

Perception and Understanding of Social Annotations in Web Search

Jennifer Fernquist
Google, Inc.
Mountain View, CA 94043 USA
jenf@google.com

Ed H. Chi
Google, Inc.
Mountain View, CA 94043 USA
edchi@google.com

ABSTRACT

As web search increasingly becomes reliant on social signals, it is imperative for us to understand the effect of these signals on users' behavior. There are multiple ways in which social signals can be used in search: (a) to surface and rank important social content; (b) to signal to users which results are more trustworthy and important by placing annotations on search results. We focus on the latter problem of understanding how social annotations affect user behavior.

In previous work, through eyetracking research we learned that users do not generally seem to fixate on social annotations when they are placed at the bottom of the search result block, with 11% probability of fixation [22]. A second eyetracking study showed that placing the annotation on top of the snippet block might mitigate this issue [22], but this study was conducted using mock-ups and with expert searchers.

In this paper, we describe a study conducted with a new eye-tracking mix-method using a live traffic search engine with the suggested design changes on real users using the same experimental procedures. The study comprised of 11 subjects with an average of 18 tasks per subject using an eyetrace-assisted retrospective think-aloud protocol. Using a funnel analysis, we found that users are indeed more likely to notice the annotations with a 60% probability of fixation (if the annotation was in view). Moreover, we found no learning effects across search sessions but found significant differences in query types, with subjects having a lower chance of fixating on annotations for queries in the news category. In the interview portion of the study, users reported interesting "wow" moments as well as usefulness in recalling or re-finding content previously shared by oneself or friends. The results not only shed light on how social annotations should be designed in search engines, but also how users make use of social annotations to make decisions about which pages are useful and potentially trustworthy.

Categories and Subject Descriptors

H.5.m. [Information Interfaces and Presentation (e.g. HCI)]: Miscellaneous

Keywords

Annotation; social search; eyetracking; user study.

Copyright is held by the International World Wide Web Conference Committee (IW3C2). IW3C2 reserves the right to provide a hyperlink to the author's site if the Material is used in electronic media. *WWW 2013*, May 13–17, 2013, Rio de Janeiro, Brazil. ACM 978-1-4503-2035-1/13/05.

1. INTRODUCTION

As more of the web involves social interactions, they produce a wealth of signals for users to utilize in searching for the most interesting and relevant information. The abundance of information suggests the importance of creating an environment in which users have the appropriate signals to make decisions on which search results are the most useful. Many search companies as well as social media sites are making investments based on the assumption that these social signals will greatly improve users' ability to make decisions online.

Utilizing these signals requires the investigation of three different research problems. (1) First, different users have different visibility to these social signals, so we must respect their privacy boundaries. (2) Second, these signals should be incorporated into the search engine's ranking algorithms using user modeling and personalization techniques. (3) Third, we need to display the appropriate annotations and explanations to the users so that they can make an informed decision about the quality of the information.

In the context of web search, there is a huge cost in curating these social signals, encrypting and storing them along with a public web crawl, as well as ranking and serving the personalized results. Much research has been done on modifying search ranking based on social information about result pages [4, 7, 14, 32, 33], but less has been done on the best way to present social information in web search results.

Here we use the term *social signals* to refer to any social information that is used to affect ranking, recommendations or presentation to the user. We use the term *social annotations* to refer to the presentation of social signals as explanations for why a search or recommendation result is being displayed. Thus, a social signal only becomes an annotation when it is presented to the user. Other researchers in the social recommendation engine literature have used the term "social explanations" or just "explanations" instead.



Figure 1: An example of a personal search result with three characteristics to indicate it's a personal result: the blue icon (a), the text annotation above the search snippet (b), and a profile picture (c).

For this paper, we are mainly interested in understanding the effectiveness of social annotations in web search and how users re-

spond to them. An important research question is whether users are actually making good use of these annotations. Figure 1 shows an example of these annotations as rendered by Google Search.

In previous work, Muralidharan et al. in an eyetracking study found that users did not generally notice these social annotations¹, with only 11% chance of fixation [22]. However, when asked in a retrospective interview, subjects reported that close relationships and knowledgeable contacts would be useful social annotations to them. Moreover, they stated that annotations would be useful for domains such as shopping, restaurants, and other social and subjective topics.

The question is, therefore, why do subjects not notice these annotations, especially since they might be useful? In a follow up Wizard-of-Oz study, Muralidharan et al. found that subjects structurally parsed the search result page using a consistent element ordering, and by moving the annotations to the top of the snippet block, there is great potential for improving the chance that users will fixate on the annotation. An important follow up question is whether the results will hold up when these changes are implemented in a real search engine.

In this paper, we conducted a new eyetracking mix-method study using a live traffic search engine with the suggested design changes on real users and their data. The study had 11 subjects with an average of 18 tasks per subject and used an eyetrace-assisted retrospective think-aloud protocol. We replicated the same experimental conditions as the previous paper and significantly extended the analysis from previous work. The contributions of this paper are:

- Using a funnel analysis, we found that users are indeed more likely to fixate on the annotations, at 60% probability of fixation, if the annotation was in view.
- We found no learning effects across search sessions but found significant differences in query types, with subjects having a lower chance of fixating on annotations for queries in the news and entertainment categories.
- In the interview portion of the study, users reported interesting “wow” moments as well as finding social annotations useful for recalling or re-finding content previously shared by oneself or others.

The results shed light on how social annotations should be presented in search results and on how users make use of social annotations to make decisions about which pages are useful and potentially trustworthy.

2. PREVIOUS WORK

2.1 Social Resources During Search

Researchers use the term ‘social search’ for several different distinct types of help that users receive from others in performing a search task. What is common between them, however, is the sense of opportunity to utilize social resources in order to help searchers

¹We use the terms notice and eye fixation somewhat interchangeably in this paper, consist with Muralidharan et al.’s work. Moreover, noticing the annotation and actually utilizing it is related but not the same. In reading, noticing and attention requires fixation, but fixation does not necessarily mean attention. Here we are observing *fixation* and interpreting the fixation somewhat generally as *noticing* the annotation but not necessarily as *utilizing* the annotation. The difference between understanding fixation vs. noticing vs. utilizing point to the depth of the problem we are trying to solve.

be more productive or examine more alternatives. Users naturally seek others for help when they are stuck. For example, Evans and Chi [11] conducted a large scale survey showing that, when searchers fail during a task, they seek greater involvement of experts as well as close ties such as family members and co-workers.

