Source Orientation in Information Retrieval:

Effects on Perceived Similarity and Recognition

Memory of Search Engine Results

David R. Danielson

Stanford University

Clifford I. Nass

Stanford University

Corresponding Author's Contact Information:

David R. Danielson
Department of Communication
Stanford University
davidd@stanford.edu

ABSTRACT

We report on two experiments indicating that information seekers (a) perceive results returned by a single search engine to be more similar to one another than if those same results are gathered from multiple, distinct search engines, and (b) process results differently in the two contexts. In Experiment 1, participants (N=32) evaluated the same Web page results ostensibly returned by either a single search engine or by multiple search engines. Participants in the multiple search engine condition perceived the results to be more heterogeneous with respect to information quality, and showed greater recognition memory for content within the results. In Experiment 2, participants (N=32) evaluated the same sets of image results, again ostensibly returned either by a single or by multiple search engines. Consistent with Experiment 1, participants in the multiple search engine condition perceived the results to be more heterogeneous in quality. However, recognition memory for the image results did not differ between the two conditions, providing evidence that increased depth of processing is not a necessary condition for increased perceived heterogeneity of the search results. We discuss implications for source orientation and information utility theories and for what dimensions of quality over which information retrieval systems might be expected to have some degree of control.

CONTENTS

1. INTRODUCTION

Computer users frequently encounter error messages, requests, advertisements, Web pages, and other content in a wide range of information-seeking environments. Who or what users believe to be responsible for these messages may influence how they are processed and responded to (Petty & Cacioppo, 1986). Thus, reasoning about the sources of information in computing environments is a critical activity for users, in large part because the resulting conclusions can have a significant impact on attitudes and behaviors.

Evaluating who or what is responsible for computer-based messages can be especially challenging, however, in part because these messages are often filtered, summarized, manipulated, or otherwise presented in ways that to some degree mask (or overwhelm cues that indicate) the content's origin. Systems that perform a gatekeeping role, such as search engines, are one prime example of such filtering. Although the sources such filters point to are ultimately responsible for their own information, the filters themselves are often salient throughout the interaction.

The focus of this paper is the extent to which search engines influence how query results are perceived and processed. In particular, we investigate the extent to which information seekers orient towards search engines as relevant in making predictive judgments about how heterogeneous returned content is likely to be. We ask two fundamental questions: (1) Will users perceive search results returned by a single search engine to be more similar to one another in terms of information quality than those same results returned by multiple, distinct search engines?, and (2) Will users process results

gathered from multiple, distinct search engines more deeply than the same results returned by a single search engine? As with gathering information from multiple people, users may expect a greater level of homogeneity in results when gathering information from a single search engine, and consequently may scrutinize such results to a lesser degree than the same content from multiple search engines.

In the remainder of this section we will: (1) discuss the theoretical framework of source orientation—which broadly asks, "Who or what do users perceive to be the sources of content?"—in relation to information-gathering in electronic environments; (2) review literature relevant to the hypotheses that users will judge results from the same search engine to be more homogenous, and will process these results less deeply, than if they were gathered from multiple search engines; and (3) briefly discuss relevant background theory regarding the two primary dependent variables in this study: perceived similarity of results and recognition memory.

1.1. Source orientation

Although it is clear that users rely upon some notion of source in evaluating online information, relatively little work has addressed how computer users determine what constitutes a source worthy of attitudinal attributions. In reviewing and explicating the concept of source with respect to computing systems, Sundar and Nass (2001) point out that several entities may be viewed by users as psychologically relevant sources, including the physical machine itself (Nass & Steuer, 1993), which is in some sense the first layer that obfuscates the human programmers and information providers themselves. Although these computational layers are often not the originators of the information they

present, computer users are nevertheless prone to orienting towards such layers as though they were (Sundar & Nass, 2000), due in large part to their relative proximity to the user. Thinking about non-proximate sources is often too cognitively demanding; it is much easier to respond to the messenger as though it were the source (Reeves & Nass, 1996; Lee & Nass, 2004; Nass & Brave, 2005).

Determining the source of online information is often a difficult task. In many cases, the source of Web content is ambiguous or not present at all (Burbules, 2001; Eastin, 2001), and in others it is present but not easy to ascertain (Toms & Taves, 2004). This problem is accompanied by Web users' generally high reliance on source identity as a criterion for assessing information quality (Rieh & Belkin, 1998). In a massive information-seeking environment such as the Web, users are likely to frequently encounter unfamiliar sources. As a result, they may rely on familiar gatekeepers and other information filters in evaluating the unfamiliar content, rather than relying exclusively on the novel sources that are actually responsible for the content.

