Embedding Search into a Conversational Platform to Support Collaborative Search

Sandeep Avula[1], Jaime Arguello[1], Robert Capra[1], Jordan Dodson[1], Yuhui Huang[1], Filip Radlinski[2]
[1]University of North Carolina at Chapel Hill, [2]Google
{asandeep,jarguello,rcapra}@unc.edu,{jojordan,yuhui97}@live.unc.edu,filiprad@google.com

ABSTRACT

Popular messaging platforms such as Slack have given rise to thousands of applications (or bots) that users can engage with individually or as a group. In this paper, we study the use of searchbots (i.e., bots that perform specific types of searches) during collaborative information-seeking tasks mediated through Slack. We report on a user study in which 27 pairs of participants were exposed to three searchbot conditions (a within-subjects design). In the first condition, participants completed the task by searching independently and coordinating through Slack (no searchbot). In the second condition, participants could only search inside of Slack using the searchbot. In the third condition, participants could both search inside of Slack using the searchbot and outside of Slack using their own independent search interfaces. We investigate four research questions focusing on the influence of the searchbot condition on outcomes associated with: (RQ1) participants' levels of workload, (RQ2) collaborative awareness, (RQ3) experiences interacting with the searchbot, and (RQ4) search behaviors. Our results suggest opportunities and challenges in designing searchbots to support collaborative search. On one hand, access to the searchbot resulted in more collaborative awareness, ease of coordination, and fewer duplicated searches. On the other hand, forcing participants to share the querying environment resulted in fewer overall queries, fewer query refinements by individuals, and greater levels of effort. We discuss the implications of our findings for designing effective searchbots to support collaborative search.

KEYWORDS

Searchbots; collaborative search; search behavior

ACM Reference Format:

Sandeep Avula, Jaime Arguello, Robert Capra, Jordan Dodson, Yuhui Huang, Filip Radlinski. 2019. Embedding Search into a Conversational Platform to Support Collaborative Search. In *2019 Conference on Human Information Interaction and Retrieval (CHIIR'19), March 10-14, 2019, Glasgow, Scotland, UK.* ACM, NY, NY, USA, 9 pages. https://doi.org/10.1145/3295750.3298928

CHIIR '19, March 10–14, 2019, Glasgow, Scotland Uk

ACM ISBN 978-1-4503-6025-8/19/03…$15.00
https://doi.org/10.1145/3295750.3298928

1 INTRODUCTION

Messaging platforms such as Slack and Microsoft Teams have become increasingly common and allow users to collaborate remotely on work-related and personal tasks. These new messaging platforms differ from older ones (e.g., IRC) in their aesthetics and ability to integrate with third-party applications (or "bots") that users can interact with directly from the communication channel. For example, Slack currently has third-party bots that perform services such as scheduling, cloud storage, version control, video conferencing, and other e-commerce services (e.g., travel booking, food delivery, and ride sharing). Additionally, so-called "searchbots" have been designed to perform specific types of searches (e.g., search for weather, news, and social media content). Currently, users mostly interact with searchbots by sending explicit requests (e.g., "@weatherbot new york city"), answering follow-up questions initiated by the searchbot, and interacting with the search results directly inside of the communication channel.

To date, much of the consideration of searchbots (and bots in general) has focused on their use by *individuals*. However, searchbots are also well-positioned to support *multiple* users working collaboratively on information-seeking tasks. Within information retrieval, the goal of *collaborative search* is to understand how people collaborate on information-seeking tasks and to design systems to support such collaborations. The most prominent approach has been to develop *dedicated* systems to support collaborative search [4, 8, 23, 25, 30, 33]. These systems have been designed with the search engine as the centerpiece component, but include peripheral tools for collaborators to communicate, share information, and gain awareness of each other's activities.

Dedicated systems for collaborative search have been found to provide benefits for users [4, 23, 25, 32]. However, in spite of these findings, stand-alone systems have not gained widespread adoption. In fact, research suggests that while people frequently engage in collaborative search, they do so using non-integrated tools—independent search systems and communication tools such as instant messaging, email, and phone [5, 21, 22]. Morris [22] and Hearst [11] highlighted these findings as a rationale to develop *lightweight search tools that can be integrated directly into existing communication platforms*. Our research is an answer to this call.

Within the context of collaborative search, integrating search tools into a communication platform such as Slack may provide benefits, but also introduce challenges. In terms of benefits, embedded search tools may help collaborators gain awareness of each other's activities and avoid duplicated work. In terms of challenges, sharing the *same* search environment may also prevent collaborators from effectively implementing "division of labor" strategies that require both joint and *independent* work. To gain insights into this novel design space, in this paper we evaluated two different searchbot designs, and compared them against a baseline scenario in which collaborators must search independently (no searchbot).

We report on a laboratory study in which 27 pairs of participants collaborated on three information-seeking tasks using the Slack messaging platform and were exposed to three searchbot conditions (a within-subjects design). In the *browser_only* condition, participants could *only* search using their own individual search interfaces. In this condition, we simulate the baseline scenario in which collaborators must search independently and use the messaging platform to share their findings and coordinate. In the *slack_only* condition, participants could *only* issue queries inside of Slack. We instrumented a searchbot that allowed participants to issue queries from Slack and returned search results directly in the Slack channel. Finally, in the *slack_browser* condition, participants were able to issue queries inside of Slack and also issue queries independently using their own individual search interfaces.

Our study investigated the following *four* research questions, in which the *searchbot condition* was the independent variable.

