Effects of Popularity and Quality on the Usage of Query Suggestions during Information Search

Diane Kelly, Amber Cushing, Maureen Dostert, Xi Niu, Karl Gyllstrom

University of North Carolina at Chapel Hill

Chapel Hill, NC 27599-3360 USA

[dianek, cushinga, dostert, xiniu @ email.unc.edu; karl@cs.unc.edu]

ABSTRACT

Many search systems provide users with recommended queries during online information seeking. Although usage statistics are often used to recommend queries, this information is usually not displayed to the user. In this study, we investigate how the presentation of this information impacts use of query suggestions. Twenty-three subjects used an experimental search system to find documents about four topics. Eight query suggestions were provided for each topic: four were high quality queries and four were low quality queries. Fake usage information indicating how many other people used the queries was also provided. For half the queries this information was high and for the other half this information was low. Results showed that subjects could distinguish between high and low quality queries and were not influenced by the usage information. Qualitative data revealed that subjects felt favorable about the suggestions, but the usage information was less important for the search task used in this study.

Author Keywords

Query suggestion, query recommendation, usage, social search, query quality, query popularity, search behavior.

ACM Classification Keywords

H5.2. Information interfaces and presentation (e.g., HCI): User interfaces.

General Terms

Human Factors.

INTRODUCTION

Social search techniques can assist searchers during information seeking and retrieval by recommending queries, information objects and search paths, and providing social tagging and rating facilities [c.f., 12, 14, 17, 24]. Freyne, et al. [12] characterize social search techniques as using search history information, such as

users' queries and selections, to adapt searching activities to people with similar interests. The basic idea behind these techniques is that retrieval can be improved for new users by incorporating information about past users' searches and behaviors. While there are currently many definitions of social search, including those that encompass collaborative information seeking [e.g., 11], in this paper we focus on social search as defined by Freyne, et al. We further focus on query suggestion features and open-ended search tasks where the user's goal is to explore a topic.

Query suggestions can be particularly beneficial for information search tasks because they provide users with alternative methods for exploring topics and can potentially help users develop better understandings of their topics. Query suggestions have the potential to provide what Bates [3] refers to as *idea tactics* or moves to help generate new ideas or solutions to information search problems. It is also the case that many models of information seeking show that when searching for difficult topics users often require additional search assistance, especially with query formulation [e.g., 4, 28]. Query suggestions can provide assistance by helping the user continue to execute searches even when they are unable to formulate their own queries.

Many current search services offer query suggestions. When users issue queries to Google, Yahoo! or Bing they receive a list of search results and a list of related queries. While the presentations of the query suggestions differ, common to all presentation methods is that no information is provided about how suggestions were selected. Query suggestions are likely generated by a combination of techniques that use query log data. However, details about the usage and frequency of various queries are not displayed to users. This type of information might help users make better choices about the potential usefulness of query suggestions. In this paper, we investigate how the display of usage data alongside query suggestions affects users' search behavior. Specifically, the impact of query suggestion popularity on search behaviors is investigated.

Query popularity provides a type of social evidence by revealing the behaviors of other users. While this information might assist users during search, it might also introduce bias into the system. Social information biases people's behaviors in many different situations and in many different ways [21, 23]. In the context of search, social

information may lead users to reproduce or imitate others' behaviors instead of exploring new search directions. For example, when users select recommended queries they see the same (or similar) results as others which may lead to a narrow view and understanding of the topic. If document popularity is also measured then the problem is magnified. Popular queries may stay popular since they are the queries that are suggested to users and available for selection. Popular documents may converge on those which are returned by popular queries since they are retrieved and viewed more often by users.

RELATED WORK

There are many theories and models in the fields of psychology, sociology and economics that describe and explain how social information influences behavior. This literature is too vast to be explored in detail in this paper, but some of the major ideas from this literature that have directly influenced the work described in this paper will be presented briefly. Bikhchandani, et al. [5] describe an informational cascade as occurring when "it is optimal for an individual, having observed the actions of those ahead of him, to follow the behavior of the preceding individual without regard to his own information" (p. 994). Informational cascades occur in situations where sequential choices are made by, and visible to, a series of people all of whom are faced with the same decision. Using formal theory, Bikhchandani, et al. [5] demonstrate how most people will prefer the social information over their own personal information and that after enough social information has accumulated, a new person's decision (even when it is contrary to the majority) is uninformative to others. Surowiecki [26] argues that collective intelligence can improve decision-making, but notes that this is only possible when individuals are diverse and each has some private, unique information and it is shared with the group.

Informational cascade is closely related to a number of other theories including herd behavior [1, 22] and conformity [7, 8]. Rook [22] states that economic descriptions of herd behavior primarily indicate that people are motivated by the sheer number of others that have already partaken in a particular action, but that psychological descriptions focus on more varied human motivations for herd behavior including conformity, a desire for consensus, and the desire to affiliate with particular reference groups. Conformity, or aligning one's behaviors with the behaviors of others [8], is often motivated by fear of sanctions or other negative social consequences. While it is unlikely that a fear of sanctions would motivate a user to imitate other's information seeking behaviors in non-collaborative settings, conformity is also influenced by what Deutsch and Gerard [10] refer to as informational social influence which is a desire to form an accurate understanding of reality. These types of influences (also referred to as social proof [7]) are particularly useful for helping people understand how to behave in ambiguous social situations, which is likely a

more useful perspective for exploring social influence in information seeking domains since the motivation to use the social information may be related to a knowledge gap.