Information filters

The issue of user source orientation becomes particularly critical on the Web, where personalized content filters, portals, search engines, and "meta" search engines present layer upon layer of information filtering. This study is concerned with whether and how users orient towards search engines in particular as psychologically relevant information sources, independent of those search engines' established reputation or a user's previous experience with them. That is, this study demonstrates that assessments of results from distinct search engines follow a general information-seeking heuristic: information from a set of distinct sources is likely to be more heterogeneous than information from a single

source.

The context of information retrieval presents a particularly interesting case for investigating source orientation effects. On the one hand, search engines are potentially good candidates for invoking user source orientation because they fulfill a traditionally human role (Reeves & Nass, 1996): filtering or gatekeeping is a computational analogue to what was once exclusively human work (Sundar & Nass, 2001). On the other hand, although search engines are proximate and salient prior to and often after interaction with the information objects to which they point, the proximate source shifts once users visit particular result pages. At this point evaluative processing of Web information sources presumably follows typical patterns (e.g., Rieh, 2002). To the extent that proximity will influence a search engine's role in shaping user attitudes towards its results, the effect may be expected to occur largely prior to interaction with the information objects—that is, during predictive assessments of the search engine results.

1.2. Predictive assessment

Information seeking frequently involves predictive judgments about the likelihood of a navigational decision leading to desired information. Predictive judgments about an information object are made prior to encountering the object, in contrast to evaluative judgments based on directly observed characteristics of the object (Hogarth, 1987). In making predictive judgments, information seekers must rely upon features of their present environment that might help indicate the likely characteristics of the distal object, because movement to a new environment—or to a new document, in the case of Web navigation—carries associated costs (Pirolli & Card, 1999). Web information seekers

frequently use a distal document's hyperlinked surrogate (in the form of an underlined text snippet) to predict its likely usefulness (Rieh, 2002), but the extent to which users also rely upon contextual factors in making such judgments has not been systematically investigated. The contextual factor of interest in this study is the number of search engines providing the set of results to be assessed by the user.

Source orientation towards search engines may have a significant impact on predictive assessments of information quality for two reasons: (1) results from the same search engine may be perceived as forming a coherent group, causing users to evaluate each result differently than if it were not part of the perceptual group, and (2) different search engines (even if the user is not familiar with them) might be expected to produce results with more varied information utility than those produced by the same search engine. The existing related literature is mixed in its support for the hypothesis that these reasons will result in assessment differences (due in part to particular characteristics of Web search engines); this study sets out to investigate whether such differences occur.

Perceived entitativity

Category membership of an object is both a powerful cause of human inductive reasoning and often a preferred cue to rely upon in making inferences about an object's likely characteristics (Heit, 2001). It is unclear to what extent users categorize information objects in the same sorts of ways they categorize individual people, and so existing work on social categorization may be of limited use in understanding the causes of users' perceptual groupings of Web pages. However, there is some evidence that hyperlink structures (such as which pages are linked to which) are used in making assessments of both Web pages and the real-world entities behind them. Stewart (2003),

for example, found hyperlinks from trusted sources to influence level of trust in the organization linked to. Moreover, credibility assessments are often partially based upon how Web documents are interconnected (Fogg, Marshall, Laraki, Osipovich, Varma, Fang, Paul, Rangnekar, Shon, Swani, & Treinen, 2001).

Stewart (2003) suggests that Web pages linked to one another may be perceived as high in *entitativity*, a term originally used by Campbell (1958) to describe the extent to which individual persons form a coherent group. We might suggest, then, that Web page results from a single search engine may be perceived as more entitative than the same collection of results collected from multiple search engines. Here, the existing theory and research is mixed. On the one hand, if users do orient towards search engines as psychologically relevant sources, the results share a common reference. On the other hand, search engine results lack a common antecedent to perceived entitativity, namely interaction amongst group members (Lickel, Hamilton, Wieczorkowska, Lewis, Sherman, & Uhles, 2000). Search engine results coexist in what could be taken by users to be a relatively idiosyncratic grouping, based upon relevance to a search query. While such a grouping typically entails some level of topical similarity, similarity among a set of objects tends not to be enough on its own to achieve perceived entitativity (Yzerbyt, Judd, & Corneille, 2004). More importantly with respect to this study, shared relevance to a particular search query occurs for results returned by a single search engine and multiple search engines alike, and thus cannot be the cause of any assessment differences between these two information-gathering contexts.