Researchers have examined many ways in which to present social signals and other information to users while they are attempting to search for information [17, 19, 28, 20, 24]. Examples of approaches taken include sharing keywords [17], search or browsing histories [13] from users who have conducted similar searches and enabling communication connections [19] between users via chat, email or voice.

One area of research focuses on synchronous collaboration between users during search. ‘Collaborative search’ has been extensively studied [19, 18, 3, 28, 20, 27, 31]. Much work in collaborative search, however, investigates two or more users carrying out the same search task synchronously, either co-located or distributed, and does not consider the lone searcher who is carrying out her own search task but who may derive benefit from the insights of others who have carried out a similar task in the past. Moreover, finding others willing to collaborate at the same time seems to be a barrier to wider adoption of this technology.

Another thread of research on social search is closely related to social question-and-answer (QA) systems, like Quora [30], that allow users to ask questions to a larger community, or Aardvark [15, 8], which connected users to individual members to whom they could ask a question. Pioneering work done by Ackerman shows the potential of social QA systems [1, 2]. Bernstein et al. [5] created a method for curating and displaying social answers for long tail web queries, presenting them as text just below the search bar and above the first search result. While the information in the answers have been crowdsourced, the answers are displayed as plain text and look as though they come from a single, potentially authoritative website, with no indication of how it was chosen. Academic research into building these systems shows the need to understand incentive structures to motivate others to answer questions [16, 24].

Morris, Teevan, and Panovich [21] characterized the types of questions that Facebook users ask each other. Similarly, Paul, Hong, and Chi investigated the use of social networking websites like Twitter to examine how users ask each other questions and the answers they receive [25]. They found that the most popular question types were rhetorical and factual and most questions did not receive a response, suggesting that queries in Twitter are not salient enough to be noticed.

One reason people turn to social networks to ask questions is because they recognize the value of expertise. Expertise has been shown to be an important factor when people decide how and who to ask for information [6, 12]. Paul, Hong, and Chi found that, on Quora, users judge expertise and reputation of answers based on previous contributions [26]. Further, Nelson et al. [23] found that expert annotations improved learning for exploratory tasks.

It is perhaps these expert annotations and signals, left over from others, that are most useful for social search since they are social signals that we can take advantage of directly in web search. For example, MrTaggy [17] presents useful keywords related to users’ search query while they carry out their task, which are derived from social bookmark tags. The tags are derived from Delicious [10], a social bookmarking web service that allows users to save and share websites. Morris et al. [19] also found that metadata associated with a webpage, like visitation history, was useful to users when carrying out the same task as a partner.

2.2 Social Annotations & Result Parsing

We have found some research for understanding how social annotations and signals are useful and utilized by searchers. Also through eye-tracking analysis, Cutrell and Guan [9] found that using longer search snippets significantly improved performance when users were conducting information-seeking tasks (e.g. the average temperature in August in San Francisco), but degraded performance for navigational tasks.

As mentioned above, we build on top of recent eyetracking research done by Muralidharan et al. [22], which in their first experiment showed, surprisingly, that searchers did not generally make use of social annotations, despite being embedded directly within the snippet blocks of search results. Importantly, in a follow-up experiment using mock-ups, they showed that by placing the annotations above the snippet block, subjects seem to pay more attention to them. However, this was never replicated in a real search engine with real data from real users.

In interpreting the results from the eyetracking study [22], Muralidharan et al. believed users perform visual structural parsing of the search result page in an interesting way. Specifically, users typically perform a linear scan through the search results from top to bottom. While performing this scan, they first look at the title and URL of each result to determine if the result reaches a certain threshold of relevance. If so, users click on the result to examine the details. If not, then users either move on to the next search result or they read through the result snippet block in reading order. Thus, consistent with this interpretation, if the social annotation is placed at the top of the snippet block, then users will fixate on it more readily than if the annotation is placed at the bottom. Experiment 2 in the paper showed evidence of this parsing behavior. Moreover, the paper concluded that an important area future work is to verify this finding using real users with real organic search tasks using their social data.

In this paper, we have conducted new eyetracking research to test the perception of social annotations with the design suggested by Muralidharan et al., with real social data from users in natural search contexts. We also explore if there are any learning effects with social annotations in search results; that is, as users begin to notice the annotations, do they then start to look for them in search results? Finally, we sought to understand for which query types social annotations are noticed more. Perhaps social annotations are more useful for some query types than others, which would be consistent with the finding from Cutrell and Guan [9] that explanatory information is more useful for informational tasks than navigational tasks.

3. RESEARCH GOALS

Overall, we wanted to investigate situations in which social annotations influence search behavior: (1) First, we wanted to identify how the visual design of social annotations affects the probability of users fixating on the annotation and how frequently. (2) Second, we were interested in whether subjects learned to look at social annotations more over the hour-long experiment, given its potential usefulness. (3) Finally, we wanted to identify and categorize types of tasks and queries that are conducive to annotations or recommendation explanations for search queries.

Given that the previous study was partially performed using simulated search result pages, we aimed to replicate the results in a real context with real users' social signals. Consistent with previous findings, we surmised that real users with organic search results will better notice the social annotations if they are placed at the top

rather than the bottom of the snippet block (below the result title and URL).

Moreover, we wanted to determine if there are learning effects over the one-hour experimental session. Presumably as the users notice more of the annotations, they will start to look for them as possible signals for result relevance. We hypothesized that users will notice more of the annotations later in the session.

Finally, we surmised that social annotations are more useful for certain query types. For example, users should find annotations for subjective queries useful, such as those relating to shopping and restaurant reviews. For goal-oriented queries, such as fact-finding, social annotations might be less useful. Therefore, we also aimed to observe subjects' search process in the presence of social annotations across search tasks of different categories, such as navigation, fact-finding, news, local services, and shopping.