There are empirical studies from the recommender system literature that investigate the impact of social influence on behaviors. In a study of social navigation in a recipe recommendation system, Svensson, et al. [27] identified a snowball effect where social trails led increasing numbers of users down paths that were ultimately not useful to them. Suchanek, et al. [25] studied the imitation rate of users in a social tagging situation and found that users were slightly more likely to imitate popular tags. However, nearly 65% of the users in this study claimed that they paid no attention to the suggested tags. In a study of movie recommendations and ratings, Cosley, et al. [9] found that users could be manipulated to rate movies in a direction consistent with the system's movie ratings, and that in the case of users re-rating previously rated movies, the system influenced some users to move from a negative to positive rating.

Researchers in the areas of business, marketing and economics have studied the extent to which social information impacts item selection [13] and purchasing decisions [6, 23]. Hanson and Putler [13] artificially inflated the number of times various shareware was downloaded at a site sponsored by AOL and found that manipulations increased the rate at which certain programs were downloaded – manipulating the download frequency in a positive direction led to more downloads when all other aspects of the shareware were held constant. In a laboratory experiment, Chen [6] investigated the impact of star ratings and sales volume on people's book purchasing decisions and found that subjects relied heavily on ratings and sales volume. Senecal and Nantel [23] examined the impact of recommendation sources (other consumers, human experts, recommender system and no recommendation) on subjects' purchasing decisions. The researchers also manipulated website type and product type. Results showed that subjects' purchasing decisions were influenced by online recommendations; however, the strongest effect was when the source of the recommendation was a recommender system. Senecal and Hantel also found an effect for product type: recommendations for experience products (products whose qualities cannot be determined prior to purchase, e.g., wine) were significantly more influential than recommendations for search products (products whose qualities can be determined prior to purchase, e.g., car).

The empirical studies suggest that social information can influence users' behaviors in some situations. However, these situations have been primarily limited to shopping tasks and leisure tasks. It is unclear how social information impacts users engaged in more cognitively demanding tasks such as information seeking and retrieval tasks. Many researchers have studied query suggestion and other social search features in the information seeking domain [20, 24, 30] and found that users find these features useful and they can led to more effective information seeking, but studies

have not been conducted to understand how social information influences behaviors in this domain.

One way that query suggestions can improve the search experiences of users is by providing what Bates [3] refers to idea tactics. Idea tactics are moves to help generate new solutions to information search problems. Bates noted that idea tactics serve a psychological purpose in that they are intended to help improve the user's thinking and creative processes (p. 280). Bates further justifies the importance of idea tactics by observing that new ideas are often "blocked or limited by one's current thinking" (p. 281). The basic idea is that the user's internal model of the information search problem can sometimes block their efforts to think of novel and useful ways to proceed with search. Bates proposed a number of tactics emphasizing idea generation and pattern-breaking. Idea generation primarily focuses on stimulating new ideas by thinking and conducting activities outside of a retrieval system. Pattern-breaking tactics help the user go beyond his or her current way of thinking about the problem and suggest moves that can be made while interacting with a retrieval system. Although some pattern-breaking tactics are intended to be used by the searcher introspectively, a number of these tactics focus on search and query behavior. However, the user's ability to use these tactics may be limited since often the user does not know a lot about the information need and requires assistance with querying, especially when searching for difficult topics.

Many information seeking models show that users need different types of interface support in different situations and that users' information seeking strategies, as well as the level of system support they desire, varies according to a number of factors including topic difficulty and search stage [4, 28]. In cases where users need more assistance with information seeking, they might use more query suggestions and rely more heavily on social information because of their lack of knowledge. There is some evidence for this: in a social psychology experiment, Baron, et al. [2] found that task difficulty increased conformity. Since subjects had less private information, they relied more on social information to guide their behaviors.

RESEARCH QUESTIONS
The research questions are: (1) Are users influenced by usage information associated with recommended queries? (2) Can users distinguish between high and low quality query suggestions? (3) What is the relationship among topic difficulty, users' willingness to take recommendations and their abilities to distinguish between good and bad queries? (4) What are users' perceptions of the usefulness of query suggestions and usage information for open search tasks?

METHOD
A controlled laboratory study was conducted with a single experimental search system implemented via the Web. The study had two within-subject independent variables: query popularity (social information) and query quality; one

quasi-independent variable: experienced topic difficulty; and two major dependent variables: use and ratings of query suggestions. Each of these components and variables is described in more detail below. Our primary reasons for conducting this study in a laboratory were so that we could control the quality and popularity information associated with suggested queries, and the search environment.

Interface and Search System
The search interface is displayed in Figure 1. This interface displays query suggestions to subjects with social information about the number of other people who have submitted the suggested queries. From this interface, subjects could view the documents they had saved, their past queries and the current search topic. Clicking on the title of a search result replaced the search results list with the full text of the document. From the full text view, subjects could save the document. We used Lemur (www.lemurproject.org) as our underlying search engine. For each query, we returned the top 100 documents, which were displayed on a single page.
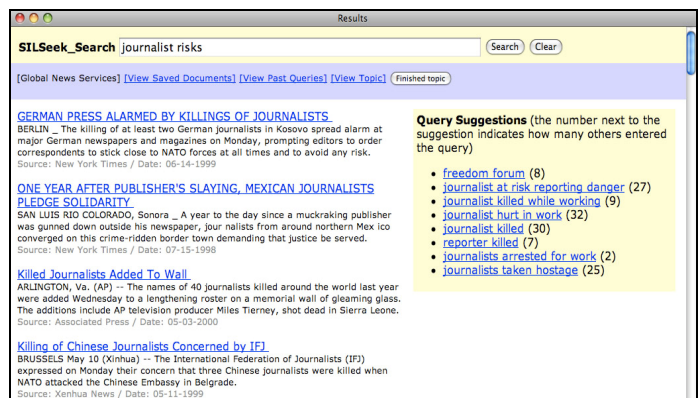

Figure 1. Interface for Search System

Document Corpus and Search Topics
A closed corpus of newspaper articles and assigned search topics were used in this study. This allowed us to better control the experimental situation and isolate the effects of query popularity and quality. A test collection from the Text Retrieval Conference (TREC) was used in this study [29]. This collection contained a 3GB corpus of newswire text (over one million documents) and 50 search topics.