Information utility

An important difference between processing information from multiple sources rather

than one source comes from the psychological literature on persuasion. Harkins and Petty (1981, 1987) have shown that information from multiple sources receives greater scrutiny than the same information presented by a single source. The researchers explained the result in terms of expected information utility, finding evidence that arguments from different sources are viewed as more likely to represent different perspectives and to come from independent pools of knowledge. They found that the persuasive advantage of multiple sources was eliminated when participants were told that the sources were part of a committee rather than being independent, but was retained when they were told that the committee members held dissimilar perspectives (Harkins & Petty, 1987). The expectation of increased information utility, they claim, leads to increased depth of processing. These findings additionally suggest that the multiple source effect does not arise out of participants' increased attention in reorienting to a new stimulus.

While the work by Harkins and Petty addresses responses to *persuasive* arguments in particular, one might reasonably predict that if users tend to view a new result from a previously unused search engine as likely to offer greater information utility than another result from a search engine for which results have previously been examined, differences in their assessments of those results will follow. However, the potential application of this theory to search engine result evaluation is by no means clear-cut. First, it is not entirely clear whether users will employ any notion analogous to the pools of knowledge that differ from individual to individual. Users might recognize that different systems often search over distinct subsets of all available documents, or they may view all search engines as having access to the same basic pool of knowledge, namely all accessible Web documents. Second, it is also unclear what, if anything, users might take a search

engine's "perspective" to be. While some users may liken differing perspectives to varying search algorithms, it seems more likely that users would rely on a general notion of search engine quality, assuming that distinct search engines are likely to differ to some degree in this respect. Even this highly general notion of differing quality may be enough, however, to result in greater depth of processing and more varied assessments for results from multiple search engines than for the same results from a single search engine.

1.3. Perceived variability

Both perceived entitativity and perceived information utility may lead to predictive assessments favoring relatively homogenous assessments of results from a single search engine and relatively heterogeneous assessments for results from multiple search engines. First, highly entitative groups are often represented in terms of a group prototype, and this representation tends to favor homogenous assessments of group members (Brewer, Weber, & Carini, 1995). Note, however, that the causal relationship between perceived similarity and entitativity may be in either direction, or the two may have simultaneous causal influence over each other (Campbell, 1958; Brewer, Weber, & Carini, 1995).

Second, perceived information utility is based upon a predictive judgment that assumes distinct information sources are more likely to produce messages based on different perspectives and pools of knowledge. Moreover, it tends to lead to increased scrutiny of messages from multiple sources, and this too may favor more heterogeneous assessments. Thus, while perceptions of entiativity and information utility are not mutually exclusive as potential reasons for expecting increased variability in assessments of results from multiple search engines, predictions of information utility explicitly lead

to increased depth of processing.

While increased depth of processing affects evaluative judgments of information objects, insofar as differences occur as a result of predictive judgments, users must first infer particular properties of a search engine result prior to accessing it.

Inferred properties

Some properties of objects tend to be inferred more readily than others, and the general underlying causes of these differences may be found in the extensive literature on inductive reasoning. Critical to this study is the importance of causal explanations connecting the object category to the inferred property (Lassaline, 1996). While results from different search engines may be viewed as more heterogeneous with respect to topical relevance than the same results from a single search engine, other properties such as how well-designed the results are may not be inferred. Properties for which users perceive no connection to the function of a search engine as information source are unlikely to be inferred; idiosyncratic properties (for example, inferring that a Web page result is likely to have typographical errors because other results from the same search engine do) generally do not lead to widespread inference (Heit, 2000). Indeed, those properties for which users are willing to make such inductive inferences may provide insight into user mental models of what search engines do and how they work.

Depth of processing

The multiple source effect found by Harkins and Petty (1981) such that encountering information from new sources leads to greater processing may itself lead to increased perceived variability. Greater perceptual attention to an attitude object, first and foremost,

can lead to increased differentiation, both across individual objects for a single property, as well as across properties for an individual object (Irwin, Tripodi, & Bieri, 1967). We may also reasonably expect that increased processing will be accompanied by a decreased reliance on general halo effects where attitudes towards an object on one dimension are inferred from previously established attitudes on other dimensions (Dion, Berscheid, & Walster, 1972; Tractinsky, Katz, & Ikar, 2000). In any case, if greater processing occurs for information gathered from multiple search engines and this increased processing has an impact on perceived variability of results, we may be reasonably confident that the effect is likely to favor increased heterogeneity.

2. A STUDY OF SOURCE ORIENTATION IN INFORMATION RETRIEVAL

The purpose of this study was to test whether information seekers (1) perceive results gathered from a single search engine to be more similar to one another in terms of information quality than the same results gathered from multiple search engines, and (2) process such results differently, as indicated by differentiated recognition memory scores for the returned results. We investigated these two questions for both Web page results and image results.