Stated more formally, we had three research hypotheses:

H1: Placement Design Annotations placed at the top of the snippet block will be perceived significantly more than those with previous designs where they are placed at the bottom.

H2: Short-term Learning Subjects will learn to appreciate the annotations during the experimental session, with higher frequencies of noticing them as they learn that they are useful.

H3: Query Types Subjects will notice the annotations more often for certain query types because those social signals are more useful to them when performing those corresponding tasks.

4. METHOD

We conducted a laboratory eye-tracking study with retrospective think-aloud protocol using the same procedure as experiment 1 in Muralidharan et al.'s paper [22].

Unlike the follow-up experiment 2 in that paper which used mock-ups and employees as subjects, we used real users with their organic personal search results blended in with the regular, live search results using a personalized ranking function.

We designed around 16 custom tasks for each individual subject, according to a fixed search task taxonomy. In all, we were able to collect eyetrace data for about 153 tasks across all subjects.

4.1 Subjects

We recruited 12 subjects through a subject pool for our study. Subjects were each compensated with a \$75 AmEx gift card. We filtered the subjects for active Google+ usage, with subjects interacting with the site several times per week (5) or at least once a day (4). Two subjects had 11-50 people in their Google+ circles, four had 51-100, and three had more than 100.

Unfortunately, one subject failed to show up for the study. For two of the subjects we had equipment failures that prevented us from capturing eye trace data, although we still used their interview data. Out of the nine subjects that we used for eye-trace analysis, four were female.

4.2 Materials

All sessions were conducted in an usability lab using a Tobii T60 eye-tracker.

Each subject had consented to allow us temporary access to their account so that we could design the custom search tasks that would trigger social and personalized search. They did not, however, know the purpose of research nor that we were designing personalized tasks for them.

4.3 Tasks

Using information from Aardvark's taxonomy of social Q&A behavior [15], we designed the search tasks for each subject so that

the tasks spanned some variety. We used their taxonomy because it was one of the most comprehensive categorizations of questions that people are likely to ask in a social search service. Figure 2 shows the top question categories posed to the Aardvark community.

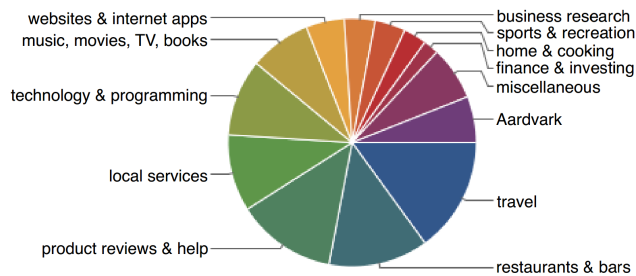


Figure 2: Categories of questions posed to the Aardvark community. Image from Horowitz and Kamvar [15].

We planned 16-20 search tasks for each subject, at least eight of which were “social search” tasks designed to organically pull up results with social annotations. The search tasks were designed to span at least 5 of the following categories:

- shopping
- local services
- how-to / recipes
- news
- navigation
- fact-finding
- entertainment

In order to ensure that personal results (relevant search results with social annotations) appeared for as many queries as required, we designed 2-4 additional social search tasks for each subject that were intended to bring up personal results. This way, if one social search task did not bring up personal results, we gave them the additional tasks to help ensure that they saw 8 tasks with personal results.

4.4 Procedure

Experimental sessions consisted of 3 parts, essentially using the same script as experiment 1’s procedure in Muralidharan et al. [22]:

4.4.1 Part 1: Search tasks

Subjects performed the 16 tasks and additional search tasks if necessary as described above. Search tasks were dictated to subjects, rather than given to them written on paper, to prevent them from entering the search task verbatim into the search box.

They were instructed to perform the task as they would at home, without performing a think-aloud or describing their actions, in order to make the behavior as natural as possible.

4.4.2 Part 2: Retrospective think-aloud

After the search tasks, we immediately conducted a retrospective review of eye-traces for those search tasks in which subjects exhibited behaviors of some interest to the experimenter. Review of eyetracking videos prompted think-aloud question/answering about subjects’ process on the entire task, particularly interesting pages, and interesting results.

During this part of the study, we often drew subjects’ attention to personal results and their social annotations, and asked them questions about their reactions to those annotations via the eyetrace data. The general script for the questions are as follows:

- Talk me through how you carried out this search task.
- Why did you click on that result?
- Did you see these (picture, annotation, icon)? What do you think they are?
- Do you know who that person is?
- Is this person a friend, co-worker, acquaintance, distant acquaintance... ?
- Do you think that person is a good reference for this query?
- What is your reaction to seeing this person’s name next to this query?
- How would that influence the likelihood of you clicking on this result?
- Are there any queries you might do where results from that person would be particularly useful?

4.4.3 Part 3: Think-aloud tasks

Finally, subjects performed two or three different search queries for which we determined ahead of time should bring up relevant personal results. For most subjects we asked them to enter a search query and then click on the ‘n personal results’ link at the top of the page, where n is the number of personal results that exist for that query, so that the entire page contained personal results.

With these personalized results present on the search page, we asked:

- How many of these people do you know?
- Are these people friends, family, acquaintances, co-workers, distant acquaintances?
- If you were explaining these results to someone else, what would you say?
- Do you see anything surprising here?
- Is there anything you’re interested in clicking on?
- Would you find it useful to see these types of results from friends?
- How might you use this feature?
- What types of searches do you think would pull up more [personal] results?

Next we present the quantitative and qualitative results in turn, focusing on specific data points that help us answer the research questions presented earlier.

5. QUANTITATIVE RESULTS

5.1 Funnel Analysis & Placement Design

In total, we collected eye-trace data for 153 tasks from nine subjects ($m=17.5$ tasks per subject, $range=[16-20]$). Each one of these eye-trace data for each task was analyzed by hand by an experimenter to understand how subjects performed the search task by formulating search queries; recording the search query used by the subject; which positions contained personal search results; whether the search result was in the field of view in the browser; and importantly, whether the subject fixated on the result and/or the social annotation. This pain-staking work was carried out for all 153 trials across all 9 subjects and then the results tallied for the analysis below.