Four topics from the original set of 50 were selected for use in this study (Table 1). We manually selected these topics by considering past users' performances with the topics from another study [20], the number of relevant documents in the corpus and whether we thought the topic would be of interest to our target subjects (undergraduates). Subjects had up to fifteen minutes to search for each topic and were instructed to save the relevant documents they found. Topics were rotated using a Latin-square.

Topic Title	Description
354: journalist risks	Identify instances where a journalist has been put at risk (e.g., killed, arrested or taken hostage) in the performance of his work.
426: law enforcement, dogs	Provide information on the use of dogs worldwide for law enforcement purposes.
393: mercy killing	Identify documents that discuss mercy killings.
638: wrongful convictions	Find documents that discuss freed prisoners who have been wrongfully convicted based on faulty forensic evidence, poor police work, or false testimony.

Table 1. Search topics used in study.

Query Popularity and Quality

For each topic, subjects were presented with eight query suggestions. Four suggestions were high quality queries and four were low quality queries. Two of the high (and low) quality queries were presented to subjects as popular queries, while two of the high (and low) quality queries were presented as unpopular queries. The only constraint was that there were always four high quality and four low quality queries and within these sets, two were described as popular and two were described as unpopular.

To identify high and low quality queries for suggestion, we examined queries submitted by subjects in another study [20] where the same topics were being used and selected the best and worst performing queries based on how many relevant documents were returned in the top 20. These suggestions were held constant and the only thing that varied was their presentation order (which was random) and the social information associated with them (which we manipulated). For any within subject-topic pair, the order of the recommended queries, as well as the associated social information, stayed the same. These factors only varied between subject-topic pairs.

No information appeared on the interface to indicate the quality of the query, but social information was provided alongside the query suggestions (see Figure 1). We defined social information as query popularity, or the number of other people who had used the query, and operationalized this by associating a value in the ranges 20-40 to denote popular queries and 1-9 to denote unpopular queries. Values were selected randomly from each range. The interface indicated that the number in parenthesis next to each suggested query described how many other users had entered the query. We used a larger range of numbers to describe popular values because we did not want subjects to become suspicious if they noticed a lot of repetition in these values. For unpopular values, we did not believe that subjects would become suspicious if they saw numbers from the set 1-9 repeated since there would naturally be

more repetition in these low values. For example, having many queries that only a single person entered is more likely that having many queries that 23 people entered.

Topic Difficulty

We asked subjects to indicate how difficult it was for them to find relevant documents for the topic after they completed each search. This was measured on a 5-point scale (1=very easy; 5=very difficult). This allowed us to investigate if subjects' experienced difficulty impacted their use of suggestions and their abilities to distinguish between good and bad quality queries.

Ratings of Suggested Queries

After subjects finished each search they were shown the eight suggested queries and asked to evaluate them according to query quality (1=very poor; 5=very good), their confidence in their quality rating (1=very unsure; 5=very confident) and the likelihood they would recommend the query (1=not very likely; 5=very likely).

System Evaluation

Subjects completed an exit questionnaire after they finished searching which asked them to evaluate the effectiveness of the search system (4 items) and their satisfaction with their performance (1 item). We are less interested in subjects' responses to these items because this was not a system evaluation. However, we included this questionnaire in the protocol because subjects were told they would be helping us evaluate a system. These results are not presented in this paper with one exception: a question was included which asked subjects to indicate the extent to which the information provided about how many people used the different queries determined whether or not they used the suggestions. We used this question, in part, to understand more about subjects' perceptions of the usefulness of query recommendation. This was assessed on a 5-point scale (1=strongly disagree; 5=strongly agree).

Subjects' Perceptions

After completing the exit questionnaire, subjects completed an exit interview which allowed us to find out more about subjects' experiences with, and perceptions of, social search features. At the start of the interview, subjects were shown some screen shots of example social search features such as Google suggestions, YouTube comments and ratings and Flickr tags. Subjects were then asked to indicate their familiarity with social search features (1=very unfamiliar and 5=very familiar) and the frequency with which they have used these types of features (never, once or twice, three or four times, too many times to remember). Subjects were asked to provide an example of when they used a social search feature and discuss the outcome. Subjects were also asked open-ended questions about their perceptions of the suggested queries in the current study and the extent to which the social information (popularity) influenced their decisions to take a suggestion. Subjects were also asked to comment on how useful, in general, they

find social information and how useful they find it in search situations. Finally, a debriefing and manipulation check was performed and subjects were asked whether they were suspicious of the query popularity information.

Summary of Procedures

Subjects were told they would be helping the researchers evaluate an experimental search engine. Subjects were not provided with a tutorial and were not told about the suggested queries. Subjects completed a consent form and demographic questionnaire. Subjects were then presented with a search topic and proceeded to search. This was followed by the post-search questionnaire. Subjects repeated this sequence for each topic. This was followed by the exit questionnaire and interview. Finally, subjects were debriefed about the deception and given a $20 USD honorarium. The individual sessions lasted about 1 hour.

Subjects

Subjects were recruited via email solicitation to the undergraduate student mailing list at our university. Twenty-three subjects participated in this study (16 females and 7 males). Subjects' mean age was 21 years (SD=1.9). All were undergraduates except for one graduate student. Thirteen percent of the subjects were humanities majors, 30% were social science majors, 22% were science majors and 35% were in a professional school. While we could have used an online recruitment service such as Amazon's Mechanical Turk, we were interested in interviewing our subjects and also wanted to have a better idea of who they were and what they did while participating in the study.