RQ1: What is the effect of the searchbot condition on participants' level of workload during the task? We used the NASA TLX [1] to measure different aspects of workload.

RQ2: What is the effect of the searchbot condition on participants' collaborative experience? We asked participants about their level of collaborative awareness, effort, and enjoyment.

RQ3: What is the effect of the searchbot condition on participants' experiences while interacting with the searchbot inside of Slack (only for the *slack_only* and *slack_browser* conditions)? We asked participants about gains obtained from the searchbot and their experiences coordinating their use of the searchbot.

RQ4: What is the effect of the searchbot condition on participants' search behaviors during the task? We investigated this question from three perspectives. First, we report on measures derived from the participant pair (e.g., total number of queries issued). Second, we report on the extent to which both participants issued queries with the same search intent versus engaging in "division of labor" strategies. Third, we report on the extent to which participants issued queries with the same intent as a previous query from the same participant (i.e., evidence of individual query refinement).

2 RELATED WORK

Our work builds upon two areas of prior research: (1) collaborative search and (2) shared control of devices.

Collaborative Search: The goal of collaborative search is to understand how people work together on information-seeking tasks and to develop systems to support such collaborations. Collaborative search has been studied in situations in which collaborators are co-located versus remote, and working synchronously versus asynchronously [31]. In this paper, we focus on remote and synchronous collaborations mediated through a messaging platform.

Many dedicated systems have been developed to support collaborative search, including SearchTogether [23], CoSense [25], Coagmento [30], CollabSearch [33], Querium [8], ResultsSpace [4], and SearchX [28]. All of these stand-alone systems include a search interface, as well as peripheral tools for collaborators to communicate, share information, and gain awareness of each other's activities.

Prior studies have found that dedicated systems can provide a wide-range of benefits for users. For example, tools to help users gain awareness of each other's activities have been found to help collaborators learn from each other [14, 23], avoid duplicated work [23, 28, 32], review each other's work [14], hand-off work to each other [25], and track their progress [32]. Capra *et al.* [4] found that awareness features also allowed participants to implement different collaborative strategies: (1) maximizing group precision by evaluating each other's findings and (2) maximizing group recall by *avoiding* each other's search paths.

While dedicated systems have been found to provide benefits, they have not gained widespread adoption. In a survey by Morris [22], about 50% of respondents reported performing collaborative searches at least once a week. In spite of this, respondents reported mostly using non-integrated search and communication tools. Interestingly, respondents also reported frustration with the lack of awareness associated with non-integrated tools, which can lead to duplicated work. Morris concluded by arguing that future research should develop "glue systems" that offer integration between existing search and communication tools.

A few systems evaluated in prior work resemble the types of "glue systems" proposed by Morris. The SearchBuddies system was developed to automatically embed search results in response to questions posted on Facebook [12]. A qualitative analysis of people's reactions to the SearchBuddies system suggests opportunities and challenges for "socially-embedded search systems". For example, some users found the embedded search results useful when they were complementary to answers provided by other "real" users. On the other hand, other users responded negatively when the embedded search results were relevant, but also obvious.

In a Wizard of Oz study, Avula *et al.* [3] explored searchbots that automatically intervened in Slack conversations centered around collaborative search. The authors evaluated two types of interventions: (1) intervening and asking questions about the needed information before providing search results and (2) intervening by "inferring" the needed information and directly providing search results. Both conditions improved participants' collaborative experience compared to a *no searchbot* condition.

Shared Control: Based on prior work on collaborative search, embedding search technology into communication platforms may provide benefits to collaborators (e.g., raising collaborative awareness). However, it may also introduce challenges associated with shared control (i.e., the need for collaborators to coordinate their use of the tool in the same workspace). Prior research has studied how *groups* coordinate their use of interactive systems, including desktop search interfaces [2], tabletop search interfaces [24], large screen applications [15, 29], mobile phone applications [18, 19], and voice-enabled personal assistants such as Siri [27]. While these studies have focused on co-located collaborations, their findings may also generalize to remote collaborations over a messaging platform. Studies have found that shared control introduces several challenges. Through a series of semi-structured interviews, Amershi and Morris [2] found that co-located web search can result in: (1) imbalanced participation, (2) pacing issues (i.e., one collaborator setting the pace for the others), and (3) difficulties implementing "division of labor" strategies (i.e., performing sub-tasks in parallel). Furthermore, studies have also found that group members often *self-select* when deciding who will interact with the system at a particular point in time [27, 29]. Self-selected (versus negotiated) interactions can lead to sub-optimal group outcomes and imbalanced engagement among collaborators. Finally, sharing a physical or virtual workspace can also be distracting when the awareness of others' activities is not beneficial [2, 24].

Prior research on group dynamics suggests that people working together tend to transition between periods of tightly coupled *group* work and more loosely coupled *individual* work (e.g., completing sub-tasks in parallel) [6, 9, 20]. More closely related to collaborative search, the CoSearch system was developed to support group searching as well as parallel searching by co-located individuals. Amershi and Morris [2] evaluated the CoSearch system against two other conditions: sharing a single computer and parallel searching on different computers. Participants reported greater awareness and sense of teamwork in the CoSearch and "shared" conditions as compared to the parallel condition. Additionally, participants reported greater ease in implementing "division of labor" strategies with the CoSearch system as compared to the shared condition.

