop an account of the rationality of search behaviour. My suggestion is that when a search engine presents a result, this can be seen as a kind of *testimony* about what are the most relevant set of pages that can be found on the web, given the user's query. The situation has the *structure* of a testimony and, following Goldman, is characterized by four stages: "(1) discovery, (2) production and

[19] Hans Marius Tonstad. *Tidenes kraftigste bitt*. *Aftenposten*, 29.11.2006. http://www.aftenposten.no/viten/article1549964.ece. Retrieved July 16, 2007.

[20] Immanuel Kant. What is Orientation in Thinking? In Kant, *Political Writings*, p. 249.

[21] Postscript in Kant, *Political Writings*. Edited by Hans Reiss. Cambridge University Press, Cambridge, 1971/1991b, p. 255.

transmission of knowledge, (3) message reception, and (4) message acceptance''.[22] In the case of search engines, the discovery (1) is the indexing of the web pages, the production of knowledge (2) is the ranking of the pages in relation to the query, the message reception (3) is the presentation of the results to the user, and the last stage (4) is the user's evaluation of the results from the standpoint of her informational needs. How should a testimony be rationally assessed? With respect to the testimony of a witness in a court case there are two pairs of questions that a judge has to ask when considering the truthfulness of the witness W's statement S.[23] The first pair of questions concerns the credibility of the witness with respect to the issue at hand.

(1.a.) How probable is it that S is true, given that W testifies that S.
(1.b.) How probable is it that S is not true, given that W testifies that S.

The judge has to consider both of these questions in his assessment of the informational value of the testimony. To answer the questions he has to take into account the witness' interests and motivations, how close he was to the situation, his general credibility, etc.

But the witness' testimony must also be assessed in the light of the independent probability of the content of W's claim:

(2.a.) How probable is it that S is true, independently of S's testimony?
(2.b.) How probable is it that S is not true, independently of S's testimony?

In order to estimate these probabilities the judge has to bring in her general knowledge of the world. But the considerations that are involved in answering the two groups of questions of course interact. A witness has to be a very credible and well informed to convince the judge of the truth of a statement that is considered to be unlikely on independent grounds.

The user's position relative to the results provided by a search engine is similar to the judge's position relative to the testimony of a witness, I suggest. One might think that because a testimony is something that is normally given by humans it is inappropriate to apply this notion to machines.[24] But the framework for the assessment of the credibility of testimonies sketched above can be applied to any source of information. Questions (1.a.) and (1.b.) can be asked with respect to any agent that can provide evidence concerning a state of affairs. Furthermore, it is possible to estimate how reliable a search engine is with respect to different tasks. Of course, a search engine cannot be held accountable in the way a human witness can. But that is not relevant to the analysis offered here. My suggestion is that the informational relationship between the user and the search engine is structurally similar to that which obtains between a judge and the testimony of a witness, even though the situations are not exactly similar. A human witness claims to be telling the truth about a particular state of affairs, as a response to a given question. A search engine, on the other hand, makes the second order claim that the information it presents is the most relevant to be found, given the query. But in both cases it is claimed that something is relevant and accurate relative to the enquirer's question. Hence, the judge and the user can be seen to be confronted with the same type of problem.

I will use this observation as a starting point for the characterization of a *rational searcher*. Like the judge assessing the witness' testimony, the rational searcher has to assess the relevance of the result in the light of both his knowledge of the search engine and of the domain of the search. The central point is that the user of a search engine is dependent on the information of the search service and has no direct access to the information himself. Of course, the user is not completely powerless. She will often be able to assess the relevance of the result relative to her informational needs. In some types of searches, often called ''lookups'', the user is searching for a particular piece of information such that she knows when she has found what she is looking for. An example would be the search for the lyrics of Dylan's ''Like a Rolling Stone''. When the text is found on bobdylan.com, you know that you are done. This is different from search with the purpose of discovering new information about a topic about which you know little – for example the search for information on a disease. In the case of search as discovery, it is much harder for the user to assess the quality of the result. But even in this case, the *rational* user has to engage in a form of triangulation.[25] She has to make an assessment of the result on the basis of (1) her general knowledge about

[22] Alvin I. Goldman. *Knowledge in a Social World*, p. 104. Clarendon Press, Oxford, 1999.