2.1 Experiment 1: Web Page Results

Experiment 1 tested whether evaluation and processing of Web page results differed in the two information-seeking contexts. We asked participants to evaluate a set of Web page results on a range of information quality dimensions, manipulating only the

purported search engine from which these Web pages were returned.

Method

In a between-participants design, 32 participants were randomly assigned to one of two experimental conditions. In condition 1, participants evaluated a set of Web page results that were presented as having been returned by the same search engine (with a color code name, e.g., "search engine Fuchsia"). In condition 2, the same content was evaluated but was presented as having been returned by different search engines (each with a different color code name; Figure 1). Participants were told that the color-coded search engines were each distinct systems.

In both conditions, participants evaluated Web page results for two separate search topics ("newborn skin care" and "snowmobiling safety"), as a repeated measures factor. The experimental task was to examine the Web page results in order to obtain information regarding the search topic. For both topics, participants assessed five Web page results. The result position (ranking) of each of the results was made explicit in the stimuli. Stimulus sampling was used for both the set of Web page results assessed and the rankings of those results, which remained constant across the two conditions.

Participants assessed the quality of the resulting Web pages by indicating how well a set of information quality terms described the result, using 10-point rating scales. For both topics, participants took a recognition memory test for material presented within the result pages. They were asked to indicate whether they recognized a set of 20 sentences that may or may not have appeared in the result pages; of the 20 items, 10 were targets and 10 were foils.

Results

Factor analysis

To summarize and reduce the information quality measures, a principal-component factor analysis was performed, using varimax rotation. Those items whose factor loadings were at least .60 on one factor and no more than .40 on any other factor were retained. The Kaiser-Meyer-Olkin statistic was .83, suggesting that the sampling adequacy was sufficient to conduct the factor analysis and the multicollinearity assumption was met. Bartlett's test of sphericity was significant ($p<.001$), indicating that the correlation matrix associated with the factor analysis was not an identity matrix. The search engine result assessments loaded on a single result quality factor, with an eigenvalue of 4.11 and accounting for 68.5% of the total variance. The factor loadings for each of the information quality measures were as follows: credible (.87), informative (.87), authoritative (.85), relevant (.83), useful (.79), and well-written (.75). These items were used to create a quality index.

Perceived similarity

For each participant, the standard deviation of the quality index across the Web page results assessed was used to indicate the extent to which the results were differentiated from one another. A low standard deviation indicated that a participant did not differentiate, attitudinally, between Web pages in terms of quality. An independent samples two-tailed *t*-test was performed to compare differentiation scores between the two experimental conditions. Differentiation scores were found to be significantly greater in the multiple search engine condition ($M = 12.8$, $SD = 2.3$) than in the single search

engine condition ($M = 10.2$, $SD = 4.2$), $t(30) = 2.2$, $p<.05$.

Memory

Between the two search topics, each participant had a recognition memory score ranging in possible values from 0-40, with 20 as a random chance score. An independent samples two-tailed *t*-test was performed to compare recognition memory scores between the two experimental conditions. Memory for material on the Web page results was found to be significantly better in the multiple search engine condition ($M = 27.9$, $SD = 3.4$) than in the single search engine condition ($M = 24.4$, $SD = 4.7$), $t(30) = 2.4$, $p<.05$.

There was not a significant correlation between the differentiation score and recognition memory ($r=.22$, $p>.23$), suggesting that increased depth of processing was not a necessary condition for increased perceived differentiation of results.

Discussion

Participants evaluating the Web pages returned by multiple search engines perceived the results to be less similar to one another than participants evaluating the same content returned by a single search engine, in terms of a set of information quality dimensions. Moreover, when the results were presented as having been returned by multiple distinct search engines, participants showed greater recognition memory for material within those Web page results. These findings are consistent with the notion that information seekers will orient towards search engines as psychologically relevant information sources, independent of any previous experience with the particular search engines used; that is, these results occurred for generic search engines with color code names.

Thus, the results suggest a simple information-seeking heuristic: if a set of results are

returned from a single search engine, then (1) they are more likely to be similar to one another in terms of information quality dimensions (such as credibility and usefulness) than if those same results are each returned by different search engines, and (2) they warrant less processing depth than if they are each returned by different search engines. The second half of this heuristic is consistent with multiple source effects in human-human interaction as described by information utility theory (Harkins & Petty, 1981). When messages are received from what is perceived to be a distinct set of individuals, those messages receive greater scrutiny. While this increased depth of processing might reasonably be expected to lead to greater differentiation, as previously mentioned, there was not a significant correlation between differentiation score and recognition memory; thus, there was some evidence that increased processing depth was not a necessary condition for greater perceived differentiation of result quality. (We investigate this relationship further in Experiment 2.)