3 METHODS

To investigate our four research questions, we conducted a laboratory study with 27 pairs of participants (44 female and 10 male). Participants were undergraduate students from our institution and were recruited in pairs. Each participant pair collaborated on three tasks that required searching for information. In this study, we did not investigate task effects. Therefore, to keep our search tasks as consistent as possible, we designed three *comparative* tasks to use in the study (Section 3.2). Participants were seated in different rooms and were each given access to three resources to use during the tasks. Resource 1: During all tasks, participants used the Slack messaging platform to communicate. Resource 2: During all tasks, participants used a shared Google Spreadsheet to create a joint written response for the task. Resource 3: For each of the three different tasks, participants were given a configuration of search tools to search for information.

The configuration of search tools available to participants varied based on our three *searchbot conditions* (Section 3.3). In the *browser_only* condition, participants were each given their own search interface to search independently. In the *slack_only* condition, participants were only able to issue queries inside of the shared Slack channel. Finally, in the *slack_browser* condition, participants were able to query the searchbot inside of Slack and also use their own independent search interface. All search tools used the same back-end system (i.e., the same document collection, retrieval model, and snippet-generation algorithm). All tools searched the New York Times (NYT) Corpus, which contains about 1.8 million NYT articles from 1987-2007.[1] As described in Section 3.2, our tasks were contextualized with this archival news collection in mind. All search tools were implemented using Lucene (v5.3.1) for retrieval and snippet-generation. For retrieval, we used Lucene's implementation of the query-likelihood model with Dirichlet smoothing and the default parameter values.

3.1 Study Protocol

Participants started the experiment by completing a demographic questionnaire. Then, they completed three search tasks, each associated with a different searchbot condition (Section 3.3). Each search task followed the same protocol. First, both participants came into the same room and the moderator explained the searchbot condition associated with the task (e.g., "In the next task, you will only be able to search inside of Slack using the searchbot"). Then, the moderator gave the participants a demonstration of how to interact with the available systems in the condition. Following this demonstration, the moderator read the search task to the participants and gave each a printed copy. Next, participants each went to their separate rooms to complete the task. Participants were given 15 minutes to complete each task. After completing each task, each participant completed a series of post-task questionnaires (Section 3.4). We used Camtasia software to record both participants' screens and Slack's API to log their communication. Additionally, we logged all query and click events inside and outside of Slack. Each participant was given $20 USD for participating in the study.

3.2 Search Tasks

Participants completed three *comparative* search tasks that we designed for the study. A comparative task is one that requires comparing different *items* (or alternatives) along different *dimensions* (or criteria). For example, a comparative "shopping task" might require comparing different products along dimensions such as their price, quality, and durability.

During each task, participants were asked to: (1) choose *three* items to consider, (2) compare the items along *three* dimensions, and (3) make a final item selection. Participants were given the exact three dimensions to consider, but *not* the items. Therefore, as part of the task, participants had to find (and agree upon) three candidate items. To compare the items (before making a selection), participants were asked to complete a 3x3 table representing the items (three rows) and dimensions (three columns). The 3x3 table was provided as a shared Google Spreadsheet. Participants were instructed to use the available search tools to find information and complete the 3x3 table (i.e., find information for each item-dimension pair). After completing the table, participants were instructed to mark their final item selection and provide a justification.

Participants completed three comparative tasks: (1) a sports personality task, (2) a lawsuits against Apple task, and (3) a sci-fi movie task. Each task included a background story (to contextualize the task) and an objective statement. Participants were assigned gender-neutral names during the experiment: Jamie and Taylor. The sports personality task had the following background story and objective:

Background: "Jamie and Taylor are journalism students and have been paired to work together on a class assignment. For this assignment, they need to write about an interesting sports personality covered in the New York Times from 1987-2007."

Objective: "Your goal is to compare/contrast THREE different sports personalities across the following THREE dimensions: (1) type of sport, (2) achievements, and (3) scandals. After completing the 3x3 table, please select the personality who you both think is the most interesting to write about based on the dimensions. Once you've made your choice, write a brief justification below the table."

Our other two tasks had the following three dimensions: Apple lawsuits task (company involved, subject of the lawsuit, outcome) and sci-fi movie task (plot, director, overall review).

[1] https://catalog.ldc.upenn.edu/ldc2008t19

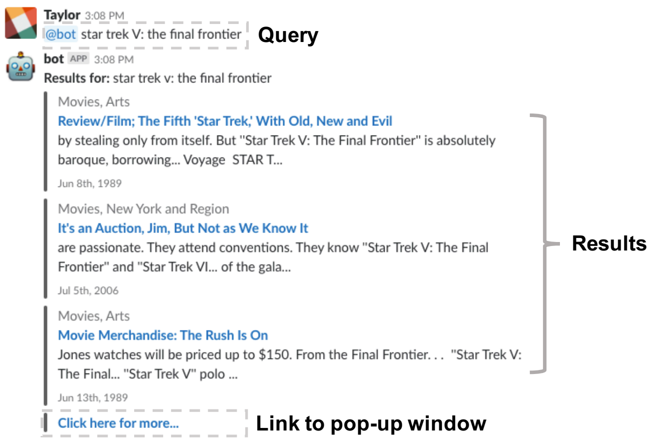


Figure 1: Searchbot query and results inside of Slack.

3.3 Searchbot conditions

Participants in our study experienced three searchbot conditions: *browser_only*, *slack_only*, and *slack_browser*. These conditions varied based on the search tools available to the participants.

In the *browser_only* condition, participants each used out-of-channel search tools. The goal of this condition was to simulate a scenario in which collaborators must use their own independent systems to search and use the messaging platform to coordinate and share their findings.