See Figure 3 for the details of the funnel analysis. The detailed results from this analysis are:

- Of the 153 tasks, we designed 83 (53%) to trigger personalized social search results, 8-10 tasks for each subject.
- During the experiment, 66 of these 83 (80%) search tasks had at least one personal result in the result set, ranging between 4 and 10 personally annotated tasks per subject (mean=6.2 tasks).

Participant	Tasks	Tasks that should contain personal results	Tasks that did contain personal results	Personal results served	Personal results in view	Personal results noticed
P1	16	8	4	4	1	0
P2	17	9	8	10	6	6
P3	18	8	6	8	6	6
P4	17	9	7	8	4	4
P5	17	9	6	7	3	2
P6	18	10	9	15	11	4
P7	18	10	9	15	11	7
P8	17	10	10	13	9	4
P9	20	10	7	8	7	2
Total	158	83	66	88	58	35
%		53%	80%		66%	60%

Figure 3: Funnel analysis showing (from left to right) for each subject: the number of tasks carried out; the number of tasks designed to have personal results; the number of tasks that actually had personal results; for those tasks, the number of personal results that were served; of those, the number of personal results that were visible to the subject; and of those, the number of personal results in which subjects fixated on the social annotation.

- These 66 tasks served up a total of 88 personal search results with social annotations (range=[4-10] results per subject, mean=9.8).
- Of these 88 personal results, by looking at the eyetrace data, 58 out of 88 (66%) were either above the fold or the user scrolled to them (range=[1,11], mean=6.4). In other words, 58 annotations in our experiment had a chance of subjects fixating on them. For the remaining tasks with personal results, 21 (24%) were below the fold and never scrolled to, while 9 (10%) were not on the first page of results and not clicked to.
- By looking at the eyetracking data in each one of the 58 annotations, we found that 35 (60%) were actually fixated on by the subjects (range=[0-7] noticed annotations, mean=3.9). By 'noticed' we mean that, while reviewing the eye gaze data, subjects fixated at least once on an annotation element (such as person name, picture, or blue avatar indicator.) Please see Figure 1 for clarification on the different elements of an annotation.

A 60% chance of fixation is a dramatic improvement from the results found previously [22]. The previous study (presented as experiment 1 in [22] using exactly the same experimental procedure described here) was conducted in August 2011. That experiment used designs which placed the annotations *below* the snippet block. In that study, out of 45 annotated results that were visible, subjects fixated on the annotations for only 5 of them (11%). A placement change increased the probability of noticing the annotation by 5.5 times!

It is worth noting that 8 out of 9 subjects noticed some attribute of the personal results for at least one of the tasks. Only one subject did not fixate on any attribute of any social annotations. Half noticed the blue icon marking a personal result (6 total fixations), 7 noticed the picture (18 total fixations), and all noticed the name (22 total fixations). Heatmap images showing the fixations for two tasks are shown in Figures 4 and 5.

We expected subjects to fixate on peoples' pictures more than names due to their saliency, yet our data suggests that pictures and names are noticed at about the same frequency (18 name fixations versus 22 picture fixations). This may be due to the fact that sub-

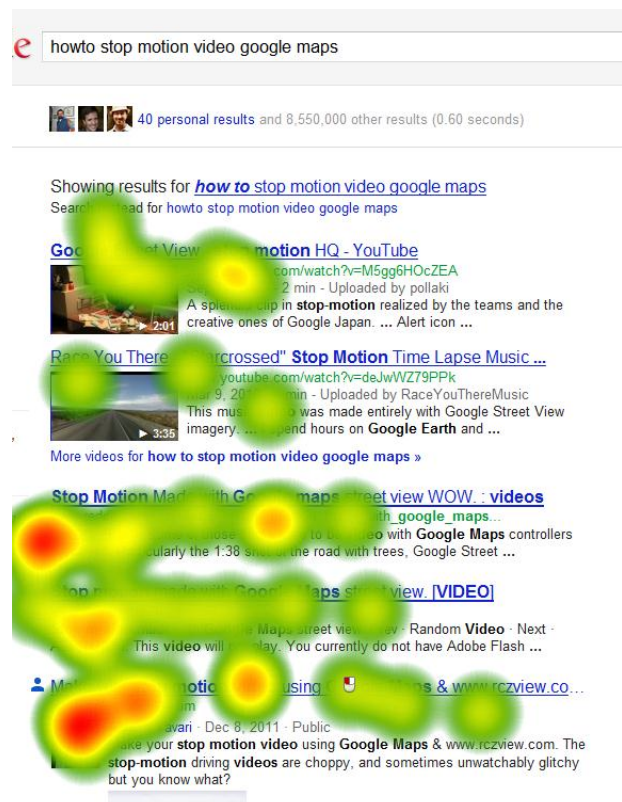


Figure 4: Heatmap generated from eyetracking data showing a subject's gaze pattern. The subject looked briefly at the first two results and then moved her eyes to results 3-5, reading the titles and URLs. She fixated on the social annotation (profile picture and text annotation) before clicking on the personal result.

jects didn't recognize their friends/colleagues in the pictures (e.g. new or unfamiliar picture, image too small). It may also be the case that, since subjects were scanning text on the results pages to find the link or answer they were looking for, they read the text in the annotations before glancing at the pictures.

An example of a gaze pattern taken from eyetracking data is shown in Figure 6. The gaze pattern demonstrates that the visual parsing strategy used by this subject is consistent with the strategy posited by Muralidharan et al., described above. That is, a vertical linear scan through the search results, using URLs and titles to make decisions about relevance, and looking in the snippet block for evidence to click through.

5.2 Short-term Learning Effects

One question about this improvement is whether it occurs due to learning effects. That is, perhaps at first users do not really notice the annotations, but later, realizing that annotations are indeed useful, users may alter their parsing strategy to pay more attention to them.

We examine how users might learn over time during the hour-long experimental session. We surmised that as they encounter the annotations organically in the session, they will find them useful and start to rely on them more and more.

