Searcher Actions and Strategies in Asynchronous Collaborative Search

Robert Capra, Annie T. Chen, Evonne McArthur, Natalie Davis
School of Information and Library Science
University of North Carolina at Chapel Hill
{rcapra, atchen, evonnemc, ndavis93}@email.unc.edu

ABSTRACT
In this paper, we present results of a laboratory study in which participants completed an asynchronous collaborative search task while thinking aloud. Based on analysis of the think-aloud data and screen recordings, we present a set of collaborative search actions and rationales that our participants employed. For each, we describe the purpose and motivations, and give illustrative examples. We also present three high-level strategies (independent, parallel, and divergent) that emerged from analysis of participants' verbalizations and discuss how participants used these strategies as part of their overall search process. Our results show that collaborators' prior work influenced search strategies and behaviors, and that participants leveraged collaborators' work at various stages of the interaction including query formulation and results examination. We discuss how the observed behaviors complement existing models of interactive information seeking and suggest ways to extend current models.

Keywords
Collaborative search, information seeking behaviors

INTRODUCTION
Collaborative information seeking (CIS), sometimes referred to as collaborative information retrieval, has been defined as "the activities that a group or team of people undertakes to identify and resolve a shared information need" (Poltrock et al., 2003, p. 239). Recently, there has been considerable interest in explicitly supporting collaboration in interactive search interfaces. For example, systems such as SearchTogether (Morris & Horvitz, 2007), Coagmento (Shah, 2012), and Querium (Golovchinsky et al., 2012) all include features to support awareness and coordination in a collaborative search. Although there has been much interest in collaborative search, existing models of information seeking behavior largely focus on the processes of individual searchers. It is unclear how dimensions of individual information seeking map onto collaborative search processes, and what new aspects need to be taken into account. Elaborating on their definition, Poltrock et al. (2003) describe both existing and new dimensions that play a role in collaborative search:

"Information retrieval involves identifying an information need, formulating a query, retrieving information, evaluating it, and applying it to address the need. Collaborative information retrieval involves these same activities but also includes communicating about the information need, sharing the retrieved information within the team, and coordinating the constituent information retrieval activities across multiple participants." (p.239)

The specific collaborative actions and strategies that users can employ are influenced by the features of a particular collaborative search system. Prior research (Reddy & Jansen 2008; Morris & Teevan, 2009; Erickson & Kellogg, 2000) has identified several important features for collaborative systems, and many CIS systems include features such as shared query histories, ratings, tags, chat/IM, filters, explicit sharing of results, and visualizations.

Our interest is in understanding the "hows and whys" of the collaborative search process. What strategies do people use? What actions do they take to realize these strategies, and how do they use the features of the system to meet their goals? How is the search process influenced by teammates' prior actions? In this paper, we investigate these questions in the context of an asynchronous collaborative search session with a specific set of interface features (query histories, shared ratings, and filter controls). Our goal is not to propose a new, generalizable model of CIS, but rather to understand the strategies and actions of users engaged in a collaborative search, and to examine these in relation to existing models of individual search processes.

We present results of a study using a contemporaneous think-aloud protocol to understand users' actions and motivations. We investigate how participants used specific collaborative features, report observations about the higher-level strategies employed, and discuss transitions among strategies. The system used in our study, *ResultsSpace* (Capra et al., 2012), was designed for asynchronous

ASIST 2013, November 1-6, 2013, Montreal, Quebec, Canada.

collaboration among small groups and includes several common CIS features: a shared query history, document-level ratings, and collaborative filter controls.

Based on analysis of the think-aloud data and screen recordings, we present a set of collaborative search actions and rationales that our participants employed. For each, we describe the purpose and motivations, and give illustrative examples. We also present three higher-level strategies that emerged from analysis of participants' verbalizations and discuss how participants realized these strategies and moved among them as part of their overall search process.

RELATED WORK

Interactive information seeking models emphasize the effect of the interaction "on searcher strategies, processes, and outcomes" (Knight & Spink, 2008, p. 216). For example, Marchionini (1995) proposed a model that incorporates eight states (e.g., formulate query, execute query, examine results, extract info, reflect/stop), with high and low probability transitions among these states. Marchionini's model highlights the interactive, iterative nature in which a searcher interacts with the information provided by the search system. These iterative components are still present in a collaborative search system, but additional information and controls may be provided based on the actions that other collaborators have taken (e.g., query histories, shared document ratings, filters). These additional components allow interactions that are not present in an individual search.

Existing information behavior models focus primarily on the individual information seeker, and there is a need to better understand the process of collaboratively seeking information (Reddy & Jansen, 2008; Shah & González-Ibáñez, 2010). Models of the information seeking process often do incorporate aspects of the information seeker's social and organizational environment, and of their work task and information needs. Collaborators in these models may have influence over a search process as persons consulted for advice, but have been considered less in the context of *shared interaction* in the search environment. However, researchers have begun to explore how models of information seeking apply to collaborative search.