In the *slack_only* condition, participants could only issue queries inside of Slack. Participants could submit queries to the searchbot using the "@bot" preamble (e.g., "@bot Roger Federer"). We used Slack's API to instrument a searchbot that was able to automatically respond to queries and embed its search results directly inside of Slack. As shown in Figure 1, the searchbot embedded the top-three search results inside of Slack and also included a "click here for more" hyperlink that opened an independent (i.e., not shared) pop-up browser window with additional results. The pop-up window included pagination controls and displayed 10 results per page. In the *slack_only* condition, the pop-up window did *not* include a search box and participants were therefore not able to issue subsequent queries from their own independent pop-up windows. In other words, participants were forced to issue all queries by invoking the searchbot from the shared Slack channel.

Finally, in the *slack_browser* condition, participants could issue queries inside and outside of Slack. Participants had to issue the first query against the searchbot inside of Slack. However, after clicking the "click here for more" hyperlink, the browser pop-up with additional search results *also* included a search box that participants could use to search independently. In other words, participants were free to search inside of Slack by using the searchbot or to search independently from their own independent pop-up windows (after the initial searchbot query).

Participants were exposed to all three searchbot conditions and all three search tasks. We used two Latin Squares to counterbalance the presentation of searchbot condition and search task, for a total of 9 treatment orders that were repeated 3 times across our 27 participant pairs.

3.4 Post-task Questionnaires

At the end of each task, participants completed a series of post-task questionnaires. Two of these were designed by us and are provided in the appendix. First, participants completed the 6-item NASA-TLX questionnaire [1], which measures six aspects of workload: (1) mental demand, (2) physical demand (3) temporal demand, (4) failure, (5) effort, and (6) frustration.

Second, participants completed a 10-item questionnaire about their collaborative experience during the task (Table 2). We designed this questionnaire to capture three different aspects of participants' collaborative experience: (1) level of awareness of each other's activities (4 items), (2) level of collaborative effort (4 items), and (3) level of collaborative enjoyment (2 items).

Finally, in the *slack_only* and *slack_browser* conditions, participants completed a 6-item questionnaire about their experiences interacting with the searchbot inside of Slack (Table 3). This questionnaire measured two aspects of participants' experiences with the searchbot: (1) gains obtained from the searchbot (2 items) and (2) ease of coordinating the use of the searchbot (4 items).

In all three questionnaires, participants responded using a seven-point scale. For all NASA-TLX questions *except* for the failure question, participants responded using a seven-point scale, from "very low" (1) to "very high" (7). For the NASA-TLX question on failure, participants rated their performance from 'perfect' (1) to 'failure' (7). Thus, for all NASA-TLX questions, higher response values indicate higher workload. For the other two questionnaires, participants responded to agreement statements using a seven-point scale, from "strongly disagree" (1) to "strongly agree" (7). Except for the bot_part_distract item, high response values for these two questionnaires indicate positive experiences (i.e., high collaborative awareness and ease of coordinating the use of the searchbot).

3.5 Search Behaviors

As part of RQ4, we investigated whether participants exhibited different search behaviors across searchbot conditions. In this section, we describe novel measures developed for this study.

Query Intent Overlap: Our *query intent overlap* measure considers the degree of overlap between the search intents of both participants. Our goal was to measure the extent to which participants searched for the same types of information (high overlap) or engaged in more "division of labor" strategies (low overlap).

To compute this measure, we performed a qualitative analysis of all queries issued by participants (inside or outside of Slack). As previously mentioned, our tasks required participants to compare items along three dimensions. We manually coded participants' queries based on the item(s) and dimension(s) mentioned in the query (i.e., a two-dimensional coding scheme). The items and dimensions were indexed based on their order of appearance in the joint query stream from both participants.

To illustrate, Table 1 shows the first four queries from a query stream and their respective item and dimension codes. The first query is coded as (*,*) because it mentions no items and dimensions. The second query is coded as (1,*) because it mentions the first item in the stream and no dimension. The third query is coded as (*,1) because it mentions the first dimension in the steam and no item. Finally, the fourth query is coded as (2,2) because it mentions the second item and dimension in the query stream.

Table 1: Example item/dimension codes

Query	Item Code	Dimension Code
sci-fi movies	*	*
ET movie	1	*
sci-fi movie reviews	*	1
the matrix director	2	2

To test the reliability of our coding scheme, two of the authors coded 10% of all query streams. The Cohen's Kappa agreement for the item and dimension codes was $\kappa = 0.989$ and $\kappa = 0.917$, respectively. Given these levels of "near perfect agreement" [17], one of the authors coded the remaining 90% of all query streams.

Given our coding of all queries, we computed the following query intent overlap measure. Let $(I, D)_1$ denote the set of item-dimension codes associated with one participant's queries and let $(I, D)_2$ denote the set of item-dimension codes associated with the *other* participant's queries. We define our query intent overlap (QIO) measure as:

$$\text{QIO} = \frac{(I, D)_1 \cap (I, D)_2}{(I, D)_1 \cup (I, D)_2}$$

QIO is in the range [0,1], with high values indicating a high search intent overlap based on the item-dimension pairs explicitly mentioned in participants' queries.

Repeated Intent: We computed one measure that considered the degree to which participants re-issued queries with the same intent as another query from the *same* participant. To this end, we used the two-dimensional coding scheme described above. Our *repeated intent* (RI) measure considers the percentage of a participant's queries that had the same item-dimension codes as another query from the *same* participant. Again, RI is in the range [0,1], with high values indicating a high degree of query reformulation with the same intent (i.e., evidence of individual query refinement).