[23] The following draws on the Bayesian analysis developed in great detail by Richard D. Friedman. Route Analysis of Credibility and Hearsay. *The Yale Law Journal*, 96(4): 667–742, 1987.

[24] This objection was raised by one of the anonymous referees of the paper.

[25] Compare the role of triangulation in Donald Davidson's philosophy of mind and language. See the papers in the collection Donald Davidson. Subjective, Intersubjective, Objective. Clarendon. 2001, in particular The Emergence of Thought.

the subject, (2) the formulation of the query, and (3) the content of the information provided by the search engine. The user has to move between these three perspectives in the process of trying to learn from the search. Her willingness to accept the result will vary depending on how much she already knows, how much she trusts the search engine, and how confident she is that the query is a precise characterization of what she is looking for.

These three aspects of the evaluation give rise to three different questions with respect to the search result. First, there is the question about how much information the search engine has. This is the question of its *coverage*, i.e. how much of the relevant part of the web it has indexed. No search engine has indexed all pages on the web. At the moment, *Google* probably has the best coverage.

But, secondly, this question about coverage should be distinguished from what could be called the problem of the *objectivity* of the search engine, i.e. whether it is non-biased and only tries to serve the informational needs of the user to the best of its abilities. The question is whether the engine is designed only with the goal in mind of providing the user with the most informative result. The issue here is about bias: the implicit claim made by a search engine is that the result it presents is the best and most relevant it can come up with, given the query. Users place trust in search engines, and assume that the ranking of the result is not biased by, say, paid hits.

A third question that is important with testimonies is how *useful* the information provided by the witness is. A search engine can have good coverage and be objective, but still be of little use if the questions the user poses are not adequate. A user of a search engine has an informational need that she translates into a query that is submitted to the search engine. The relevance of the result set is evaluated relative to this need. But a user of course also will evaluate the *query* in the light of the result. An irrelevant result will often be an indication that the query itself should be changed. This is why information scientist compares search behaviour to "berry picking"[26]: the user starts by formulating a query as an expression of a need for information, learns from the result, updates her information, and poses a new query in another direction, etc. A rational searcher will not only evaluate the relevance of the result relative to her need for information, but also evaluate her own query with respect to how the search engine works. Hence, if the user gets results that do not seem very relevant, she should not – and probably would not – automatically conclude that the search engine is no good. An alternative interpretation to be considered is that the problem was the formulation of the query.

It is thus analytically important to distinguish three different questions that the user of a search engine, just like the judge in a court case, is concerned with: coverage, objectivity, and usefulness. And objectivity is the key question. Only if the search engine can be assumed to be honest is it possible for the user to learn through experience whether her queries are efficient ways to characterize the information in which she is interested. Conversely, with a biased search engine it could be very difficult to find out whether an irrelevant result is caused by a badly formulated query or by the interests built into the algorithm. The same point holds true with respect to the question of coverage: if the search engine is biased, it is very difficult to know whether the absence of relevant information stems from lack of coverage or is caused by the bias itself. Note that I am not claiming either that it is possible to have a completely unbiased search engine or that with an objective search engine it is easy to find an optimal query. What I *am* claiming is that people have to assume that search engines are objective in the sense that these engines try to be consistent with their own stated policies. Most people would feel that their trust was betrayed, I think, if it were discovered that *Google* allowed paid hits in their result sets. This is an indication that people actually assume that the search engines to be objective in the sense I have indicated.

The central point that emerges from this analysis of search engine as providers of testimony is one with strong Kantian affinities. In the discussion above it was suggested that we could characterize the idea of a rational searcher in terms of the need to employ three kinds of information: 1) knowledge of the search program, 2) knowledge of the search domain, 3) information about the result set. The problem with search engines that are biased, I have suggested, is that this makes it difficult to carry out this process of rational triangulation. Not only does it undermine the user's trust in the search service: it is also difficult to use one's world-knowledge to evaluate the result, as well as to assess the adequacy of the query. Hence, a search engine that is biased in ways not known to the user will undermine her ability to reason about the information provided and thus, to put it in Kantian terms, to undermine the user's freedom to think rationally.