Kuhlthau's (1991) Information Search Process (ISP) model has been studied in the context of collaborative search by several researchers. Shah and González-Ibáñez (2010) studied pairs of participants engaged in a synchronous collaborative search. They used action logs and chat logs to map the participants' processes onto Kuhlthau's stages. They concluded that exploration, formulation, and collection were not distinct stages in the collaborative searches observed, and noted that participants moved fluidly between individual search activities and interactions with collaborators. Hyldegård (2009) conducted a study of group members' information behaviors and found similarities to the ISP model stages. Like Shah and González-Ibáñez, Hyldegård noted that group-based problem solving seemed to be a dynamic process that shifted between a group perspective and an individual perspective. Reddy and Jansen (2008) studied the collaborative information behaviors of patient care teams in hospitals. They developed a model of collaborative information behavior along two dimensions: a behavior axis from searching to seeking, and a context axis from individual to collaborative. They found that shifts from individual to collaborative information behaviors were often prompted by certain types of events, and identified four specific triggers: complexity of the information need, fragmented resources, lack of domain expertise, and lack of immediately accessible information.

Other information seeking models have also been considered in the context of collaborative search. Shah (2012) examines how several existing models might be applied in a collaborative search context. For example, he reviews Marchionini's (1995) eight stages and describes how collaboration could be present in each phase. Evans and Chi (2010) developed a model of "social search" that considers coordination with others before, during, and after a search episode. Yue and He (2010) examined stages of the collaborative search process in an e-discovery task and found that participants rated individual, asynchronous parts of a larger collaborative task as the most difficult. They concluded that in these situations, users could benefit from "implicit collaboration support like recommendations and support for relevancy judgment." (p.4)

From the perspective of understanding searchers actions, Xie and Joo (2010) studied the use of tactics and strategies at different phases of an individual search. Common strategies included iterative result-evaluation, iterative exploration and query-initiation. Yue at al. (2012) used Hidden Markov Models to model search tactics in individual and collaborative settings, finding that collaborators' actions can influence query reformulations.

In this paper, we examine participants' think-aloud descriptions of how they conducted a collaborative search and why they took specific actions. To study this, we used the ResultsSpace collaborative search system.

RESULTS SPACE

ResultsSpace is a collaborative search system we developed to explore how awareness of collaborators' actions can be used in a collaborative search. ResultsSpace is designed to support collaboration among two to six people working together to find information for a shared task. The main ResultsSpace interface is shown in Figure 1.

ResultsSpace augments the traditional query box and result list with additional displays and controls for collaborative filtering and awareness. Results returned by ResultsSpace can be rated as either "relevant", "maybe relevant", or "not relevant" using the controls to the left of each result (i.e., the green up arrow, yellow box, and red down arrow, respectively). These controls also serve as a collaborative awareness display by including the number of teammates

Figure 1. Results Space Collaborative Search Interface

who have previously given the document a particular rating. The black underline shows the rating that the currently logged in user has given. For example, in Figure 1, we can see that the second result has been rated as "relevant" by one teammate and that the current user (Betty) has also rated it as "maybe relevant". Mousing over the control displays a pop-up to show who made the ratings (e.g., Martin and Betty have rated the fifth result as "not relevant"). Result titles and snippets are highlighted or diminished based on the net ratings of the group. Results with a net positive rating are highlighted with a light green background (e.g., results 2, 3, and 7 in Figure 1). Results with a net negative rating are "faded out" as shown for results 5 and 6 in Figure 1. Results with no net positive or negative rating are displayed normally. Using these displays, users can quickly see which results have been evaluated by their collaborators and what ratings were given. The displays can be helpful in different ways depending on the goals of the user. For example, if a user wanted to understand what their collaborators had found useful, they could focus on the results highlighted in green. However, if they wanted to focus on finding "new" articles, they could use the display to avoid previously rated items.

ResultSpace shows users a list of the 10 most recent queries issued by collaborators and the user's own 10 most recent queries. These can be clicked at any time to re-issue the query. ResultsSpace also includes controls to filter results based on whether or not a particular collaborator has rated the result, and based on the relevance ratings (left side of Figure 1). Additional details about ResultsSpace are outlined in a previous paper (Capra et al., 2012).

METHOD

We conducted a laboratory study using a think-aloud protocol in which participants worked on a task in a simulated collaborative setting. We began by recruiting three initial participants (C1-C3) from our school. Data from these three participants was used as the "collaborator" data for the subsequent participants and was built up incrementally. For example, C1 started with no collaborative data, C2 could see C1's actions (e.g., ratings and queries), and C3 could see actions from both C1 and C2. These three participants were used only to build up the collaborator data; their actions are not included in the analyses presented here. After the initial three participants, we recruited a main set of 11 participants (P1-P11) from a university-wide opt-in mailing list. All participants were graduate students, and came from a variety of disciplines.

For the simulated asynchronous collaboration, we used a method similar those used by Paul and Morris (2009) and Nelson et al. (2009). We asked each of the 11 participants to imagine that they were working on a collaborative project with three collaborators. The "collaborators" consisted of the prior data from C1, C2, and C3 and were referred to in the scenario by the pseudonyms Betty, Luis, and Martin. Thus, each of the 11 main participants came to the lab individually and worked on the task with C1, C2, and C3 as their "collaborators."