RESULTS

We first present descriptive results showing the usage of the suggestions. Next, we present an analysis of the effects of the query popularity and query quality on subjects' selection of query suggestions. Following this, we present data describing subjects' ratings of topic difficulty and an analysis of how topic difficulty influences use of suggestions. We further examine differences in the query popularity and query quality associated with the query suggestions subjects selected. In the final section, we present qualitative data from our interviews with subjects which provide insight into subjects' general use and perceptions of social search features, as well as their experiences and use of suggestions in this study.

Use of Suggestions

Subjects submitted a total of 722 queries for all topics combined (32 queries on average, or about 8 queries per subject per topic). Four-hundred twenty-five (59%) of these queries were of their own creation while 297 (41%) were suggestions. Each subject was shown a total of 32 suggestions (8 suggestions per topic * 4 topics) and selected an average of 13.70 (SD=7.02). One subject did not select any suggested queries, while another selected 24.

We examined the overlap between the queries subjects entered manually and the suggested queries to see how many duplicates occurred. That is, how many times a manually entered query matched exactly a suggested query. We found that 113 of the 425 subject-created queries were exactly the same as one of the suggested queries. Of these 113 queries, only 7 were the first queries typed by subjects (in which case they would not be duplicates since subjects did not receive any suggestions until they entered one query). Although there is no way to ascertain that subjects did not naturally type the 106 duplicate queries on their own, these results suggest that some subjects may have taken suggestions by manually entering the queries.

Figures 2 and 3 show that subjects integrated query suggestions into their searching fairly quickly. Figure 2 counts suggestions as only those that were clicked, while Figure 3 includes both clicked suggestions and manually entered duplicates. Figure 2 shows that about half of the second queries entered by subjects were suggestions. Figure 3 shows that subjects took roughly twice as many suggestions for their second queries. Counting manually entered queries that duplicate query suggestions provides a different picture of the use of the suggestions. Ultimately, these data show that using a simple click metric to measure uptake may not tell the whole story. However, in the interest of space, we only use clicks to measure usage in the remainder of analyses reported in this paper. We conducted all analyses using clicked suggestions and a combination of clicked+typed suggestions, but this did not change the results drastically.
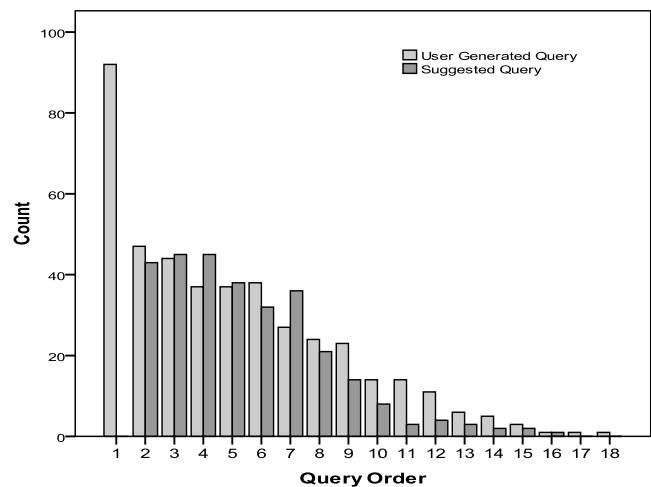

Figure 2. Frequency and source of query (self-created or suggested) according to order of submission during search. Only queries clicked by subjects are counted as suggestions.
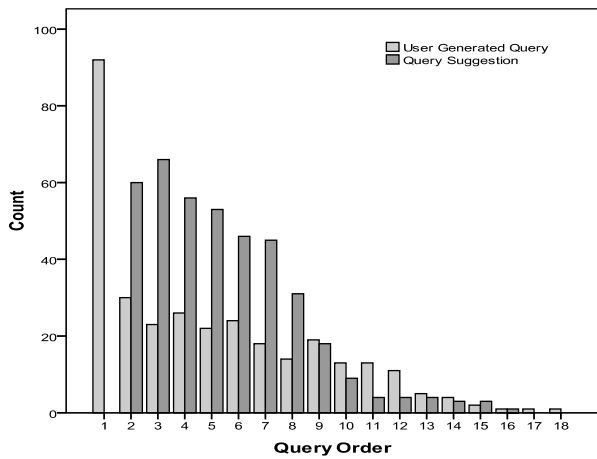
Figure 3. Frequency and source of query (self-created or suggested) according to order of submission during search. Suggestions include queries clicked on and subjects' manually entered queries that matched the suggestions exactly.

Recall that query suggestions were presented randomly to subjects to control for presentation order effects. Figure 4 shows the relationship between query position and selection. Of the 92 queries that were presented in position 1, 47 of them were selected. Overall, more queries presented in position 1 were selected than those presented in other positions. For position 8, only 27 of the 92 suggested queries were selected. Positional bias has been shown to exist in many situations, most recently in the context of ranked search results [18, 19]. These results suggest that a bias may also exist in the context of query suggestion.
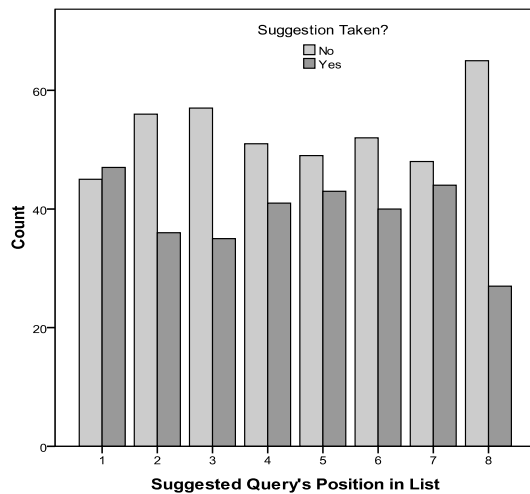

Figure 4. Frequency of suggested queries being taken (i.e., clicked on) based on position in recommendation list.