Compare this with Kant's second formulation of the categorical imperative: "Act in such a way that you treat humanity, whether in your own person or in

[26] Marcia Bates. The Design of Browsing and Berrypicking Techniques for the Online Search Interface. *Online Review*, 13(5): 407–424, 1989.

the person of another, always at the same time as an end and never simply as a means".[27] By "humanity" Kant meant the property that makes us human, i.e. our ability to reason rationally. For Kant, to exercise one's practical rationality involves the formulation of a maxim for the action to be undertaken. This process of formulating a maxim requires information about the conditions of the successful exercise of the action. Kant's own examples show that to lie or to give false promises are prime examples of violations of this imperative because the liar, by giving false information, makes it impossible for the victim to exercise his rationality. The problem is that the person who gives false promises acts on a maxim that the other person cannot consistently include in *her* own maxim. A search result that is biased in ways not known to the user will similarly provide an input to the agent's reasoning that he cannot rationally process.

Discussion

On the analysis developed here search engine companies are only morally required not to undermine the rational user's search strategies. Others have argued that the moral duties of the search engine companies go farther than this. In a recent paper[28] Introna and Nissenbaum have argued that the details of the algorithms of the search engines should be made public. Their central argument for their suggestion is that search engines are *public goods*.

> [S]earch engines […] raise political concerns not simply because of the way they function, but because the way they function seems at odds with the compelling ideology of the Web as a public good. This ideology portrays the fundamental nature and ethos of the Web as a public good of a particular kind, a rich array of commercial activity, political activity, artistic activity, associations of all kinds, communication of all kinds, and a virtually endless supply of information. In this regard the Web was, and is still seen by many as democratic medium that can circumvent the hegemony of the traditional media market, even government control.[29]

The implication of this perspective on the web as a democratic technology is that it is politically problematic if the public services for accessing the web are systematically biased: "If search mechanisms systematically narrow the scope of what seekers may find and what sites might be found, they will diminish the overall value of the Web as a public forum, as well as a broadly inclusive source of information".[30]

The claim that the major search engines are public services is in my opinion clearly correct. They serve as central entry points for people's access to information on the web. It is also easy to agree with the suggestion that search engines should be thought of in political terms. Still, there are important objections to the suggestion that search engine algorithms should be made public. This is the serious risk that the publication of the details of the algorithms itself would create new bias. The problem is that web masters try to manipulate the search engines to improve their rank. And search engine spamming – spamdexing – is already a huge problem. Spam can take a number of forms.[31] One classical method is 'link-farming', i.e. the creation of pages that link to your own target page to manipulate the search engines to give it a higher ranking. There is currently a war-like situation between the search engines and the so-called search engine optimizers. As the search engines continuously try to change their algorithms to prevent spamming, the spammers tries to find out how to manipulate the service. In 2004 it was estimated that as much as 10–15% of the web pages was spam.[32]

In 2003, it was even claimed by some bloggers that the central Google algorithm, PageRank, no longer worked because of spamming. It was claimed that

> …the algorithm was no longer useful because bloggers and CEOs had learned too much about it and had, in effect, changed the nature of the Web. Since PageRank is based on an optimistic assumption that all links are conceived in good faith with no ulterior motives, an assumption that no longer holds, then PageRank is no longer useful.[33]

Fortunately, PageRank is not dead, but the problem of spamdexing has caused a lot of effort to prevent manipulation. And the major search engines will in

[27] Immanuel Kant. *Grounding for the Metaphysics of Morals*, p. 36. Hackett Publishing, Indianapolis, 1981/[1785].

[28] Lucas Introna and Helen Nissenbaum. Shaping the Web: Why the Politics of Search Engines Matters.

[29] Lucas Introna and Helen Nissenbaum. Shaping the Web: Why the Politics of Search Engines Matters, p. 178.

[30] Lucas Introna and Helen Nissenbaum. Shaping the Web: Why the Politics of Search Engines Matters, p. 178.

[31] For an introduction and overview over methods of search engine spam, see Ian Witten, Marco Gori and Teresa Numerico. Web Dragons. Chapter 5, 'The Web Wars'.

[32] Ian Witten, Marco Gori and Teresa Numerico. Web Dragons, p. 164.