Task Scenario and Corpus

Studying search behavior in a laboratory setting always presents challenges. For our study, we wanted to develop a collaborative scenario that would: 1) be familiar to our participants who were all graduate students, 2) be flexible enough to allow participants to make choices about how they wanted to approach the task and their role in the

collaboration, and 3) involve a task that was accessible to participants from a variety of academic departments. To support this, we developed the following scenario:

For this task, imagine that you are taking an Environmental Studies class here at UNC. As part of the class, your instructor has given you a research assignment to do in small groups. The goal of the assignment is to find articles that will help you write a research paper on an assigned topic (shown below). Your instructor has given you access to a database of news articles from 1996 to 2000 to be used for the assignment.

You are in a group with three other people (Luis, Martin, and Betty). Your team agreed that everyone would do some searches on the database to find articles that may be useful in writing your research paper. Your other team members may have already done some searches and the group has agreed to meet tomorrow to talk about what everyone found. Your task today is to find and rate articles that will help your group with the assignment.

We adapted this search task from the TREC 2004 Robust Track collection (Voorhees, 2006). The tasks in this collection have multiple dimensions and are largely exploratory in nature, making them well-suited to collaborative search. These tasks are designed to be used with the AQUAINT collection of news articles. We loaded the ResultsSpace system with a total of 856,941 AQUAINT articles from three major sources (AP, NYTimes, and Xinhua) after filtering to remove duplicates and ensuring that articles had titles. We adapted task (#435), which had a "medium" number of relevant documents in the corpus, meeting our goal of not being too easy or too difficult:

Use the news article database to find articles that will help you write a research paper on the topic below.

What measures have been taken worldwide and what countries have been effective in curbing population growth? While researching this topic, keep in mind that your paper will be stronger if you support your thesis with actual cases in which population measures have been taken and the results are known. For this assignment, reduction measures to control growth are defined as those that are being actively pursued. Passive events such as disease or famine that involuntarily reduce population should not be cited.

Protocol

Upon arrival, participants were given an overview of the study and asked to read and sign an informed consent form. Next, they completed a demographic questionnaire and watched a video that introduced the ResultsSpace system, including an overview of its collaborative features. Participants were then given a written copy of the scenario and task description and asked to read it out loud. Reading the scenario aloud helped to ensure that participants had read all the information and also prepared them for thinking aloud while doing the task. Next, we answered any questions they had, and gave them 30 minutes to work on the task. Participants were told they could stop earlier if they reached a point where they would naturally stop. We instructed participants to "think aloud" as they worked to explain what they were doing and thinking. We emphasized that this was an important aspect of the study, and gently reminded them to keep talking if they lapsed. At the 25 minute mark, we let them know they had five minutes left.

ANALYSIS

For the first phase of our analysis, we wanted to understand what strategies and actions were used by participants to engage in the collaborative search process. We outlined areas and actions of interest and two of the authors completed an initial phase of open-coding based on watching the first 15 minutes of two participants' sessions. Through this step, initial codes were created for both the main actions taken and for the participants' descriptions of their actions. As part of this step, we also noticed that participants described several higher-level strategies, or "states" in which they worked: 1) working independently, 2) working on areas similar to their teammates, and 3) explicitly working in areas motived by their team, but different from areas already explored by teammates.

After developing the initial coding scheme, the two coders did additional independent coding to revise and clarify the code definitions. A two-level coding scheme for the actions was finalized. At the top level, six actions were coded: *issue query*, *used ratings*, *used filters*, *make a rating*, *read document*, and make a *strategic comment*. Each action was associated with one of three strategy states: independent, parallel, or divergent (explained in detail below). For three of the actions (issued query, used ratings, strategic comment), the verbal think-aloud data revealed additional, clear distinctions based on the participants' reasons for performing the action. We incorporated these as a second-level in the coding scheme that we refer to as the *descriptions* for the actions.

Utilizing the final coding scheme, the two researchers both independently coded the remaining participant recordings. Cohen's kappa statistic was 0.87 for the first-level action codes and was 0.69 for a combination of the second-level description codes and strategy codes, indicating substantial agreement. Discrepancies were resolved by consensus.

COLLABORATIVE SEARCH STRATEGIES

Three collaborative strategies emerged from our analysis of the data: 1) *independent* work, 2) collaborative work *parallel* to collaborators' prior work, and 3) collaborative work exploring in directions *divergent* from those previously explored by collaborators.

Independent State

When in the independent strategy state, participants were acting on their own and were not immediately focused on their collaborators' work. While in the independent strategy state, the collaborative interface features were not directly influencing the participant even though they may have used these features at other points in the search process.

Parallel State

The parallel state describes a mode in which the participant was using awareness of their collaborators' previous work to do additional work in the same space. A participant in this state might have explored the results of a collaborators'

previous query in more depth, read/rated documents from collaborators' prior searches, or issued related queries to find results on the same sub-topic.