Query Popularity, Quality and Use

Were subjects more likely to take popular query suggestions than unpopular suggestions? Figure 5 shows the relationship between the social information associated with a query suggestion and whether the suggestion was taken by subjects. Of the query suggestions taken, 148 (47%) were associated with low social information and 165 (53%) were associated with high social information.
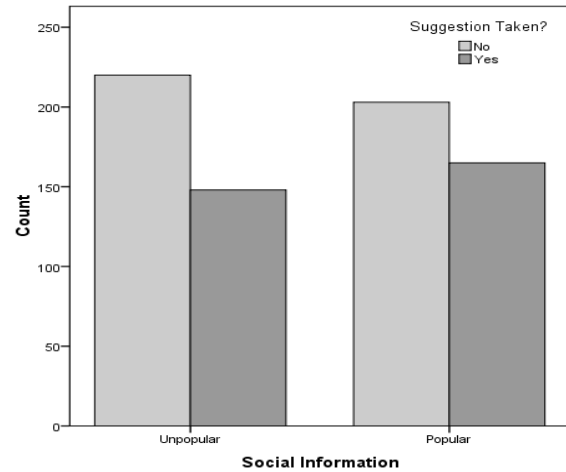

Figure 5. Frequency of popular and unpopular queries taken (i.e., clicked on) by subjects.

Were subjects more likely to select a high-quality query than a low-quality query? Figure 6 shows the distribution of query suggestions clicked according to quality. Of the suggestions taken, 170 (54%) were high-quality query suggestions and 143 (46%) were low-quality suggestions.
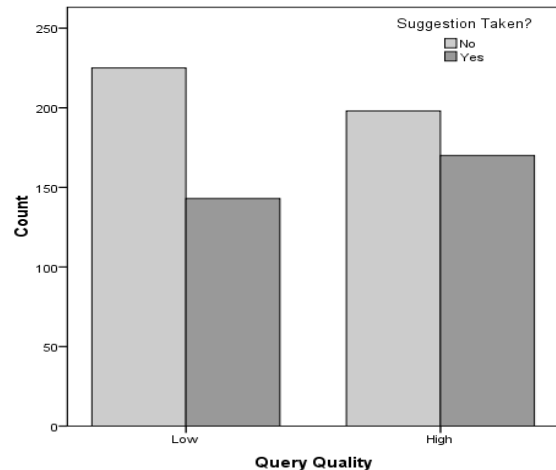

Figure 6. Frequency of high and low quality queries being taken (i.e., clicked on) by subjects.

A logistic regression analysis was conducted to determine if the observed differences were statistically significant. The outcome measure – selection of query suggestion – was a nominal, binary variable which is why logistic regression was selected over ANOVA or multiple regression (each of which assumes a continuous outcome measure).

The logistic regression showed that query quality was a significant predictor of suggestion usage, but query popularity was not. The Wald chi-square statistic associated with each predictor was 1.61, $p=.203$ (query popularity) and

4.05, p=.044 (query quality). However, for each variable, the odds ratios were similar: 1.20 for query popularity and 1.35 for query quality. These ratios show that the odds of a popular query suggestion being clicked is .813, while the odds of an unpopular query suggestion being clicked is .672 (the odds ratio is computed as .813/.672). The odds of a high quality query suggestion being clicked is .859, while the odds of a low quality query suggestion being clicked is .636. The Nagelkerke R square for the predictor model using query quality was .010. This is quite low and shows that the model does not explain a lot of variance.

Subjects' Evaluatoins of Query Quality
Subjects were asked to evaluate the quality of each suggested query after they finished searching. Table 2 shows these mean ratings according to the social information associated with the queries. There was very little difference in quality, confidence and willingness to recommend ratings according to the social information.

	Popularity	Mean (SD)	t-test
Quality	Low	2.87 (1.15)	$t(734)$=-.665,
	High	2.93 (1.18)	p=.506
Confidence	Low	2.90 (1.11)	$t(734)$=-.370,
	High	2.93 (1.08)	p=.711
Recommend	Low	2.84 (1.19)	$t(734)$=-.457,
	High	2.88 (1.23)	p=.647

Table 2. Subjects' evaluations of query suggestions according to popularity.

High quality queries were rated significantly higher than low quality queries and subjects were more likely to recommend these queries (Table 3). Subjects were slightly more confident in their evaluations of high quality queries, but this difference was not significant.

	Query Quality	Mean (SD)	t-test
Quality	Low	2.80 (1.18)	$t(734)$=-2.38,
	High	3.00 (1.15)	p=.017
Confidence	Low	2.84 (1.09)	$t(734)$=-1.79,
	High	2.99 (1.09)	p=.074
Recommend	Low	2.74 (1.20)	$t(734)$=-2.73,
	High	2.89 (1.20)	p=.007

Table 3. Subjects' evaluations of query suggestions according to query quality.

Topic Difficulty and Selection
Table 4 displays subjects' average self-reported difficulty ratings of each topic. Overall, Topic 354 was rated the easiest, while 426 was rated the most difficult. These differences were statistically significant, [$F(3, 91)$=8.36, p<.01]. Scheffe's post-hoc tests showed that there were significant differences between several pairs of topics: Topic 354 was significantly less difficult than Topics 393 and 426, and Topic 426 was significantly more difficult than Topics 354 and 638.