3.6 Data Analysis

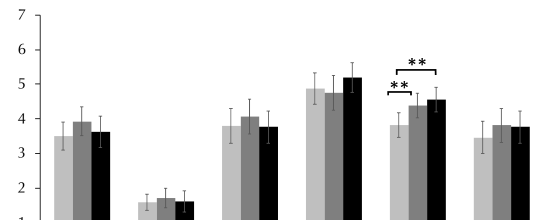
For RQ1-RQ4, we investigate the effects of the searchbot condition on different outcomes. We used mixed-effects regression models to analyze the effect of the searchbot condition on each outcome. Using mixed-effect models provided two advantages. First, in a few instances, participants did not answer every questionnaire item, resulting in imbalanced data. Mixed-effects models are well-suited for imbalanced repeated measures data [10]. Second, using mixed-effects models allowed us to account for random variations due to the search task ($n = 3$), the participant pair ($n = 27$), and the participant ($n = 54$). For outcome measures at the participant-pair level (e.g., total queries issued), we included task ID and the participant-pair ID as random effects. For outcome measures at the participant level (e.g., mental demand), we included as random effects the task ID and the participant ID nested within the participant-pair ID. We tested the statistical significance of each mixed-effects model by computing the χ^2 statistic using the likelihood-ratio test against a null model (i.e., one without the searchbot condition as a co-variate). To compare between all pairs of searchbot conditions, we ran analyses using the *browser_only* and *slack_only* conditions as baselines, and used a Holm-Bonferroni correction [13].

4 RESULTS

In the following sections, we present our results for RQ1-RQ4. In Section 5, we discuss the main trends and implications.

4.1 RQ1: Workload

To investigate RQ1, we analyzed participants' post-task responses to the NASA-TLX questionnaire, which measured six aspects of workload: (1) mental demand, (2) physical demand, (3) temporal demand, (4) failure, (5) effort, and (6) frustration. Figure 2 shows the means and 95% confidence intervals of participants' responses across searchbot conditions. Searchbot condition had a significant effect for effort ($\chi^2(2) = 11.757$, $p < 0.01$). Participants reported greater levels of effort during the *slack_only* ($\beta = 0.574$, S.E. = 0.220, $p < 0.01$) and *slack_browser* condition ($\beta = 0.741$, S.E. = 0.220, $p < 0.01$) as compared to the *browser_only* condition. One possible explanation is that the novelty of the searchbot required participants to expend more effort than when completing the task using familiar, non-integrated tools. Our results for RQ2-RQ4 provide insights about *why* the *slack_only* and *slack_browser* conditions required more effort.



	Mental Demand	Physical Demand	Temporal Demand	Failure	Effort	Frustration
browser_only	3.500	1.592	3.796	4.870	3.814	3.462
slack_only	3.925	1.703	4.074	4.759	4.388	3.814
slack_browser	3.629	1.611	3.759	5.185	4.555	3.759

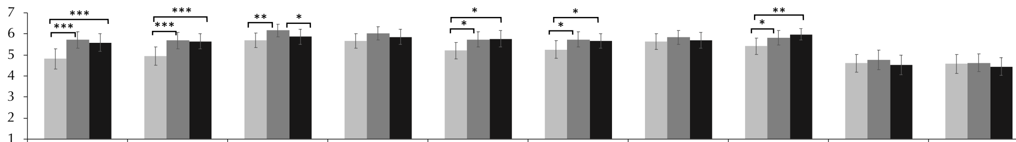
Figure 2: Post-task responses about workload across the searchbot conditions. '' denotes significant differences at $p < .01$ level.**

4.2 RQ2: Collaborative Experience

To investigate RQ2, we analyzed participants' responses to our post-task questionnaire about their collaborative experience: (1) awareness of each other's activities; (2) ease of collaboration (sharing information, coordinating, communicating, reaching consensus); and (3) enjoyment. Figure 3 shows the means and 95% confidence intervals of participants' responses about their collaborative experience across searchbot conditions.

Collaborative Awareness: Searchbot condition had a significant effect for three measures of awareness: (1) aware_partner ($\chi^2(2) = 19.718$, $p < 0.001$), (2) aware_self ($\chi^2(2) = 15.035$, $p < 0.001$), and (3) aware_part_info ($\chi^2(2) = 9.545$, $p < 0.01$).

In terms of aware_partner, participants reported greater awareness of their partner's activities in the *slack_only* ($\beta = 0.907$, S.E. = 0.210, $p < 0.001$) and *slack_browser* condition ($\beta = 0.778$, S.E. = 0.210, $p < 0.001$) as compared to the *browser_only* condition. In terms of aware_self, participants reported greater confidence that their partner was aware of their own activities in the *slack_only* ($\beta = 0.740$, S.E. = 0.207, $p < 0.001$) and *slack_browser* condition ($\beta = 0.704$, S.E. = 0.207, $p < 0.001$) as compared to the *browser_only* condition. Finally, in terms of aware_part_info, participants reported greater confidence that their partner was considering useful information in the *slack_only* condition as compared to the *browser_only* ($\beta = -0.463$, S.E. = 0.148, $p < 0.01$) and *slack_browser* condition ($\beta = -0.296$, S.E. = 0.148, $p < 0.05$).



	aware_partner	aware_self	aware_part_info	aware_self_info	ease_share	ease_cord	ease_comm	ease_cons	enjoy_me	enjoy_part
browser_only	4.814	4.944	5.703	5.667	5.203	5.259	5.629	5.425	4.611	4.574
slack_only	5.722	5.685	6.166	6.037	5.740	5.740	5.833	5.814	4.759	4.629
slack_browser	5.592	5.648	5.870	5.851	5.759	5.667	5.698	5.981	4.537	4.444

Figure 3: Post-task responses about collaborative experience across the searchbot conditions. '*', '', and '***' denote significant differences at $p < .05$, $p < .01$ and $p < .001$ level.**

Collaborative Effort: Searchbot condition had a significant effect for three measures of collaborative effort: (1) ease_share ($\chi^2(2) = 8.987$, $p < 0.05$), (2) ease_cord ($\chi^2(2) = 7.712$, $p < 0.05$), and (3) ease_cons ($\chi^2(2) = 9.447$, $p < 0.01$).