There are two important notes with respect to the information quality index. First, the index indicates that users expected differences along a range of quality dimensions that went beyond relevance to a query. The index included dimensions relevant to how believable the resulting content was perceived to be (credible and authoritative). Thus, information seekers expect that results returned by a single search engine will be more similar to one another not only in terms of traditional information retrieval dimensions, but in terms of credibility as well. Second, while one item on the index was somewhat idiosyncratic with respect to information retrieval (well-written), the index gives some indication of how information seekers perceive search engines to work and what dimensions of quality over which information seekers expect search engines to have

some degree of control.

2.2 Experiment 2: Image Results

Experiment 2 tested whether evaluation and processing of image results differed in the two information-seeking contexts. We asked participants to evaluate five sets of image results on a range of result quality dimensions, manipulating only the purported search engine from which these images were returned.

Method

In a between-participants design, 32 participants were randomly assigned to one of two experimental conditions. In this experiment, participants were asked to evaluate sets of image results; as in Experiment 1, the image result sets were presented either as having all been returned by the same search engine, or as having been returned by different search engines (Figure 2).

In both conditions, participants evaluated five sets of image results (for five different search terms). The experimental task was to examine the results in order to gather images that might be useful for teaching a group of children about the California outdoors. Images for five search topics were evaluated: "California coast," "California wildlife," "California lakes,", "California redwoods," and "California mountains." (As part of this task, participants chose images they believed would be particularly useful.)

Participants assessed the quality of the image result sets by indicating how well a set of terms described the results, using 10-point rating scales; to avoid participant fatigue, participants were asked to evaluate the quality of each set of returned images as a whole, rather than each individual image. Participants then took a recognition memory test for

the images. They were asked to indicate whether they recognized a set of 40 images that may or may not have appeared in the search engine result sets; of the 40 items, 20 were targets and 20 were foils.

Results

Factor analysis

To summarize and reduce the image set quality measures, a principal-component factor analysis was performed, using varimax rotation. Those items whose factor loadings were at least .60 on one factor and no more than .40 on any other factor were retained. The Kaiser-Meyer-Olkin statistic was .80, suggesting that the sampling adequacy was sufficient to conduct the factor analysis and the multicollinearity assumption was met. Bartlett's test of sphericity was significant ($p<.001$), indicating that the correlation matrix associated with the factor analysis was not an identity matrix. The image set assessments loaded on a single result quality factor, with an eigenvalue of 4.14 and accounting for 68.9% of the total variance. The factor loadings for each of the image set quality measures were as follows: visually stunning (.90), beautiful (.85), useful (.84), comprehensive (.81), high quality (.81), and relevant (.77). These items were used to create a quality index.

Perceived similarity

For each participant, the standard deviation of the result quality index across the sets of images assessed was used to indicate the extent to which the result sets were differentiated from one another. A low standard deviation indicated that a participant did

not differentiate, attitudinally, between image sets in terms of quality. An independent samples two-tailed *t*-test was performed to compare differentiation scores between the two experimental conditions. Differentiation scores were found to be significantly greater in the multiple search engine condition ($M = 9.2$, $SD = 2.9$) than in the single search engine condition ($M = 6.4$, $SD = 2.2$), $t(30) = 3.0$, $p<.01$.

Memory

Each participant had a recognition memory score ranging in possible values from 0-40, with 20 as a random chance score. An independent samples two-tailed *t*-test was performed to compare recognition memory scores between the two experimental conditions. Memory for the image results was not found to differ significantly as a function of number of search engines used to gather the results. In the multiple search engine condition, the mean memory score was 29.2 ($SD = 4.1$), compared to 28.2 ($SD = 5.4$) for the single search engine condition. To test for floor and ceiling effects, we compared memory scores in the two conditions to both chance (a score of 20) and the maximum possible score (40). To determine whether memory was significantly different than chance in each of the conditions, we determined how many standard errors away from chance the memory scores were; significance was determined using the *z*-score. Participants in both the multiple search engine ($z = 9.0$, $p<.001$) and single search engine ($z = 6.1$, $p<.001$) conditions did significantly better than chance. Using a similar strategy to check for ceiling effects, participants in both the multiple search engine ($z = 8.7$, $p<.001$) and single search engine ($z = 10.5$, $p<.001$) conditions were significantly below the maximum possible score. Thus, the tests indicate that the lack of significance between the two conditions is not a result of floor or ceiling effects.