Divergent State

In the divergent state, collaborative awareness was used specifically to forge new ground rather than to extend existing work. When in this state, participants were trying to use knowledge of what their collaborators had already done to go in new directions. They were intentionally trying to avoid overlap and duplication. The divergent strategy represents collaborative awareness with the intent to work in an alternative direction with respect to the team's existing results (e.g. to increase breadth).

COLLABORATIVE SEARCH ACTIONS

Participants took specific actions as supported by the ResultsSpace interface to achieve their search strategies and goals. The main actions were: (i) issue query, (ii) make rating, (iii) use rating, (iv) apply filter, and (v) read document. As part of the think-aloud, participants sometimes made specific comments about their strategies and we included these in our top-level coding as (vi) strategic comments. Below, we describe each of these actions and give examples to illustrate their use.

Issue Query

Participants issued many queries and query formulation is a natural place for people to think-aloud. Based on this we were able to refine the "issue query" action into three second-level codes based on the verbal descriptions. We also were able to associate a high-level strategy code (independent, parallel, divergent) for all the queries in our dataset. These are presented below.

Creation of new query (independent)

Some queries were entered as "new" queries that were not influenced by the work of collaborators or by prior queries issued by the participant themselves. Participants may have generated these new queries from personal knowledge, or from information contained in the task. We coded these types of queries as employing the independent strategy. For example, P2 described generating a query based on the written task description:

P2: "Since the topic is the measures that have been taken worldwide and… population growth, I might start with 'Population growth worldwide' as a search term"

Revision based on information (independent)

In individual search, users often revise their own queries and our participants also did so. They revised queries based on information gathered and based on their prior knowledge. For this code, we did not include cases where collaborators' work was used to revise a query, and thus we count this an independent strategy. Below, P1 revises a query based on prior knowledge, and P2 does so based on knowledge gained while searching:

P1: "Since 'legal controls' [query] didn't work, I'm focusing on the one country that I know about the legal controls in to see if there's anything helpful there"

P3: "A couple of the articles I've been finding are about family planning so I'm going to change my search"

Revision based on lack of information (independent)

We noted situations in which participants reformulated queries because they were unsatisfied with the original result, either in terms of novelty or topical relevance. Participants in these cases did not use their collaborators' prior work, so this is also an independent strategy:

P8: "There's a lot of stuff about gun control; maybe control isn't the right word."

Use collaborator query (parallel)

Participants could use the ResultsSpace query history to issue similar or identical queries to those issued by their teammates. For this code, we focus on situations where the goal was to use a collaborators' prior query to explore a *similar* area. Example goals could be to refine the precision of a collaborator's results, or to make sense of the area where a collaborator was searching. Specific actions could include clicking on a collaborators' prior query, or purposively creating a similar query. These types of actions are examples of the parallel strategy.

P4: "Let's do [a query for] family planning because all of them [previous collaborators] have been searching with family planning"

Use collaborator query (divergent)

Participants could also use their collaborators' queries to create intentionally *dissimilar* queries in order to increase the coverage (recall) of the team or avoid work overlap. Since this action involves using collaborators work as a jumping off point and purposely branching into a different area, is considered a divergent strategy.

P3: "I see that other people in my group have used sex education and family planning… those sound like interesting ideas to use in a search but since they've already done them I don't want to just replicate what they've done. So, maybe I can use a different country."

Used Ratings

Participants made use of their collaborators' ratings while engaged in their search process. Frequently, ratings were used to triage results returned from a search and to make decisions about whether to view a specific document.

Operationalizing the coding for this action required making distinctions about specific verbalizations and observed behaviors. We applied this code if a participant explicitly verbalized that they were using their collaborators' ratings to make decisions about what documents to view or avoid. We also included cases where it was very clear that the participant was engaged in a sequence of actions that involved deliberately viewing only documents that were rated (or not rated), even if the participant did not verbally

describe the behavior. Based on the verbal data, we identified three second-level codes described below. All three focus on using the ratings to make decisions about viewing a document (or not).

Viewed documents others have rated (parallel)
Participants viewed documents that their teammates had rated as relevant as part of a process of sense-making, or to continue down the same path based on what their collaborators viewed as relevant.

P2: "People thought this was useful--it's not a very descriptive title but let's see what it is"

P8: "I'd… skip to the one that already has a green on it"

Viewed documents others have NOT rated (divergent)
Conversely, participants used the ratings to choose documents that their collaborators had not already rated. This can be viewed as a choice to diverge – the participant specifically decided to focus on unrated documents.

P1: "Seems like a lot of people have already checked out these articles; it doesn't seem helpful to look at articles other people have already agreed upon"

Viewed documents others disagreed on (parallel)
Participants also used ratings to identify documents that had already been rated, but that could benefit from additional input. For example, a participant might choose to view documents that their teammates had rated, but disagreed about (i.e., that had conflicting ratings), for the purpose of trying to resolve the conflict. Similarly, they might view a document because a teammate had given it a "maybe relevant" rating and they wished to make a deciding vote. This type of "scale tipping" was described by some participants as a conscious contribution to the group by adding their opinion. Since the contribution furthers an existing line of exploration by the teammates, we categorize this as an action as parallel strategy. The quotation below illustrates how this type of action was viewed as productive and re-assuring to P3, especially since they did not feel that they were finding new information on their own.