[33] Amy N. Langville and Carl D. Meyer. *Google's PageRank and Beyond*, p. 140. Princeton University Press, Princeton, 2006.

fact respond to serious attempts to manipulate the ranking by removing the site from its index. There is thus every reason to believe if the indexing algorithms were made public, this problem would be so significant that the usefulness of the search engines would be considerably reduced. Nissenbaum and Introna mention this problem – only to dismiss it:

> Search engine operators are loath to give out details of their ranking algorithms for fear that spammers will use this knowledge to trick them. Yet, ethical Web-page designers can legitimately defend a need to know how to design for, or indicate relevance to, the ranking algorithm so that those who search find what is genuinely relevant to their searches.[34]

This response does not give the problem the attention it deserves. The main problem with the response is that the manipulation of indexing is construed as a problem only for the search engine companies, i.e. "that spammers will trick them". This is of course also true. But this is not the central problem if we see search engines as public services, as Introna and Nissenbaum of course do. Then the serious problem is that biased search engines provide less useful services to *the public*.

This shows, I believe, that the search engine companies cannot be morally required to publish the details of their algorithms. However, I believe Introna and Nissenbaum are right in insisting on the value of open standards on the web. Their arguments lend support to initiatives like the Wikia Search-project which tries to develop an open source search engine on the model of Wikipedia. The initiator of this project is Jimbo Wales, the founder of Wikipedia. The motivation for this project is exactly to increase the transparency of search engines:

> Search is part of the fundamental infrastructure of the Internet. And, it is currently broken.
>
> Why is it broken? It is broken for the same reason that proprietary software is always broken: lack of freedom, lack of community, lack of accountability, lack of transparency. **Here, we will change all that**.[35]

The idea is thus to develop a search engine that is based on human contributions to the ranking of search engine results, using an open algorithm. The project is still in its initial phases, and the search engine performs very badly compared to the commercial search engines. It is an open question whether the project will succeed,[36] but of course it would be a good thing if it did. But still this does not imply that all the other search engine companies are morally required to operate on the same model: the enforcement of such a requirement would mean that the quality of the existing search services would go down because if spam.

This would be a loss not only for the searchers but also for small players on the web because on a web with more spam the findability of a web page would depend even more than it does today on the resources the web site have available for search engine optimization. Hence, smaller and less resourceful web sites could end up being more marginalized. This would in my opinion also be a less democratic web.

One suggestion might be that the algorithms of search engines could be controlled by a group of experts on behalf of the public. This would lift the veil of secrecy while voiding the problem of spam, it might be argued. Such an arrangement could perhaps make sure that the users are not deceived and that paid hits are not secretly allowed. But note that this would not meet the standards of openness advocated by Wales, and Introna and Nissenbaum because the details of the algorithm would still not be public knowledge. The arrangement would not provide information that could help users of the web to more rationally exploit its resources.

Conclusions

Two points emerge from this discussion. First, the big search engine companies are morally required to make their policies known to their users and to follow them. More generally, they are morally required to act in such a way that they do not undermine their user's efforts to act as rational searchers. These moral requirements arise because of the important roles the big search engines play as providers of access to information on the web, and hence as contributors to the public use of reason. This conclusion is supported by considerations that are broadly Kantian in nature and seems to follow if we apply his principles for the

[34] Lucas Introna and Helen Nissenbaum. Shaping the Web: Why the Politics of Search Engines Matters, p. 174.

[35] Jimbo Wales. Statement on the Wikia Search Project Home Page. http://search.wikia.com/wiki/Search_Wikia. Retrieved July 16, 2007.

[36] In an interview with New Scientist, open source programmer Ben Laurie says: "By publishing its search algorithm, it's going to be pretty obvious to spammers how to get to the top of the search hits, risking a huge spamfest. Some genius might come up with algorithms that, despite being published, are resistant to that. But it strikes me as unlikely." Marks, Paul. Open-source search engine gangs up on Google. New Scientist. May 2007.

public use of reason to informational intermediaries like search engines.