As in Experiment 1, there was not a significant correlation between differentiation score and recognition memory ($r=.17$, $p>.34$), further suggesting that increased depth of processing was not a necessary condition for increased perceived differentiation of results.

Discussion

Participants evaluating the image results returned by multiple search engines perceived the results to be less similar to one another than participants evaluating the same images returned by a single search engine, in terms of a set of result quality dimensions. Thus, increased perceived differentiation appears to occur for both Web page and image results when multiple search engines are used. Additionally, differences in perceived similarity occurred for results across multiple (albeit related) search topics, indicating that the effects of source orientation towards search engines persist even when (slightly) specialized searches are performed with a set of search engines. Whether the result would persist given a highly unrelated set of search topics is unknown, but this experiment at least suggests that information seekers will perceive results across multiple searches from a single engine to be more similar in quality than from multiple search engines.

Differences in perceived similarity occurred in spite of no differences in memory for the images between the two contexts. In contrast to Experiment 1, participants did not appear to process the results in greater depth when returned by multiple, distinct search engines, likely because there is pragmatically less room for variation in the extent to which one studies a simple image in depth as compared to text. As with Experiment 1, a significant correlation was not found between differentiation score and recognition

memory. Experiment 2 therefore provided strong evidence that increased depth of processing is not a necessary condition for greater perceived differentiation of the search results.

As in Experiment 1, the quality index provides some indication of the result quality dimensions over which users may expect information retrieval systems to have some control. Somewhat interestingly, among these dimensions were aesthetic concerns (visually stunning and beautiful). However, this may depend in part on the tasks used in this study; to the extent that the usefulness of image results for a particular task is unrelated to aesthetic concerns, users may not expect aesthetic differences in images from one search engine to another.

3. GENERAL DISCUSSION

At its core, this study demonstrates an instance in which precisely the same content may be evaluated and perceived differently based upon the information-seeking context in which it is presented. Just as receiving messages from different people—each with varying points of view and expertise—can cause information seekers to process and respond to the content differently than if it came from a single individual, the context in which search results are gathered can influence how deeply users process the results, and ultimately how they will perceive the quality of what they have found.

While this study indicates in part that information retrieval can in some ways echo information gathering in human-human interaction, it also addresses a fundamental difficulty in assessing multiple source effects (see, e.g., Lee & Nass, 2004). If multiple individuals each present a single piece of content on a given topic, the recipient might

reasonably believe each individual presented their highest quality content (i.e., each individual's "best argument" or "most relevant information"). If the same content was presented by a single individual, on the other hand, the recipient might instead reasonably believe the content differed in quality (i.e., the individual's "top five arguments"). This reasoning may well be a cause of multiple source effects found in the existing literature. While *persuasive* effects of multiple sources, in particular, were not the focus of this study, we believe our approach can account for an experimental difficulty in assessing multiple source effects more generally. This study indicates that after accounting for the ranking of the information presented, users perceive results gathered from a single search engine to be significantly more similar to one another in quality than the same results gathered from multiple search engines—with the effect occurring for both textual Web page and image results. Moreover, in the case of Web page results, where information seekers can differ greatly in the extent to which they examine and thoroughly process the resulting content, users show better memory for results when gathered via multiple search engines; in the case of image results, where there is pragmatically less room for variation in processing depth between results, this effect is eliminated.

While information retrieval systems are of course in a real sense not responsible for the content they filter, users will nonetheless process and respond to retrieved documents as though their shared "origin" (i.e., the search engine used to retrieve them) ensures a likely degree of similarity. The dimensions along which results are perceived to be more similar to one another can help shed light on how users believe such systems work. For example, the result quality index for the Web page results in this study included several information quality dimensions, ranging from relevance and usefulness to credibility and

authority. That the index included credibility and authority is particularly instructive (see, e.g., Fogg, 2003). If users expect results from a single search engine to be relatively similar in credibility, a mismatch between these expectations and how search engines actually filter and organize results can be disastrous. This study suggests that gathering information from a single search engine could cause users to more readily believe (or more readily disbelieve) a document, based upon the credibility of documents accompanying it in a search results list. In short, the documents, to a degree, become associated with one another, and users may consequently infer properties that are not necessarily appropriate to infer. This appears to be a general information-seeking strategy, in that it is not dependent upon any previous experience with the specific search engine used.