P3: "There's an article that one of my group people says is irrelevant and one says is relevant so maybe I can be the tie-breaker because I'm not being so successful finding things on my own."

Filter
ResultsSpace provides faceted controls on the left side of the interface that can be used to filter the set of results. Users can filter the results to show only specific ratings ("relevant", "maybe relevant", "not relevant") or to show only results that have been rated by specific teammates.

Participants applied the "User" filters to narrow results to show only items that their collaborators had rated. Use of this type of filter was used when the participant wanted to get an overview of the types of documents that a particular teammate had found, and how they had rated them.

P8: "Then I'd probably try to do another method and maybe look each member up and see what they thought"

P11: "Let's look at what Martin has done."

Participants used the "Relevance" filters to find the documents that teammates had rated as relevant:

P11: "Let's see what all of them found as relevant."

Both these types of filtering focus on documents that teammates had already rated, helping participants to understand the progress of the team so far. However, reviewing documents that teammates had already rated could be part of a broader, multi-stage strategy to first understand what collaborators' had found and then work in a different direction. Such a multi-stage strategy would require a second distinct action in which the filter was cleared and a new search issued. We view these separately and consider the application of a filter for its immediate help in sense-making which falls into the parallel strategy.

Strategic Comment
At times, participants made think-aloud comments about their strategy which were not tied to a specific actions in the interface. We coded these as "strategic comments", noting whether they were independent, parallel, or divergent.

Independent – P6: "I'm sort of looking at common words [on the SERP]. You know, common authors, you know, anything that is just repeated multiple times to get a sense of [..] what the major keywords are"

Parallel – P5: "It looks like my group members are really just focusing on family planning and sex education, so that's probably what I should focus on, too."

Divergent – P1: "It's about Kenya… the other people have just been looking at Asia and not really branching beyond there so it might be worth looking in this direction"

Make a Rating
Rating documents was a common action. Participants did not always describe why they were making a particular rating, but when they did it was often directly related to text they found in the document that indicated relevance (or not). For example, P6 describes a situation with debate about making a rating:

P6: This [article] is also about a concern. Looking back at this [the sheet with the task], they want to specifically know which measures. So, this might be relevant, I'm not 100% sure. [Marks yellow]"

We also noticed cases where a participant was making ratings and engaged in a type of "parallel" strategy that indicated a degree of social influence about the ratings.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	Total
Issue Query	**10**	**4**	**11**	**13**	**25**	**22**	**9**	**3**	**23**	**7**	**10**	**137**
Creation of new query (i)		1		3	2	2	2	1	3	1	2	**17**
Revision based on information (i)	4		3		1	2	2		2	1	1	**16**
Revision based on lack of information (i)	4	1	5	6	21	12	4	1	11	1	7	**73**
Use collaborator query (p)		2	1	4		5	1	1	7	4		**25**
Use collaborator query (d)	2		2		1	1						**6**
Used Ratings	**6**	**3**	**2**	**0**	**2**	**0**	**4**	**6**	**1**	**0**	**1**	**25**
Viewed documents others have rated (p)		2					4	4	1		1	**12**
Viewed documents others have not rated (d)	3	1	1		2			2				**9**
Viewed documents others disagreed on (p)	3		1									**4**
Strategic Comment	**7**		**2**	**1**	**4**	**2**	**1**	**3**	**1**	**1**	**5**	**27**
Parallel (p)	1		2		1			2	1			**7**
Divergent (d)	4				3		1				3	**11**
Independent (i)	2			1		2		1		1	2	**9**
Filter	**1**	**2**	**0**	**0**	**0**	**0**	**6**	**5**	**0**	**2**	**2**	**18**
Made Rating	**25**	**14**	**10**	**15**	**30**	**29**	**13**	**21**	**30**	**15**	**32**	**226**
Reading	**29**	**15**	**16**	**15**	**28**	**28**	**18**	**21**	**45**	**15**	**24**	**262**
Total	**78**	**38**	**41**	**44**	**89**	**81**	**51**	**59**	**100**	**40**	**74**	**884**

(i)=independent, (p)=parallel, (d)=divergent

Table 1. Summary of Coded Participant Actions. Bolded rows show the top-level codes. Second-level codes are shown for issue query, used ratings, and strategic comment.

P8: "And all of these already have other people that said yellow, so I feel better when I also put yellow."

Participants also made ratings to help resolve conflicts, and to help their group members in future searches:

P1: "I am finding that I keep forgetting to click on the negative [rating] for things that aren't helpful…which is a useful part of a program like this to help your group members know that something isn't worth looking at."

Reading

Reading documents was also a common action. Similar to the ratings, participants did not always verbalize why they choose to read an article, but when they did it was usually because they had seen some text in the result title or snippet that led them to think the document was relevant. For example, P6 describes using a title to make a decision:

P6: "the next article... says 'Population Control, Hard Nut to Crack' which implies that it's talking about issues – steps that have been taken to... limit population growth."