For each subject, we first ordered tasks according to the order they were received during the hour-long study. We tabulated both the number of personal results served for that task number (Fig-

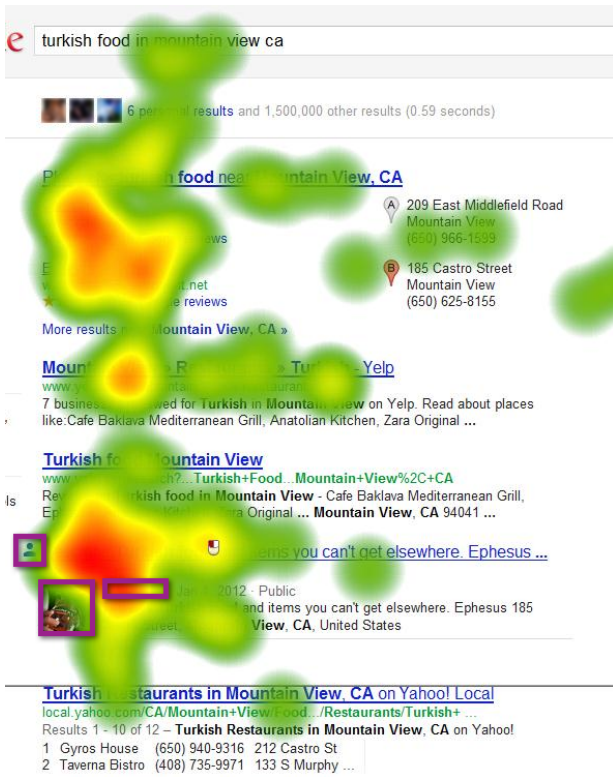


Figure 5: Heatmap generated from eyetracking. The three personal result characteristics are highlighted. This subject looked primarily at the title, URL and text annotation of the personal result.

ure 7), as well as the number of times the subject fixated on an annotation during that task. The results of this tabulation is shown in Figure 8.

As might be seen in Figure 8, we did not find a learning effect in our data, with the chance of fixating on the annotations hovering around 50% over the hour-long study. We do not know from this experiment whether longer term exposure to social annotations might change behavior, but simply an hour's worth of exposure does not seem to change the probability of noticing.

5.3 Query Type Differences

Another hypothesis we had was that subjects would notice social annotations more for *subjective* queries for informational tasks compared to *goal-oriented* queries for navigational tasks. In this section, we examine how the annotations performed across different task/query types. By using the taxonomy that was used to plan the personalized tasks, we first tabulated the results by query type categories. The result of the tabulation is shown in Table 1.

We can see that for task types such as navigation, local services, shopping and how-to, subjects had a higher chance of noticing annotations. On the other hand, the news query type had a surprisingly low chance of noticing. We were also surprised by the differences between the navigation and fact-finding categories, despite their semantic similarity.

Visually examination of the table suggests the differences between query types are significantly different from each other. However, since many of the cells in the table has expected frequencies below 5, we could not directly perform a Chi-squared test of independence, as that would violate the sample size require-

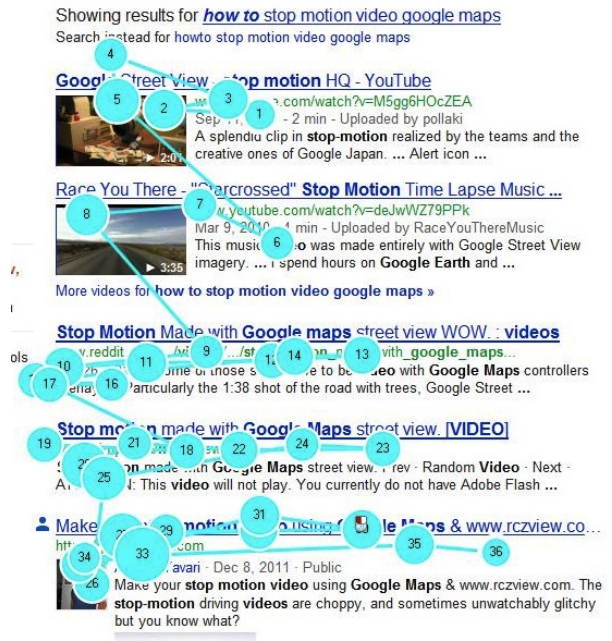


Figure 6: This image shows the gaze pattern for the personal result from the heatmap figure from Figure 4. First, the image depicts the subject's vertical linear scan of a search result page, and that the bottom personal result is the one the user eventually clicked on. For the horizontal scan of the personal result, we see that the subject scans the result title first, then the URL and finally looks at the social annotation.

		Participant								
		P1	P2	P3	P4	P5	P6	P7	P8	P9
First 8 tasks with personal results in view	T1	1	2	1	1	0	0	0	0	0
	T2		2	1	1	1	1	1	0	0
	T3		1	2	1	1	0	1	0	1
	T4		1	1	1		1	1	0	0
	T5			1			0	0	1	0
	T6						1	2	1	0
	T7						0	0	1	1
	T8						1	2	1	

Figure 7: This figure shows, for each subject and for the first eight tasks that had personal results in view, how many personal results were fixated on. Blank entries indicate the subject had no additional tasks with personal results in view.

Task	Personal results served	Personal results noticed	% noticed
T1	2	1	50%
T2	5	2	40%
T3	7	4	57%
T4	9	5	56%
T5	11	5	45%

Figure 8: This figure shows, for the first five tasks, the number of personal results that were served and fixated on for that task (over all subjects).

Task Category	Personal Result Noticed?		Total	% Noticed
	Yes	No		
Navigation	5	0	5	100%
Local services	7	0	7	100%
How-to	6	3	9	67%
Shopping	7	4	11	64%
Entertainment	5	4	9	56%
Fact-finding	4	6	10	40%
News	1	6	7	14%
	35	23	58	

Table 1: Chance of fixating on the annotation analyzed by query type.

ment. Instead, we created three combined semantic categories of (a) Local/Shopping; (b) How-to/Fact-finding/Navigation; and (c) Entertainment/News. We found the differences to be marginally significant ($\chi^2(2, N = 58) = 5.82, p = .054$). We combined Navigation with Fact-finding, despite their difference in probability, because of their semantic similarity. Clearly, more subjects are needed to definitively understand the differences between query types at a finer granularity.