Topic	Mean	Standard Deviation
[354] Journalist risks	0.74	.81
[638] Wrongful convictions	1.26	1.01
[393] Mercy killing	1.74	1.05
[426] Law enforcement, dogs	2.09	.99

Table 4. Subjects' self-reported topic difficulty ratings on a 5-point scale where 1=very easy; 5=very difficult.

Figure 7 shows the number of suggestions taken per topic. The most query suggestions were used for Topic 426 (the most difficult topic) and the least for Topic 354 (the easiest topic). A Chi-Square test showed that subjects were more likely to use suggestions when the topic for which they were searching was hard [$\chi^2(3, 736)$=24.12, p<.01]. The odds of a suggestion being taken according to topic difficulty (with the most difficult topic presented first) was 1.24, .786, .672 and .444, respectively.
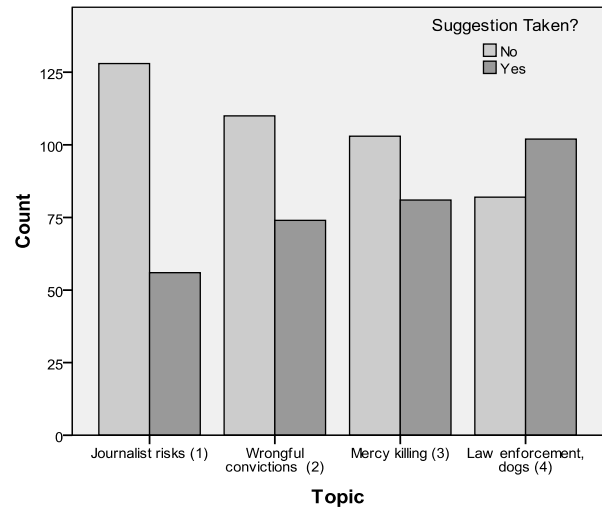

Figure 7. Number of suggestions taken per topic. Number in parenthesis next to each topic number is the difficulty rank according to subjects' ratings (where 1=easiest; 4=hardest).

We further investigated whether subjects were more likely to select queries that were associated with high social information than low social information with respect to topic difficulty. In this analysis, we only consider those suggestions that subjects selected. Figure 8 shows the distribution of subjects' selections. Although subjects selected more unpopular queries than popular ones for Topic 426, which was the most difficult topic, these differences were not significant [$\chi^2(3, 313)$=2.05, p=.56]. When we only focus on the suggestions that subjects took and examine the odds of a subject selecting a popular suggestion, it appears for almost all topics the odds of selecting a popular suggestion is relatively high: journalist risks (1.333), wrongful convictions (1.242), mercy killing (1.189) and law enforcement, dogs (.888). Interestingly, the lowest odds are associated with the most difficult topic.
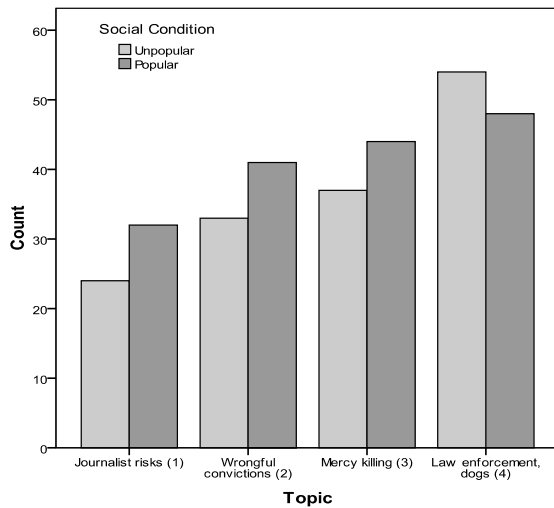
Figure 8. Number of popular and unpopular query suggestions taken per topic.

Figure 9 shows the distribution of subjects' selections according to query quality. Overall, subjects selected more good quality queries than bad quality queries even for the most difficult topic. These differences were not significant [χ^2(3, 313)=1.22, p=.75]. Odds were computed to examine the likelihood of subjects selecting a high quality query for each topic. The odds were: journalist risks (.931), wrongful convictions (1.312), mercy killing (1.314) and law enforcement, dogs (1.170). The lowest odds were associated with the easiest topic, while the highest odds were associated with topics of medium level difficulty.
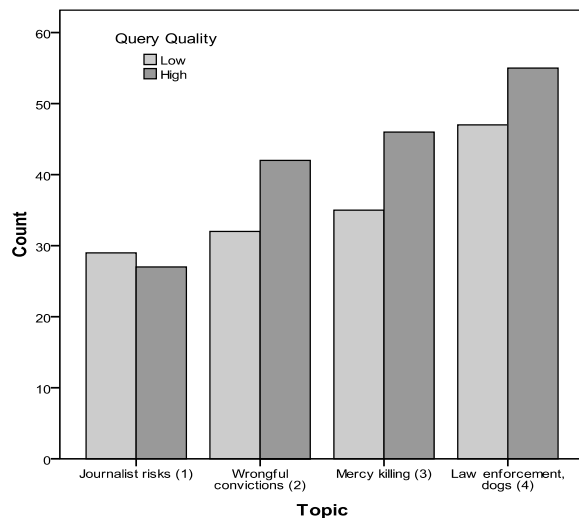

Figure 9. Number of high and low quality query suggestions taken per topic.

Feedback from Subjects

On average, subjects indicated that they were somewhat familiar with social search features (*M*=3.13, *SD*=0.97, where 1=very unfamiliar and 5=very familiar). Few subjects used the middle value to describe their familiarities. Six subjects used 1 or 2 to describe their

familiarities, while 15 subjects used 4 or 5. Nearly half of all subjects (48%) indicated that they have used social search features too many times to remember. Five (22%) indicated three or four times, 6 (26%) indicated once or twice, while 1 indicated never. Overall, most subjects were familiar with social search features and had used them in the past. Most subjects described their experiences with social search features as good and positive. Many subjects stated that in the past they have used social search features for academic research tasks, such as conducting research for a class paper. In addition, a few subjects stated they have used social search features during online shopping and one subject stated that he uses social search features for distraction and as a procrastination technique.