Participants also read documents as part of other strategies. For example, P3 describes reading a document based on a parallel strategy, and P5 as part of a divergent goal:

P3: "There was...an article that one of my partners tagged as relevant that was about [this program] being successful" (P3 then reads another article about the same program)

P5: "Here's an article about Kenya; I don't think my group members have found anything about Kenya."

Summary of Actions

Based on participants' interactions and think-aloud comments, we identified and coded the actions and strategies described in the previous sections. Table 1 shows a summary of the actions for each participant. The rows in bold summarize the actions as coded at the top-level, and the non-bolded rows below show the second-level actions and strategy codes. As noted in the previous section, we did not separate out second-level action codes for the filter, made rating, and reading actions because participants did not provide enough verbal data for those actions to support that level of analysis. Since the actions summarized in Table 1 are quite different in nature, we did not normalize them as percentages for each participant.

From Table 1, we can see that participants were varied in their actions and strategies. All participants issued queries independently, and all but one participant (P11) made some use of their collaborators' prior queries. All but three participants (P4, P6, and P10) made use of their collaborators' prior ratings to make decisions about what documents to view. Participants also made "strategic comments" that described utilizing their collaborators' prior work. Five participants (P1, P3, P5, P8, P9) made comments that indicated they did some work in parallel to their collaborators, and four participants (P1, P5, P7, P11) described strategies that were intentionally divergent. All participants read and rated documents and in most cases these numbers were similar, indicating that most documents that were read were also rated. A more detailed analysis of the reading and rating actions is given in (Capra et al., 2012). In general, the filter controls were not used heavily, but two participants (P7, P8) made considerable use of them to help understand what their collaborators' had rated.

Taken in the context of the task, the actions we observed and that are summarized in Table 1 suggest that while

participants took many actions in the independent state, the collaborative states (parallel and divergent) played important roles overall. We observed influence of the collaborative awareness information at key points in the search process – when making choices about what queries to issue next, what documents to view, and how to best contribute to the work of the group.

In the study, participants were given a full 30 minutes to work. The wording of the task allowed them to make their own choices about how to allocate that time and what actions to take. Our results illustrate that participants leveraged the work that collaborators had done previously and that it influenced their search strategies and behaviors.

STATE DURATIONS AND TRANSITIONS

In addition to looking at frequency of actions, we were also interested in the amount of time participants spent in each of the three strategy states (independent, parallel, and divergent), and how and why they transitioned between them. Since the strategy states were coded at the level of the actions, we considered a participant as being in a particular state until they provided evidence of being in a different state through a subsequent action or a comment. For example, if a participant described making a query to diverge from their collaborators', and then rated documents based on that query, we considered those "made ratings" actions to also be in the divergent state unless they indicated otherwise.

Figure 2 shows the total amount of time in seconds that each participant spent in each strategy state. We can see that overall, participants spent the greatest proportion of time in the independent state. However, many participants spent considerable time in the parallel state, indicating the important role that understanding collaborators' prior work played in the sensemaking and search process. Participants generally spent less time in the divergent state, but four (P1, P3, P5, P11) worked in this state for over 20% of their time.

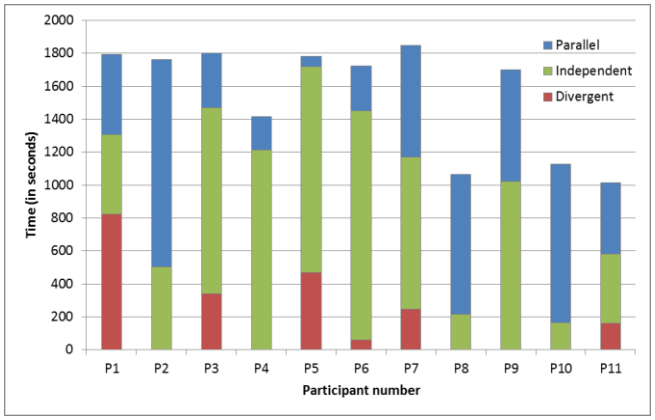

Figure 2. Participant Time in Strategy States

Considering Figure 2, we can begin to identify several collaborative search styles:

Parallel-oriented: This group tended to work in parallel with their teammates, and included those who spent a significant amount of time understanding what their collaborators had done (e.g., P2, P8, P9, P10).

Independent-oriented: These participants spent most of their time exploring and searching for relevant documents on their own (e.g., P4, P6).

Team-aware: This group spent the majority of their time working in a mixture of the parallel and divergent states, with little time in the independent state (e.g., P1, P11).

Mixtures: Some participants used mixtures of independent, parallel and/or divergent strategies (e.g., P3, P5, P7).

The task we gave participants allowed them considerable choice in deciding how to spend their time. Individual characteristics and preferences likely played a role in the strategies and actions taken. We did not explicitly ask participants to reflect on their own behaviors, but based on our observations, we comment on several possible influences. First, we note that people may have natural tendencies to allocate their time differently to these respective uses, i.e. some individuals may tend towards working in parallel with collaborators, others to diverge, and still others may prefer working independently. When faced with all of the awareness mechanisms presented by the ResultsSpace system, some curious participants may have wondered what others had done, and spent time investigating, while others who were more self-efficacious in search, may have been more confident venturing out mostly on their own. Lastly, individuals may have seen their own role in the search task or as part of the group differently; in particular, individuals may have felt that their role was to bring new information to the group, and thus spent more time in the divergent strategy state.