3.1 Limitations

While this study indicates differences in perceived result similarity and depth of processing for two different information-seeking contexts, it does not attempt to explore whether such differences persist across searches performed on highly disparate topics, or over time. The two experiments in this study examined search contexts in which results all pertained to the same search topic (Experiment 1) or to a set of fairly related topics (Experiment 2), within a single information-seeking session lasting roughly 20 to 30 minutes.

Additionally, while this study provides evidence that increased depth of processing is not the cause of increased perceived differentiation of the search results, it cannot pinpoint the process by which users make these assessments. It is unknown whether users

assume properties of the search results prior to accessing them, or if they change their evaluative strategies and behaviors (in some way other than increased processing depth), leading to the differences in perceived similarity.

3.2 Future work

Investigating information seeking behavior over time and across several topics, recording user expectations prior to accessing search results, and measuring user behaviors as they evaluate the results (for example, via eye tracking) could address limitations in what this study can indicate about user assessments of search engine results. Additionally, future work can both extend the experiments described here and investigate potential mediating variables in the assessment differences observed when information seekers use one or multiple search engines. First, because the dimensions along which such differences occur provide insight into user models of what search engines do and how they work, future work might attempt to map out precisely which qualities tend to be inferred from one result to another, and which tend to be viewed as idiosyncratic.

Second, the types of search objects for which these differences occur may also indicate the contexts under which users will orient towards search engines as psychologically relevant information sources. This study has shown the effect to occur for both textual Web page and image results. Whether user perceptions and assessment strategies will be influenced for products, services, people, locations, news items, or any other type of information search engines might filter remains an open area for investigation.

Third, future work may extend these experiments to the assessment of various modalities, including audio and video files. In these cases, there frequently is a highly salient source (a voice and/or a face) that may overwhelm source orientation towards search engines and other information filters. More generally, whether or not search engines would still influence user processing and attitudes towards results following switches in modality (as opposed to search engines using the same input and/or presentation modality as the results) is unclear.

Finally, it is important to determine the extent to which source salience interacts with the effects discussed here. Online information filters can differ substantially in the extent to which result branding (relative to the identity of the filter itself) is both salient and reinforced during information seeking; this can determine which factors tend to influence user assessments of online information (Fogg, Soohoo, Danielson, Marable, Stanford, & Tauber, 2003). If the brand of the gatekeeper (for example, a travel site) is less prominent due to the salience of the result brands (airlines, hotels, etc.), for example, we may reasonably expect less reliance on the information filter itself in assessing result quality.

3.3 Conclusion

Determining who or what is responsible for messages in human-computer interaction becomes more and more difficult for users as content is filtered and presented in ways that obfuscate its origin. This study demonstrates how such filtering can ultimately influence user attitudes and behaviors towards information in a ubiquitous context— information seeking on the Web. As with human information sources, users expect a relative level of heterogeneity in the quality of information presented by distinct search

engines, and when the objects allow for significant variation in processing depth, users

will place greater scrutiny on information when gathered from multiple search engines.

The particular dimensions for which greater heterogeneity of result quality is expected

can provide insight into user models of how information retrieval systems work. This

study suggests that such user expectations extend beyond traditional information retrieval

dimensions.

REFERENCES

Brewer, M.B., Weber, J.G., & Carini, B. (1995). Person memory in intergroup contexts: Categorization versus individuation. *Journal of Personality and Social Psychology, 69*, 29-40.

Burbules, N.C. (2001). Paradoxes of the Web: The ethical dimensions of credibility. *Library Trends, 49*, 441-453.

Campbell, D.T. (1958). Common fate, similarity, and other indices of the status of aggregates of persons as social entities. *Behavioral Science, 3*, 14-25.

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology, 24*, 285-290.

Eastin, M.S. (2001). Credibility assessments of online health information: The effects of source expertise and knowledge of content. *Journal of Computer-Mediated Communication, 6(4)*.

Fogg, B.J. (2003). *Persuasive Technology: Using computers to change what we think and do*. San Francisco: Morgan Kaufmann.

Fogg, B.J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P., & Treinen, M. (2001). What makes Web sites credible? A report on a large quantitative study. *Proceedings of ACM CHI'01, Human Factors in Computing Systems*, 61-68.

Fogg, B.J., Soohoo, C., Danielson, D.R., Marable, L., Stanford, J., & Tauber, E.R.

(2003). How do users evaluate the credibility of Web sites? A study with over 2,500 participants. *Proceedings of DUX2003, Designing for User Experiences Conference.*

Harkins, S. and Petty, R.E. (1981). Effects of source magnification of cognitive effort on attitudes: An information processing view. *Journal of Personality and Social Psychology, 40*, 401-413.