In terms of ease_share, participants reported greater ease in sharing information in the *slack_only* ($\beta = 0.537$, S.E. = 0.206, $p < 0.05$) and *slack_browser* condition ($\beta = 0.556$, S.E. = 0.206, $p < 0.05$) as compared to the *browser_only* condition. In terms of ease_cord, participants reported greater ease in coordinating with their partner in the *slack_only* ($\beta = 0.482$, S.E. = 0.183, $p < 0.05$) and *slack_browser* condition ($\beta = 0.407$, S.E. = 0.183, $p < 0.05$) as compared to the *browser_only* condition. Finally, in terms of ease_cons, participants reported greater ease in reaching consensus with their partner in the *slack_only* ($\beta = 0.389$, S.E. = 0.184, $p < 0.05$) and *slack_browser* condition ($\beta = 0.571$, S.E. = 0.184, $p < 0.01$) as compared to the *browser_only* condition.

Collaborative Enjoyment: Searchbot condition did not have a significant effect on either measure of collaborative enjoyment (enjoy_me, enjoy_part).

4.3 RQ3: Searchbot Experience

To investigate RQ3, we analyzed participants' responses to our post-task questionnaire about their experiences using the searchbot. This 6-item questionnaire was only given to participants in the *slack_only* and *slack_browser* conditions, and focused on two main themes: (1) ease of coordinating the use of the searchbot and (2) gains obtained from the searchbot. Figure 4 shows the means and 95% confidence intervals of participants' responses for the *slack_only* and *slack_browser* conditions.

Coordinating Searchbot Use: Searchbot condition had a significant effect on three measures associated with participants' experiences coordinating the use of the searchbot: (1) bot_part_intent ($\chi^2(1) = 4.676$, $p < 0.05$), (2) bot_part_info ($\chi^2(1) = 7.744$, $p < 0.01$), and (3) bot_part_distract ($\chi^2(1) = 7.884$, $p < 0.01$).

In terms of bot_part_intent, participants reported greater awareness of their partner's goals when interacting with the searchbot in the *slack_only* versus *slack_browser* condition ($\beta = 0.463$, S.E. = 0.209, $p < 0.05$). In terms of bot_part_info, participants reported greater awareness of the information their partner was requesting from the searchbot in the *slack_only* versus *slack_browser* condition ($\beta = 0.759$, S.E. = 0.263, $p < 0.01$). In terms of bot_part_distract, participants reported being more distracted by their partner's interactions with the searchbot in the *slack_only* versus *slack_browser* condition ($\beta = 0.574$, S.E. = 0.197, $p < 0.01$).

Gains: Searchbot condition did not have a significant effect on the gains from the searchbot (bot_useful_info, bot_saved_time).

4.4 RQ4: Search Behaviors

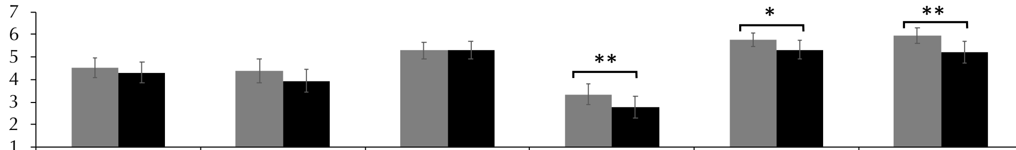
In RQ4, we investigate whether participants exhibited different search behaviors across searchbot conditions. We approach this question from three different perspectives. First, we consider measures that combine the activities of *both* participants (e.g., the total number of queries issued). Second, we consider the extent to which both participants issued queries with the same search intent. Third, we consider the extent to which individual participants issued queries with the same intent as a previous query from the *same* participant (i.e., evidence of individual query refinement).

Aggregate Measures: We investigate the effect of the searchbot condition on four aggregate measures: (1) total number of queries, (2) abandoned queries, (3) clicks, and (4) clicks per query. As shown in Figures 5a-5d, searchbot condition had a significant effect on three measures: (1) queries ($\chi^2(2) = 10.837$, $p < 0.01$), (2) abandoned queries ($\chi^2(2) = 14.581$, $p < 0.001$), and (3) clicks per query ($\chi^2(2) = 6.673$, $p < 0.05$).

In terms of queries, participants issued fewer queries in the *slack_only* condition as compared to the *browser_only* ($\beta = 2.481$, S.E. = 1.042, $p < 0.05$) and *slack_browser* condition ($\beta = 3.519$, S.E. = 1.042, $p < 0.01$). Similarly, participants had fewer abandoned queries in the *slack_only* condition as compared to the *browser_only* ($\beta = 2.111$, S.E. = 0.754, $p < 0.05$) and *slack_browser* condition ($\beta = 3.00$, S.E. = 0.754, $p < 0.001$). In terms of clicks per query, participants had more clicks per query in the *slack_only* condition as compared to the *browser_only* ($\beta = 0.476$, S.E. = 0.235, $p < 0.05$) and *slack_browser* condition ($\beta = 0.593$, S.E. = 0.235, $p < 0.05$).