COLLABORATIVE STRATEGY STATE TRANSITIONS

We wanted to understand how participants moved among the strategy states as part of their overall search process. Table 2 summarizes the transitions across the entire sample. We only considered transitions from one state (left) to a different state (top), so the diagonal is empty.

from \ to	**independent**	**parallel**	**divergent**
independent	-	28	9
parallel	22	-	3
divergent	13	1	-

Table 2. Transitions to a Different Strategy State

Table 2 shows some interesting patterns. Transitions between independent and parallel were the most common, and occurred frequently in both directions (independent to parallel and vice-versa). Similarly, transitions from independent to divergent (and vice-versa) were also common, but less so, occurring about half as often as the independent-parallel transitions. Though not represented in this matrix, participants most often started the task in independent state. Direct transitions between parallel and divergent strategies were rare – occurring only four times across all our data. The prevalence of transitions with the independent state can be understood in part by considering

the context provided in Figure 2. All participants spent time in the independent and parallel states, but overall, less time was spent in the divergent state.

Transitioning between states occurred for several different reasons and at different points in the search process. Many participants, after running out of independent ideas, reached out to use their collaborators' prior work. For example, approximately 10 minutes into the task, P6 indicated unhappiness with their search results and said:

P6: "So I think I am going to go back... I'm looking at what Martin searched."

Transitions out of the independent state could occur for reasons other than running out of relevant results or ideas. For example, P1 was working independently and noticed that many articles they were finding had already been rated:

P1: "It seems like other people have already checked out these articles. It doesn't seem that helpful to look at articles that other people have already agreed upon."

After this, P1 modified their strategy to avoid overlap with collaborators, putting them in the divergent strategy state.

Going the other direction, transitions from the parallel and divergent states to the independent one frequently occurred because of inspiration, keywords, or connections noticed in documents or in a results list. One of these events might then take the participant in a new direction, leading to work independent of collaborators' direct influence. For instance, P6 transitioned from parallel to independent after finding key words in a document:

P6: "I might... try using those sentences... 'reduction of birth rate in developing countries' and 'adoption of contraception in developing countries' as search terms."

DISCUSSION

In this study, it was necessary for participants to negotiate a tension between their own contributions to the collaborative search process and coordination with their teammates. We observed actions used, along with three strategy states (independent, parallel, and divergent) that participants moved among as part of their search process.

Similar tension between individual and team aspects of groupwork has been described in other CSCW settings (e.g., Dourish &Bellotti, 1992) and also found in studies of collaborative information seeking (Reddy & Jansen 2008; Shah & González-Ibáñez, 2010). In Reddy and Jansen's (2008) work, they identified specific triggers that contributed to a person's decision to shift from individual to collaborative information behaviors including topic complexity, fragmented resources, lack of expertise, and lack of accessible information. In our study, we also observed specific rationales and triggers that influenced when participants moved from an individual to a collaborative strategy. Similar to the results of Reddy and Jansen, our participants described reasons such as a lack of ideas for search terms and having trouble finding good results as reasons for turning to the collaborative features.

Many of the actions observed in our study could map onto existing individual interactive information seeking models such as Marchionini's eight stages (1995). We observed query formulation and execution, results examination, information extraction, and reflection. Thus, we see an extension to Marchionini's model that would account for collaborative user strategies at each state. For example, query formulation can be accomplished either as an independent, parallel, or divergent action as supported by CIS systems and as seen in our data. Similarly, if a CIS interface supports shared ratings, they can influence the results examination stage. Marchionini's high and low probability transitions could be complemented with transitions from our stage+strategy combinations. Our analysis of strategy transitions gives insights about which transitions would have high and low probability. New states would be needed to capture the additional interactions available as part of a collaborative system such as making ratings, and sensemaking of teammates' work.

From our data, we also see connections to Pirolli and Card's (1999) information foraging model in which searchers make decisions about what results to investigate based on "information scent". In a collaborative setting, our participants used teammates' ratings as part of the scent for a set of results, and they obtained ideas for getting to new "patches" by reviewing their teammates' work.

The collaborative awareness features of ResultsSpace benefited participants' sensemaking behaviors. Participants were able to use information about what their collaborators had rated not only to decide which documents to read, but also to understand their teammates' strategies and the overall topic space. Thus, rather than a process in which searching is followed by understanding and analysis of documents, the system both facilitates and encourages switches between searching and sensemaking. This type of "integrated search and sensemaking cycle" was also observed in Morris, Lombardo and Wigdor's (2010) evaluation of WeSearch, a tabletop display that supported synchronous collaborative search. Thus, one promising direction for future development would be the incorporation of additional features for sensemaking.