Despite this, the results here suggest that there is query type variation and that subjects did pay more attention to social annotations for some query types, such as Local/Shopping.

6. INTERVIEW RESULTS

6.1 Expertise and Influence

During the retrospective think-aloud, 23 tasks were reviewed in total across all subjects. In 18 of the reviewed tasks, subjects knew the person from the personal result. P4 did not know anyone in the three personal results we reviewed with her.

In all but one of the 18 cases in which the subject knew the person in the annotation, the subjects said that that person was a good reference for the query. The people in the personal results that were reviewed were mostly friends (13), people they followed on Google+ (4) or former classmates (2).

When asked how the annotations might influence the likelihood of them clicking on the result, subjects had mostly positive responses. In 8 out of the 13 reviewed tasks where we asked the question, subjects said the annotation would make them more likely to click on the result, especially if the individual in the annotation was a friend or expert on the result subject.

“Highly influence. [Annotated person] is the go-to guy on the story.” - P1

[discussing personal results for a query on beer] “If [friend] were a beer drinker then maybe [I would click the result]. The others, I don’t know if they know beer.” - P1

“It’s probably a good video to watch anyways because friends recommended it.” - P10

“if [friend] shares a link, I’ll click it. ...If I think someone is cool and generally shares interesting stuff, I’ll click on it.” - P11

However, we also heard subjects say that they trust the ranking presented by Google and that the relevancy of the result is still the overriding factor in deciding to click through. For example:

“I don’t know that it would [influence me]. If I thought the result was useful anyways I’d click on it, [regardless of] whether it had been shared or had anything to do with anyone else ... It’s kinda nice to know he shared this particular one.” - P5

“Yes, I wanted to click [the personal result] more. But I might have clicked it anyways since it’s the first link.” - P6

These quotes suggest users have a nuanced way of looking at relevance and the decision of whether to click through can be multifaceted.

6.2 Query Types

For each reviewed query in the retrospective interview, we asked the subject whether they knew the person in the personal result. If they did, we categorized the topic of that query. The topics covered the following domains: tech (7), music (3), videos (3), food (2), travel (2), news (2), entertainment and social media.

We then asked subjects what other query types that person would be a good reference for. The additional query types that the people would be good references for were: tech (6), music (2), food (2), anything they +1 or share (2), news, videos, video games, design, photography, social media. P1 said he “would trust [personal] results on the subject I want if I know the [annotated person] has expertise in that area.”

During the interviews, subjects mentioned uses for personalized search results. Subjects thought they might be useful for finding a variety of information, including:

- Recommendations from friends
- Recommendations from local strangers
- Reviews
- Product searches
- Funny things
- Wedding entertainment links (one subject’s profession)

Examining this list, it is worth noting that much of what subjects mentioned were ‘subjective’ queries involving searching for someone else’s opinion.

P8 thought personal results would be useful for “Social stuff: partying, eating, stuff you do with people. Weird stuff.” This is an interesting use case that we had not anticipated. For example, at the end of the session, P11 tried to use personalized search to see what his friends were up to that weekend. His query of [weekend events] did not turn up relevant results. He then tried [January 6 weekend events] which actually brought up fewer personalized results because it had more keywords. He thought that personal search would be smart enough to understand what he meant by his query and find the relevant information, even if the keywords did not exactly match.

6.3 Re-finding information

Social search and personal results could be useful for re-finding information. Some search tasks we designed asked subjects questions that could be answered by specific articles that they or their friends had shared on Google+. Subjects had previously read some of these posts and, as they carried out the their search task, they mentioned having recently read about that topic.

For example, when P2 was given the task “How many people are still using AOL dial-up?”, he mentioned that he had read about that recently. When he conducted the search for the task, the third result was a personal result of the post he had read, and he recognized

and clicked on it. However, he did not notice any attribute of the annotation (icon/picture/name), as evidenced by the eye gaze data and the retrospective review. Ranking these results higher could be very useful when re-finding previously shared/seen posts.

6.4 Serendipitous and “Wow” moments

P2 though it was “pretty crazy” how quickly his data appeared in personal search results. He had taken a picture in our building’s lobby just before we began our session, and the picture serendipitously appeared in personal search results during his session.

P8 tried a search for [pizza san francisco] and saw some posts from the CEO of his company. He was surprised to learn that his CEO had such strong opinions about pizza.

Both of these cases illustrate the potential for “wow” moments for users, because the search engine exceeded their expectations of what is possible. These moments might quickly become part of the new user expectation and are a potential user experience win for a search engine.

6.5 Diversification and Freshness

Diversification could be important in social searches. For example, P10 wanted to see what wedding related results he could pull up from friends. As he searched for various things, he noticed that in the inline image results, 3/4 of the pictures were from the same person, even if there were in fact image results from many contacts. It seemed that he would expect results to come from a diverse range of social contacts.

Some subjects looked at the dates of search results when deciding which to click on. In some cases, personal results had been given a higher ranking even if they were quite stale. For example, P2 said that he had remembered reading a particular news story in the past that was relevant to the task, but he didn’t want to read it again; he only wanted to learn about the conclusion to the news story. In another example, during the talk-aloud section, P10 entered the query [cool], saw the results, and then remarked: “My old post is coming up first—I don’t like that.”

Care must be taken to ensure personal results only appear when highly relevant. P11 expressed concern that seeing too many social annotations could be “like noise”.

7. SUMMARY OF FINDINGS

In our study, we aimed to first verify the findings from a previous Wizard-of-Oz study which used mock-ups to test the placement of social annotations [22]. We used real users with real social contexts and custom tasks that were derived from a taxonomy of social Q&A query types. We found that subjects had a 60% chance of fixating on the annotations when they were placed at the top of the snippet block. This is a great improvement when compared to an 11% chance when the annotations are placed at the bottom of the snippet block [22]. Thus, hypothesis H1 is confirmed.

Second, we surmised that subjects would learn to appreciate the social annotations over time and make use of them more over the course of our hour-long experiment. Instead, we found no such evidence over the hour-long study, and so H2 is rejected.