Harkins, S. and Petty, R.E. (1987). Information utility and the multiple source effect. *Journal of Personality and Social Psychology, 52*, 260-268.

Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin & Review, 7*, 569-592.

Heit, E. (2001). Background knowledge and models of categorization. In U. Hahn & M. Ramscar (Eds.), *Similarity and Categorization*. Oxford: Oxford University Press, 155-176.

Hogarth, R.M. (1987). *Judgment and Choice: The psychology of decision (2nd ed.)*. New York: John Wiley & Sons.

Irwin, M., Tripodi, T., & Bieri, J. (1967). Affective stimulus value and cognitive complexity. *Journal of Personality and Social Psychology, 5*, 444-448.

Lassaline, M.E. (1996). Structural alignment in induction and similarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 754-770.

Lee, K.M. & Nass, C. (2004). The multiple source effect and synthesized speech:

Doubly-disembodied language as a conceptual framework. *Human Communication Research, 30*, 182-207.

Lickel, B., Hamilton, D.L., Wieczorkowska, G., Lewis, A., Sherman, S.J., & Uhles, A.N. (2000). Varieties of groups and the perception of group entitativity. *Journal of Personality and Social Psychology, 78*, 223-246.

Nass, C. & Brave, S. (2005). *Wired for Speech: How voice activates and advances the human-computer relationship.* Cambridge, MA: MIT Press.

Nass, C. & Steuer, J. (1993). Voices, boxes, and sources of messages: Computers and social actors. *Human Communication Research, 19*, 504-527.

Petty, R.E. & Cacioppo, J.T. (1986). The Elaboration Likelihood Model of persuasion. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology, Vol. 19*. San Diego: Academic Press, 123-205.

Pirolli, P. and Card, S.K. (1999). Information foraging. *Psychological Review, 106*, 643-675.

Reeves, B. & Nass, C. (1996). *The Media Equation: How people treat computers, television, and new media like real people and places*. New York: Cambridge University Press.

Rieh, S.Y. (2002). Judgment of information quality and cognitive authority in the Web. *Journal of the American Society for Information Science and Technology, 53*, 145-161.

Rieh, S.Y. & Belkin, N.J. (1998). Understanding judgment of information quality and cognitive authority in the WWW. *Proceedings of the 61st Annual Meeting of the American Society for Information Science, 35*, 279-283.

Stewart, K.J. (2003). Trust transfer on the World Wide Web. *Organization Science, 14*, 5-17.

Sundar, S.S. & Nass, C. (2000). Source orientation in human-computer interaction: Programmer, networker, or independent social actor? *Communication Research, 27*, 683-703.

Sundar, S.S. and Nass, C. (2001). Conceptualizing sources in online news. *Journal of Communication, 51*, 52-72.

Toms, E.G. & Taves, A.R. (2004). Measuring user perceptions of Web site reputation. *Information Processing and Management, 40*, 291-317.

Tractinsky, N., Katz, A.S., & Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers, 13*, 127-145.

Yzerbyt, V., Judd, C.M., & Corneille, O. (2004). Perceived variability, entitativity, and essentialism: Introduction and overview. In V. Yzerbyt, C.M. Judd, & O. Corneille (Eds.), *The Psychology of Group Perceptions: Homogenity, entitativity, and essentialism*. Philadelphia: Psychology Press, 1-24.
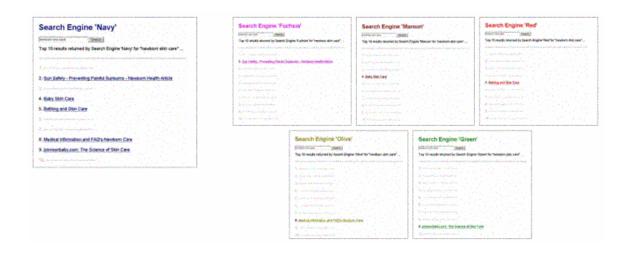
Figure 1

Experiment 1 Conditions: In the single search engine conditions (left), participants evaluated five of the "top 10 results" returned by a color code-named search engine. In the multiple search engine condition (right), participants evaluated the same results, with the same result positions, returned by five different color code-named search engines. Web pages and result positions were stimulus sampled, with the samplings being held constant between conditions.

Figure 2

Experiment 2 Conditions: In the single search engine condition (top), participants evaluated five sets of image results (for five related queries) returned by a color code-named search engine. In the multiple search engine condition (bottom), participants evaluated the same results returned by five different color code-named search engines.