All subjects, except one, reported they selected some of the suggested queries in this study, which is in agreement with the log data. In response to an exit questionnaire item that asked subjects to indicate the extent to which they agreed that the information about how many other people used the different queries determined whether they used the suggestions (where 1=strongly disagree; 5=strongly agree), most subjects selected 4 (35%), followed by 2 (26%), 3 (22%) and 1 (17%). No subject selected 5 (strongly agree). When asked in the exit interview how important the social information was in their decision to use a query suggestion, most subjects reported the social information was unimportant and a few subjects indicated that they ignored it. These results seem to contradict subjects' responses to the quantitative item. Subjects' comments also indicated that they performed some local testing to determine the extent to which the social information could be trusted.

Most subjects reported that they find social search features useful, especially when they need a starting point to enter keywords for a search or for shopping tasks. One subject reported that reviewing suggested queries was a good way to stimulate "thinking outside the box." Another subject stated that social search was a useful method to narrow the search, while another indicated that suggestions were especially useful in search situations where she could not think of any keywords. One subject reported that social information was not important because she was uninterested in what other people thought.

DISCUSSION AND CONCLUSION

In this study, subjects selected a large number of query suggestions during their searching, although their choices of which queries to select were not influenced by the social information associated with the queries. While subjects selected more query suggestions that were associated with high social information, this difference was not significant. Subjects did, however, select significantly more high quality query suggestions than low quality query suggestions and in their post-search evaluations of query quality they rated good quality queries as significantly better than low quality queries. These results seem to suggest that subjects relied more on their own judgments

than the judgments of previous users and could distinguish between good and bad quality query suggestions.

Our results differ slightly from previous studies. One potential reason for this is the type of task in which subjects engaged. The task was cognitively intense and because of this, subjects may have been more critical of social information. One subject stated that for academic search tasks (of which he considered the experimental task to be an example) he relies less on social information and more on his own knowledge. In response to an interview question about the importance of the social information, most subjects maintained that this information did not influence them and some even indicated they ignored it (this is similar to Suchanek et al.'s [25] results). These responses were slightly contradictory to subjects' responses to the closed question about the impact of the social information on their choices of which query suggestions to take. The results seem to indicate that subjects may have noticed the information, but that ultimately, they used their own judgments. Given that these subjects were undergraduate students, this is a nice finding for it suggests that they may be more sophisticated consumers of information than what is generally thought. However, it may also be the case that subjects felt more pressure to appear as independent thinkers and therefore responded in a way that made it seem as if the social information was unimportant to them.

It is important to distinguish between the usefulness of the suggestions and the usefulness of the social information: most subjects indicated that the query suggestions were useful when they ran out of ideas or faced a cold-start problem, and overall, subjects took a lot of suggestions. Another important finding was that subjects took more query suggestions when searching for difficult topics. Although we have not analyzed subjects' search performances yet, it may be the case that the query suggestions helped subjects be more successful. It is also the case that our study was not set-up to directly compare the effectiveness of query suggestions for tasks of varying levels of difficulty. Instead, topic difficulty was a quasi-independent variable since it was generated by subjects rather than controlled by the researchers. Future studies should use topic difficulty as an independent variable.

Overall, these results provide some evidence that query suggestions are a useful type of idea tactic and can assist subjects during information seeking. First, subjects integrated the suggestions into their searching fairly quickly. In about half of the cases, the second query submitted by a subject was a suggestion (this number grows if we consider manually entered queries that duplicated a suggestion). Subjects took significantly more suggestions when searching for difficult topics and expressed in the interviews that the suggestions stimulated their thinking and, in particular, "thinking outside the box." This sentiment is aligned with Bates' [3] notion of idea tactics. One of the primary goals of idea tactics, Bates stated, is to help searchers break-out of their current way of thinking

about a search problem. The query suggestions seemed to do this at least for some subjects. Furthermore, subjects commented that the suggestions helped them when they exhausted their own ideas, again demonstrating support for the notion of query suggestions as a type of idea tactic. Subjects also took significantly more suggestions for difficult topics, which provides evidence that query suggestions are most useful when subjects face difficulties when searching. This result is aligned with Vakkari's [28] findings that topic difficulty can influence searchers' choice of search tactics, and in particular, that searchers might move from a search approach where they enter queries when searching is easy to one where they take suggestions when searching becomes more difficult.

One interesting challenge to evaluating the extent to which interface features can support idea tactics is defining and capturing usage of these features. In this study, we used selection (clicking) as a measure of usage, but after analyzing subjects' manually entered queries found many queries (n=106) that duplicated a suggestion. As stated earlier, it might be the case that subjects entered these queries naturally. It might also be the case that subjects preferred typing to clicking. Regardless, it can be argued that subjects' usage of the suggestions is not completely detectable with click-through data. Eye-tracking data can provide some insight into whether subjects gaze at the suggestions and future research might incorporate this as a data collection technique, but even then it is difficult to determine if the suggestions helped cognitively. Subjects may select particular terms from the suggested queries or the suggested queries may give subjects ideas for whole new queries. The suggestions might also help subjects gain a better understanding of the different facets of the topic.