Third, we hypothesized that subjects would be more likely to look for social signals for certain query types. We found some evidence of differences in the probabilities of fixating on the annotations between different query types. For example, surprisingly, subjects noticed annotations for queries in the news category less often, even though social signals might help users select more interesting news to read. In the interview portion of the study, subjects mentioned some query types more often than others. So H3 is at least partially supported by the evidence.

Interestingly, in the interviews, users also mentioned serendipitous discoveries, re-finding of information, and other potential uses of social search.

8. DISCUSSION

The results indicate that we can greatly affect how much a social annotation is fixated on by simply changing its placement. This appears to be further evidence of the visual parsing strategy used by searchers in examining search result pages. Search result snippets function as ranking and recommendation explanations. The findings suggest that the order of information within that snippet block can affect what information users will use to decide on the importance of a result.

We can make use of this design knowledge if we have some guesses as to the importance of the annotation and the likelihood that it will persuade users to click through to the result. For more important annotations, we can place them at the top of the snippet block. Less important annotations can go toward the bottom of the snippet block.

One also wonders how users came to learn how to parse the search result pages in this way. An information foraging explanation would suggest that they have optimized their searching behavior to fit the environment [29]. We are at the cusp of teaching users how to parse social search results by designing where and how the annotations will appear. This relationship between design and the user is a two-way street: with their behavior informing our design, as well as our design changing their behavior. It is quite likely that users’ behavior will continue to evolve as search engines change, and search engines will have to evolve along with them. After all, search engines are including social search results and its associated annotation only because of the belief that they will be useful to the users.

The fact that the users did not learn to pay more attention to the annotations over the course of the hour-long experiment does not necessarily mean that there is not some other longer-term learning that is happening. The results, however, do suggest that users only look for social signals when it is appropriate to the task context. Users did not start to look for these annotations more often simply because the previous search task might have returned a useful annotation. They only looked for them when the task (and the corresponding query type) suggested social signals would be useful for decision making. For example, the results suggest users make use of annotations when searching for local services and shopping results, but less so when they are searching for interesting entertainment or news articles.

The fact that users did not pay attention to social annotations during news and entertainment tasks was a surprise to us and deserves further study. Our small sample size, a limitation of our study, is due to the detailed nature of our experiment using eyetrackers. A much more expanded study using different experimental methods is needed to investigate under what situational contexts users will make use of social annotations for selecting news to pay attention to. It may be that desktop search is the wrong context, but tablet experiences are much more conducive to social annotations for news selection. Or it could be that the selection of news is deeply personal and topical, with users selecting what news to pay attention to simply based on the headlines and the content; and what makes news truly *social* is the comments made by friends and connected others.

Finally, the query type analysis from both the quantitative section and the interview section somewhat suggests that users pay attention to annotations more when the query types are more subjective, such as local, shopping, products, music, tech, food, reviews,

and general recommendations. This is somewhat inline with previous eyetracking results from Cutrell and Guan [9]. We hypothesize that personal and social results may be more useful for exploratory tasks; that is, tasks which are open-ended, ill-defined and multi-faceted and that seek to foster knowledge or inform some action. These tasks often involve a domain that the searcher is less familiar with or unfamiliar in the manner in which they might achieve their goal. One or more of the searcher's contacts may possess such knowledge that they lack and it could appear among personal results.

9. CONCLUSION

As the web becomes increasingly social, more and more information signals come from social sources. Users expect and demand to be able to search for socially-sourced information. Appropriately making use of social annotations to explain these social search results, therefore, becomes a necessity. While much work has gone into using social *signals* for search, much less is known about how users can and will make use of social *annotations* to make decisions during search sessions.

Previous research suggests the way in which users make use of annotations during search is situated within the structured visual parsing of search result pages [22]. Users linearly scan through the search results one by one, and for each result, they first evaluate the title and the URL to decide whether or not to click through. Only when they are not sure whether the result is relevant do they start to read the result snippet, looking for explanations and reasons why they should click through. In a previous experiment with mockups, subjects appear to be sensitive to the placement of social annotations within the snippet [22], with an 11% chance of fixating on the annotation when it is placed at the *bottom* of the snippet block, but it was unclear how much improvement is actually possible with real users.

We built upon this previous finding and conducted a new study using real users and their real social data in the newly launched social search service. Analyzing 153 sequences of eyetrace data with their associated tasks essentially frame by frame, we found that users had a 60% chance of fixating on the annotation when placed at the *top* of the snippet block. This dramatic improvement is consistent with the structural visual parsing interpretation.

We also examined whether subjects learned to make more use of annotations over the duration of the hour-long experiment. The results show users did not notice the annotations more later in the sessions. More research is needed to understand whether users are learning to utilize social annotations over a longer period of time.

Finally, via query type analysis as well as interviews conducted with retrospective think-aloud protocol, subjects appeared to make use of annotations for certain query types more often than others. In particular, we found users made use of annotations for local and shopping query types more than the fact-finding and news query types. We believe this suggests users make use of annotations only when the task context and situation fit with the need to look for social resources, such as subjective queries like shopping, products, and restaurant reviews.

Search has always been contextual. Each user brings different experiences and backgrounds to formulating and reformulating search terms. Their ability to interpret information as well as make sense of what they have learned also differs greatly. As web users become more social, they bring their social context to search and they expect and demand search engines to make use of this social context. Our research is a step toward this direction by investigating how social annotations affect user search behavior. Further research is needed to fully understand the situational contexts where

social search and the associated annotations are helpful to users, to conduct larger A/B tests and to analyze log data of searches with personal results. It also remains to be explored for results where social annotations exist, in what circumstances it is most appropriate to show them.

10. ACKNOWLEDGMENTS

We would like to thank Zoltan Gyongi and Matthew Kulick for their comments and encouragement in carrying out this research.