Second, on the analysis offered here the search engine companies are not required to publish the details of their algorithms. Two kinds of arguments are offered for this conclusion. One reason is that the publication of the algorithm would make users worse of in terms of information. Hence, it is very likely that the flood of spam would effectively undermine the searchers rational strategies. However, it is not clear that the fact that the algorithms are not made public do undermine the user's rational search strategies.

Acknowledgements

Parts of this paper were presented at the workshop 'Kant Revisited in Light of New Technology', held at NTNU, Trondheim, on March 19 and 20, 2007, and at ETHICOMP2007 conference, held in Tokyo, March 27–29, 2007. I have benefited from the discussion and comments from the participants on both occasions. In particular, I am grateful to Charles Ess and Bjørn Myskja for their comments.

References

Marcia Bates. The Design of Browsing and Berrypicking Techniques for the Online Search Interface. *Online Review*, 13(5): 407–424, 1989.

Tara Calishan and Rael Dornfest, *Google Hacks*. O'Reilly, Sebastopol, CA, 2003.

Donald Davidson, *Subjective, Intersubjective, Objective*. Clarendon Press, Oxford, 2001.

Deborah Fallows. Search Engine Users. Internet Searchers are Confident, Satisfied and Trusting – but They are also Unaware and Naïve. PEW/INTERNET Report, p. 1, 2005.

Richard D. Friedman. Route Analysis of Credibility and Hearsay. *The Yale Law Journal*, 96(4): 667–742, 1987.

Alvin I. Goldman, *Knowledge in a Social World*. Clarendon Press, Oxford, 1999.

Mark Henderson. The Fish with the Most Powerful Jaws in History. *The Times*, 29.11.2006. Retrieved July 16, 2007 from http://www.timesonline.co.uk/tol/news/world/article653453.ece.

Jack Hirschleifer. The Blind Leading the Blind: Social Influences, Fads, and Cascades. In Matiano Tommasi and Kathryn Ierulli, editors, *The Economics of Human Behavior*. Cambridge University Press, Cambridge, 1995.

Lucas Introna and Helen Nissenbaum. Shaping the Web: Why the Politics of Search Engines Matters. *Information Society*, 16: 169–185, 2000.

Immanuel Kant. *Grounding for the Metaphysics of Morals*. Hackett Publishing, Indianapolis, 1981/1785.

Immanuel Kant. *Political Writings*. Edited by Hans Reiss. Cambridge University Press, Cambridge, 1971/1991a.

Immanuel Kant. What is Orientation in Thinking? In Hans Reiss, editor, *Kant Political Writings*. Cambridge University Press, Cambridge, 1971/1991b.

Amy N. Langville and Carl D. Meyer, *Google's PageRank and Beyond*. Princeton University Press, Princeton, NJ, 2006.

Marcell Machill and Marcus Beiler, editors, *Die Macht von Suchmaschinen*. Herbert von Halem Verlag, Koln, 2007.

Paul Marks. *Open-Source Search Engine Gangs up on Google*. New Scientist, 2007.

Onora O'Neill, *Constructions of Reason. Explorations of Kant's Practical Philosophy*. Cambridge University Press, Cambridge, 1989.

Wolfgang Schultz, Thorsten Held and Arne Laudien. *Suchmaschienen als Gatekeeper in der öffentliche Kommunikation*. Vistas, 2005.

Hans Marius Tonstad. *Tidenes kraftigste bitt*. *Aftenposten*, 29.11.2006. Retrieved from July 16, 2007, http://www.aftenposten.no/viten/article1549964.ece.

David A. Vise. *The Google Story*. Pan Books, Basingstoke and Oxford, 2005.

Jimbo Wales. Statement on the *Wikia Search* Project Home Page. Retrieved from July 16, 2007, http://search.wikia.com/wiki/Search_Wikia.

Vinzenz Wyss and Guio Keel. Google as Trojanisches Pferd? Konsequenzen der Internt-Recherche von Journalisten für die journalistische Qualität. In: Die Macht von Suchmaschinen. Edited by Marcell Machill and Marcus Beiler, pp. 143–162. Herbert von Halem Verlag, Koln.

Ian Witten, Marco Gori and Teresa Numerico, *Web Dragons. Inside the Myths of Search Engine Technology*. Morgan Kaufman, Amsterdam, 2007.