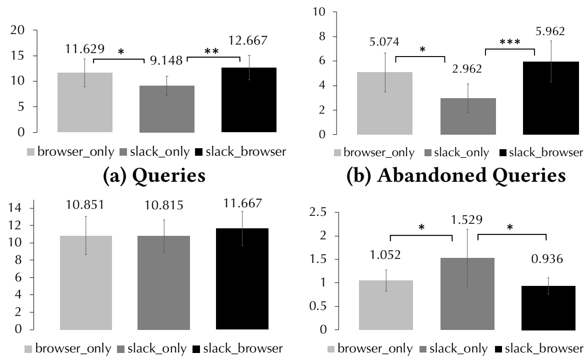
Query Intent Overlap: Our query intent overlap (QIO) measure considered the extent to which both participants issued queries with same search intent or engaged in more "division of labor" strategies. As shown in Figure 6, searchbot condition had a significant effect on QIO ($\chi^2(2) = 12.983$, $p < 0.01$). Participants' queries had a greater intent overlap in the *browser_only* condition as compared to the *slack_only* ($\beta = 0.162$, S.E. = 0.045, $p < 0.001$) and *slack_browser* condition ($\beta = 0.133$, S.E. = 0.045, $p < 0.01$).

Repeated Intent queries: Our repeated intent (RI) measure considered the extent to which individual participants issued queries with the same search intent (i.e., same item-dimension codes) as another query issued by the same participant, which we treat as evidence of individual query refinement. As shown in Figure 7, searchbot condition had a significant effect on RI ($\chi^2(2) = 14.966$, $p < 0.01$). Participants issued a greater percentage of queries with the same intent as another query in the *browser_only* ($\beta = 0.167$, S.E. = 0.053, $p < 0.001$) and *slack_browser* condition ($\beta = 0.198$, S.E. = 0.053, $p < 0.001$) as compared to the *slack_only* condition.



	bot_useful_info	bot_saved_time	coordinate_bot_use	bot_part_distract	bot_part_intent	bot_part_info
slack_only	4.518	4.388	5.296	3.351	5.778	5.962
slack_browser	4.314	3.944	5.296	2.778	5.314	5.203

Figure 4: Post-task responses about participants' experiences with the searchbot in the *slack_only* and *slack_browser* conditions. '*' and '' denote significant differences at $p < .05$, and $p < .01$ level.**



(a) Queries **(b) Abandoned Queries** **(c) Clicks** **(d) Clicks per Query**

Figure 5: Aggregate measures across searchbot conditions. '*', '', and '***' denote significant differences at $p < .05$, $p < .01$, and $p < .001$ level.**

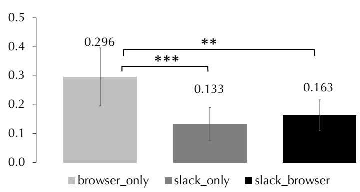


Figure 6: Query Intent Overlap. '' and '***' denote significant differences at $p < .01$ and $p < .001$ level.**

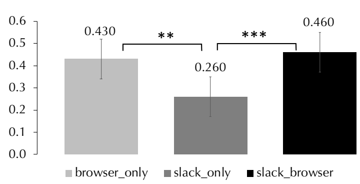


Figure 7: Repeated Intent (RI). '' and '***' denote significant differences at $p < .01$ and $p < .001$ level.**

5 DISCUSSION

Our goal in this study was to investigate different ways of embedding search tools into a communication platform (i.e., Slack). We investigated three conditions: *browser_only*, *slack_only*, and *slack_browser*. Here, we discuss our findings and implications.

RQ1: In terms of workload, participants reported greater effort in the *slack_only* and *slack_browser* conditions as compared to the *browser_only* condition (Figure 2). In other words, participants reported the least amount of effort in the condition where they completed the task using familiar, non-integrated tools (i.e., searching independently and coordinating through Slack). One possibility is that the novelty of the searchbot caused participants to expend more effort. Additionally, our results for RQ2-RQ4 (discussed below) provide insights as to *how* participants expended more effort in the *slack_only* and *slack_browser* conditions.

RQ2: Our results suggest that the searchbot improved participants' collaborative awareness and experience (Figure 3). In terms of awareness, participants reported significantly greater awareness of their partner's activities and greater confidence that their partner was aware of their own activities in the *slack_only* and *slack_browser* conditions. Between these two conditions, awareness was slightly greater in the *slack_only* condition. Specifically, participants reported greater confidence that their partner was searching for valuable information in the *slack_only* condition. In terms of their collaborative experience, participants reported significantly greater ease in sharing information, coordinating, and reaching consensus in the *slack_only* and *slack_browser* conditions. In relation to RQ1, these results suggest that the effort increase in the *slack_only* and *slack_browser* conditions did *not* originate from communication and coordination efforts.

Our RQ2 results suggest that integrating search tools into communication channels such as Slack provides some of the same benefits as standalone collaborative search systems [4, 8, 23, 25, 30, 33]. These findings are also consistent with prior work on searchbots that intervene in collaborations [3]. In that study, the searchbot raised collaborators' awareness of their activities and made it easier to share information, coordinate, and reach consensus.

RQ3: In terms of participants' experiences with the searchbot, our results suggest two main trends (Figure 4). First, in the *slack_only* condition, participants reported greater awareness of their partner's motivations for interacting with the searchbot. Most likely, this greater awareness came from participants being forced to share the same querying environment.