11. REFERENCES

- [1] Ackerman, M. S., and Malone, T. W. Answer garden: a tool for growing organizational memory. *SIGOIS Bull.* 11, 2-3 (Mar. 1990), 31–39.
- [2] Ackerman, M. S., and McDonald, D. W. Answer garden 2: merging organizational memory with collaborative help. In *Proceedings of the 1996 ACM conference on Computer supported cooperative work*, CSCW '96 (1996), 97–105.
- [3] Amershi, S., and Morris, M. R. Cosearch: a system for co-located collaborative web search. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, CHI '08 (2008), 1647–1656.
- [4] Bao, S., Xue, G., Wu, X., Yu, Y., Fei, B., and Su, Z. Optimizing web search using social annotations. In *Proceedings of the 16th international conference on World Wide Web*, WWW '07 (2007), 501–510.
- [5] Bernstein, M. S., Teevan, J., Dumais, S., Liebling, D., and Horvitz, E. Direct answers for search queries in the long tail. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, CHI '12 (2012), 237–246.
- [6] Borgatti, S. P., and Cross, R. A relational view of information seeking and learning in social networks. *Manage. Sci.* 49, 4 (Apr. 2003), 432–445.
- [7] Carmel, D., Zwerdling, N., Guy, I., Ofek-Koifman, S., Har'el, N., Ronen, I., Uziel, E., Yogev, S., and Chernov, S. Personalized social search based on the user's social network. In *Proceedings of the 18th ACM conference on Information and knowledge management*, CIKM '09 (2009), 1227–1236.
- [8] Chi, E. H. Who knows?: searching for expertise on the social web: technical perspective. *Commun. ACM* 55, 4 (Apr. 2012), 110–110.
- [9] Cutrell, E., and Guan, Z. What are you looking for?: an eye-tracking study of information usage in web search. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '07 (2007), 407–416.
- [10] Delicious. <http://www.delicious.com/>.
- [11] Evans, B. M., and Chi, E. H. An elaborated model of social search. *Inf. Process. Manage.* 46, 6 (2010), 656–678.
- [12] Evans, B. M., Kairam, S., and Pirolli, P. Do your friends make you smarter?: An analysis of social strategies in online information seeking. *Inf. Process. Manage.* 46, 6 (Nov. 2010), 679–692.
- [13] Farzan, R., and Brusilovsky, P. Social navigation support for information seeking: If you build it, will they come? In *Proceedings of the 17th International Conference on User Modeling, Adaptation, and Personalization: formerly UM and AH*, UMAP '09 (2009), 66–77.
- [14] Heymann, P., Koutrika, G., and Garcia-Molina, H. Can social bookmarking improve web search? In *Proceedings of the*

- international conference on Web search and web data mining*, WSDM '08 (2008), 195–206.
- [15] Horowitz, D., and Kamvar, S. D. The anatomy of a large-scale social search engine. In *Proceedings of the 19th international conference on World wide web*, WWW '10 (2010), 431–440.
- [16] Hsieh, G., and Counts, S. mimir: a market-based real-time question and answer service. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, ACM (New York, NY, USA, 2009), 769–778.
- [17] Kammerer, Y., Nairn, R., Pirolli, P., and Chi, E. H. Signpost from the masses: learning effects in an exploratory social tag search browser. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09 (2009), 625–634.
- [18] Morris, M. R. A survey of collaborative web search practices. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, CHI '08 (2008), 1657–1660.
- [19] Morris, M. R., and Horvitz, E. SearchTogether: an interface for collaborative web search. In *Proceedings of the 20th annual ACM symposium on User interface software and technology*, UIST '07 (2007), 3–12.
- [20] Morris, M. R., Lombardo, J., and Wigdor, D. WeSearch: supporting collaborative search and sensemaking on a tabletop display. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, CSCW '10 (2010), 401–410.
- [21] Morris, M. R., Teevan, J., and Panovich, K. What do people ask their social networks, and why?: a survey study of status message q&a behavior. In *Proceedings of the 28th international conference on Human factors in computing systems*, CHI '10, ACM (New York, NY, USA, 2010), 1739–1748.
- [22] Muralidharan, A., Gyongyi, Z., and Chi, E. H. Social annotations in web search. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems (CHI '12)* (New York, NY, 2012), 1085–1094.
- [23] Nelson, L., Held, C., Pirolli, P., Hong, L., Schiano, D., and Chi, E. H. With a little help from my friends: examining the impact of social annotations in sensemaking tasks. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09 (2009), 1795–1798.
- [24] Nichols, J., and Kang, J.-H. Asking questions of targeted strangers on social networks. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, CSCW '12, ACM (New York, NY, USA, 2012), 999–1002.
- [25] Paul, S. A., Hong, L., and Chi, E. H. Is twitter a good place for asking questions? a characterization study. In *ICWSM* (2011).
- [26] Paul, S. A., Hong, L., and Chi, E. H. Who is authoritative? understanding reputation mechanisms in quora. In *Collective Intelligence conference* (2012).
- [27] Paul, S. A., and Reddy, M. C. Understanding together: sensemaking in collaborative information seeking. In *In Proceedings of CSCW2010* (2010), 321–330.
- [28] Pickens, J., Golovchinsky, G., Shah, C., Qvarfordt, P., and Back, M. Algorithmic mediation for collaborative exploratory search. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '08 (2008), 315–322.
- [29] Pirolli, P., Card, S. K., and Van Der Wege, M. M. Visual information foraging in a focus + context visualization. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '01, ACM (New York, NY, USA, 2001), 506–513.
- [30] Quora. <http://www.quora.com/>.
- [31] Vogt, K., Bradel, L., Andrews, C., North, C., Ender, A., and Hutchings, D. Co-located collaborative sensemaking on a large high-resolution display with multiple input devices. In *Proceedings of the 13th IFIP TC 13 international conference on Human-computer interaction - Volume Part II*, INTERACT'11 (2011), 589–604.
- [32] Yanbe, Y., Jatowt, A., Nakamura, S., and Tanaka, K. Can social bookmarking enhance search in the web? In *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*, JCDL '07 (2007), 107–116.
- [33] Zanardi, V., and Capra, L. Social ranking: uncovering relevant content using tag-based recommender systems. In *Proceedings of the 2008 ACM conference on Recommender systems*, RecSys '08 (2008), 51–58.