It is difficult to determine which theory of social behavior and influence provides the most useful lens for interpreting the results of this study. The results of this study suggest that task type mediates the effects of social influence on information seeking behavior. Future studies can be designed to explicitly test this relationship by including additional task types. These theories also provide interesting ways to proceed with future research. For example, much work has examined how traits, states and situational variables such as uniqueness [15] and resistance [9] mediate the effects of social influence. Ludford, et al. [21] has shown that uniqueness can positively impact people's participation in online discussions; providing information about uniqueness and consensus might be used to create more diversity in social search features. Finally, many theories of social influence were developed in the context of group decision-making or in situations where people's choices were made in the presence of others. Application of these theories to new search situations, such as collaborative information seeking (where groups of people work together on the same task) may prove useful.

REFERENCES

1. Banerjee, A.V. A simple model of herd behavior. *Quarterly Journal of Economics 107,* 3 (1992), 797-817.

2. Baron, R. S., Vandello, J. A. & Brunsman, B. The forgotten variable in conformity research: Impact of task importance on social influence. *Journal of Personality and Social Psychology 71*, 5 (1996), 915-927.

3. Bates, M. J. Idea tactics. *Journal of the American Society for Information Science, 30* (1979), 280-289.

4. Belkin, N. J., Cool, C., Stein, A., & Thiel, U. Cases, scripts, and information-seeking strategies: on the design of interactive information retrieval systems. *Expert Systems with Applications 9*, 3 (1995), 379-395.

5. Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom and cultural change as informational cascades. *Journal of Political Economy 100, 5* (1992), 992-1026.

6. Chen, Y.-F. Herd behavior in purchasing books online. *Computers in Human Behavior 24*, (2008), 1977-1992.

7. Cialdini, R. B. *Influence: Science and practice* (4th ed.), Boston, MA: Allyn & Bacon, 2001.

8. Cialdini, R. B., & Goldstein, N. J. Social influence: Compliance and conformity. *Annual Review of Psychology 55*, (2004), 591-621.

9. Cosley, D., Lam, S. K., Albert, I., Konstan, J. A., & Riedl, J. Is seeing believing? How recommender interfaces affect users' opinions. *Proc. CHI 2003*, ACM Press (2003), 585-592.

10. Deutsch, M. & Gerard, H. B. A study of normative and informational social influences upon individual judgment. *Journal of Abnormal Psychology 51*, (1955), 629-636.

11. Evans, B. & Chi, E. H. Towards a model of understanding social search. *Proc. CSCW 2008*, ACM Press (2008), 485-494.

12. Freyne, J., Farzan, R., Brusilovsky, P., Smyth, B., & Coyle, M. Collecting community wisdom: Integrating social search & social navigation. *Proc. IUI 2007*, (2007), 52-61.

13. Hanson, W. A. & Putler, D. S. Hits and misses: Herd behavior and online product popularity. *Marketing Letters 7*, 4 (1996), 297-305.

14. Hill, W., Stead, L., Rosenstein, M., & Furnas, G. Recommending and evaluating choices in a virtual community of use. *Proc. CHI 1995*, ACM Press, (1995), 194-201.

15. Imhoff, R., & Erb, H.-P. What motivates nonconformity? Uniqueness seeking blocks majority influence. *Personality and Social Psychology Bulletin 35*, (2009), 309-320.

16. Järvelin, K. & Kekäläinen, J. Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems (TOIS) 20*, (2002), 422-446.

17. Joachims, T., Freitag, D., & Mitchell, T. WebWatcher: A tour guide for the World Wide Web. *Proc. ICML* (1997), 770-777.

18. Joachims, T., Granka, L., Pan, B., Hembrooke, H. & Gay, G. Accurately interpreting clickthrough data as implicit feedback. *Proc. SIGIR 2005*, ACM Press (2005), 154-161.

19. Keane, M. T., O'Brien, M. & Smyth, B. Are people biased in their use of search engines? *CACM 51,* 2 (2008), 49-52.

20. Kelly, D., Gyllstrom, K., & Bailey, E. W. A comparison of term and query suggestion features for interactive searching. *Proc. SIGIR 2009,* ACM Press, (2009), 371-378.

21. Ludford, P. J., Cosley, D., Frankowski, D., & Terveen, L. Think different: Increasing online community participation using uniqueness and group dissimilarity. *Proc. CHI 2004*, ACM Press (2004), 631-638.

22. Rook, L. An economic psychological approach to herd behavior. *Journal of Economic Issues XL*, 1 (2006), 75-95.

23. Senecal, S. & Nantel, J. The influence of online product recommendations on consumers' online choices. *Journal of Retailing 80*, (2004), 159-169.

24. Smyth, B., Balfe, E., Freyne, J., Briggs, P., Coyle, M., & Boydell, O. Exploiting query repetition and regularity in an adaptive community-based Web search Engine. *User Modeling and User-Adapted Interaction 14*,5 (2004), 382-423.

25. Suchanek, F. M., Vojnovic, M., & Gunawardena, D. Social tags: Meaning and suggestions. *Proc. CIKM 2008*, ACM Press (2008), 223-232.

26. Surowiecki, J. *The wisdom of crowds*. New York, NY: Doubleday, 2004.

27. Svensson, M., Höök, K., Laaksolahti, J., & Waern, A. Social navigation of food recipes. *Proc. CHI 2001*, ACM Press (2001), 341-348.

28. Vakkari, P. Changes in search tactics and relevance judgments when preparing a research proposal: A summary of the findings of a longitudinal study. *Information Retrieval 4,* (3-4) (2004), 295-310.

29. Voorhees, E. M. Overview of the TREC 2005 Robust Retrieval Track. *Proceedings of 14th Annual Text Retrieval Conference, (TREC-14)*, 2006.

30. White, R. W., Bilenko, M., & Cucerzan, S. Studying the use of popular destinations to enhance web search interaction. *Proc. SIGIR 2007*, ACM Press (2007), 159-166.