CONCLUSION

In this paper, we presented results of a laboratory study that used a contemporaneous think-aloud protocol to gain insights into the actions and strategies that participants employed while working on an asynchronous collaborative search. Based on the data, we identified a set of collaborative search actions and described motivations and situations where each action was used. We also identified three higher-level strategies that participants used to guide their overall search process (independent, parallel, and divergent) and found that transitions often involved the independent state. Participants in our study used

combinations of strategies, suggesting that CIS systems need to provide support for them, as well as for fluid transitions between them. We showed how these actions and strategies complement existing models of interactive information seeking, and how they suggest ways to extend existing models to support collaborative search.

REFERENCES

Capra, R., Chen, A. T., Hawthorne, K., Arguello, J., Shaw, L., and Marchionini, G. (2012). Design and evaluation of a system to support collaborative search. In *Proceedings of the American Society for Information Science and Technology, Volume 49 (ASIS&T 12)* (p. 1-10). Silver Spring: ASIS&T.

Dourish, P. & Bellotti, V. (1992). Awareness and coordination in shared workspaces. In *Proceedings of the ACM conference on computer-supported cooperative work (CSCW '92)*. (pp. 107-114). New York: ACM.

Erickson, T., and Kellogg, W.A. (2000). Social translucence: an approach to designing systems that support social processes. *ACM Transactions on Computer-Human Interaction 7*(1), 59-83.

Evans, B. M., & Chi, E. H. (2010). An elaborated model of social search. *Information Processing & Management, 46*(6), 656–678.

Golovchinsky, G., Dunnigan, A., and Diriye, A. (2012). Designing a tool for exploratory information seeking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Extended Abstracts (CHI EA '12)*. (pp. 1799-1804). New York: ACM.

Hyldegård, J. (2009). Beyond the search process – Exploring group members' information behavior in context. *Information Processing & Management, 45*(1), 142–158.

Knight, S. A., & Spink, A. (2008). Toward a Web Search Information Behavior Model. In A. Spink & M. Zimmer (Eds.), *Web Search: Multidisciplinary Perspectives* (pp. 209–234). Berlin: Springer-Verlag.

Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science, 42*(5), 361–371.

Marchionini, G. (1995). Information Seeking in Electronic Environments. Cambridge: Cambridge University Press.

Morris, M. R., & Horvitz, E. (2007). SearchTogether : An Interface for Collaborative Web Search. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology (UIST '07)* (pp. 3–12). New York: ACM Press.

Morris, M. R., Lombardo, J., & Wigdor, D. (2010). WeSearch: Supporting collaborative search and sensemaking on a tabletop display. In *Proceedings of the ACM conference on computer supported cooperative work (CSCW '10)* (pp.401-410). New York: ACM.

Morris, M. R., & Teevan, J. (2009). Collaborative Web Search: Who, What, Where, When, and Why. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, Lecture #14. Morgan & Claypool.

Nelson, L., Held, C., Pirolli, P., Hong, L., Schiano, D., and Chi, E.H. (2009). With a little help from my friends: examining the impact of social annotations in sensemaking tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. (pp. 1795-1798). New York: ACM.

Paul, S. & Morris, M. (2009). CoSense: enhancing sensemaking for collaborative web search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09*). (pp. 1771-1780). New York: ACM.

Poltrock, S., Grudin, J., Dumais, S., Fidel, R., Bruce, H., & Pejtersen, A.M. (2003). Information seeking and sharing in design teams. In *Proceedings of the 2003 international ACM SIGGROUP conference on supporting group work (GROUP '03)* (pp. 239-247). New York: ACM.

Pirolli, P., & Card, S. (1999). Information Foraging. *Psychological Review 106*(4), 643-675.

Reddy, M. C., & Jansen, B. J. (2008). A model for understanding collaborative information behavior in context: A study of two healthcare teams. *Information Processing & Management, 44*(1), 256–273.

Shah, C. (2012). *Collaborative Information Seeking: The Art of Making the Whole Greater than the Sum of All.* Berlin: Springer-Verlag.

Shah, C., & González-Ibáñez, R. (2010). Exploring information seeking processes in collaborative search tasks. In *Proceedings of the 73rd Annual Meeting of the American Society for Information Science and Technology (ASIST '10)*, Vol. 47, Article 60, 10 pages. Silver Spring: ASIST.

Voorhees, E. M. (2006). Overview of the TREC 2005 Robust Retrieval Track. In *Proceedings of TREC-14*.

Wilson, T.D. (2000). Human information behavior. *Informing Science 3*(2), 49-55.

Xie, I., & Joo, S. (2010). Transitions in search tactics during the Web-based search process. *Journal of the American Society for Information Science and Technology, 61*(11), 2188-2205.

Yue, Z., and He, D. (2010). Exploring Collaborative Information Behavior in Context: A Case Study of E-discovery. In *CSCW 2010 Workshop on Collaborative Information Seeking*, Feb, 7, 2010, Savannah, Georgia.

Yue, Z., Han, S., Jiang, J., & He, D. (2012). Search tactics as a means of examining search processes in collaborative exploratory Web search. In *Proceedings of the 5th Ph.D. workshop on information and knowledge (PIKM '12)* (pp. 59-66). New York: ACM.