Second, in the *slack_only* condition, participants reported being more distracted by their partner's interactions with the searchbot. This result resonates with prior research on group dynamics and shared control of devices. Prior work shows that teamwork involves periods of tightly coupled *group* work and loosely coupled *individual* work [6, 9, 20]. Furthermore, prior studies have found that being aware of other's activities can be burdensome at times when it is not beneficial (e.g., during periods of individual work) [2, 24]. Our RQ3 results suggest that forcing users to search within the communication channel raises awareness, but can also be distracting during periods of individual work. Again, in relation to RQ1, our results suggest a possible source of effort associated with the searchbot—participants were distracted by their partner's interactions with the searchbot during periods of individual work.

RQ4: Our results suggest that the searchbot condition had an effect on participants' behaviors, both as partners and individual searchers. First, based on our qualitative coding of all queries issued, participants' search intents had less overlap in the *slack_only* and *slack_browser* conditions (Figure 6). One possibility is that the greater awareness in these two conditions allowed participants to avoid searching for the same information. Capra *et al.* [4] also observed that awareness features in a dedicated collaborative search system allowed participants to avoid each other's search paths.

Second, in the *slack_only* condition, participants issued significantly fewer queries, had fewer abandoned queries, and more clicks per query than in the other two conditions (Figures 5). One possibility is that participants in the *slack_only* condition issued fewer queries to avoid distracting their partners. As a negative outcome of issuing fewer queries, individual participants in the *slack_only* condition issued fewer queries with a repeated intent (Figure 7). This result points to another source of effort introduced by the searchbot (RQ1)—participants in the *slack_only* condition tried to do more with fewer (potentially sub-optimal) queries. Greater levels of query refinement in the *browser_only* and *slack_browser* conditions might have led to better search results and less effort.

Summary and Implications: Our results for RQ1-RQ4 suggest opportunities and challenges for designing search tools that can be embedded into communication channels such as Slack.

In terms of opportunities, access to the searchbot provided several benefits. When given access to the searchbot, participants reported greater awareness of each other's activities and less effort in sharing information, coordinating, and reaching consensus. Moreover, the increased awareness from the searchbot may have allowed participants to avoid searching for the same type of information (i.e., avoiding duplicated work).

In terms of challenges, access to the searchbot also had negative consequences. First, in some cases, participants found their partner's interactions with the searchbot to be distracting. Based on prior work, these distractions probably occurred at times when participants were engaged in individual work and the awareness of their partner's activities was not beneficial [2, 24]. Second, our results suggest that participants in the *slack_only* condition completed the task using fewer queries that were possibly sub-optimal. Participants in the *slack_only* condition (vs. both other conditions) issued fewer queries, had fewer abandoned queries, more clicks per query, and fewer query-reformulations with the same intent (i.e., less query refinement). Finally, the novelty of the searchbot might have also added some overhead for participants. Participants reported the least amount of overall effort in the baseline condition, in which they had to search independently (no searchbot).

Overall, our results suggest that the *slack_browser* condition provided important benefits compared to the other two conditions. Compared to the *slack_only* condition, individual participants in the *slack_browser* condition issued more queries with a repeated intent, which possibly led to better search results. Compared to the *browser_only* condition, participants in the *slack_browser* condition reported greater awareness, lower coordination effort, and had lower levels of query intent overlap (i.e., less duplicated work). While the *slack_browser* condition provided benefits, it also required more effort than the *browser_only* condition. Future work should build on this design and identify sources of effort that can be alleviated.

The increasing popularity of messaging platforms such as Slack is an opportunity to develop lightweight search tools that users can interact with directly from the platform. Tools such as searchbots are not only well-positioned to assist individual users, but also *multiple* users collaborating on tasks involving search. Our results have implications for designing searchbots to support collaborative search. First, our results suggest that it is important to support *both* tightly coupled group work and more loosely coupled individual work. In our case, forcing collaborators to share the same querying environment resulted in less trial-and-error at the individual work level (i.e., less individual query refinement). Less trial-and-error may have negative consequences—sub-optimal search results in the short-term and lower levels of exploration and learning in the long-term. Second, greater collaborative awareness has benefits (e.g., less duplicated work), but also comes at a cost (e.g., can be distracting at times when it is not beneficial). Future research should develop systems that dynamically adjust their awareness features based on collaborators' current activities. In this respect, a system might be able to adjust the *amount* of collaborative awareness provided (i.e., which activities are communicated), how *saliently* the awareness is communicated, and whether the awareness is provided in real-time versus "cached" for collaborators to use later.

6 CONCLUSION

We reported on a study that investigated different ways of integrating search tools into messaging platforms such as Slack. Specifically, we investigated different searchbot designs. In our study, 27 participant-pairs collaborated on tasks using Slack and were exposed to three searchbot conditions: (1) only searching independently outside of Slack, (2) only searching inside of Slack using the searchbot, and (3) searching both inside of Slack (using the searchbot) and outside of Slack (using independent interfaces).

Our results suggest opportunities and challenges in designing searchbots to support collaborative search. In terms of opportunities, access to the searchbot increased participants' awareness of each other's activities, made it easier to communicate and coordinate, and resulted in fewer duplicated searches. In terms of challenges, our results suggest that an effective searchbot design must satisfy different needs. It must raise awareness of collaborators' activities in cases where the awareness is beneficial. Additionally, the design must allow users to pursue their individual low-level strategies without intruding on each other's "space". In our study, forcing participants to share the same querying environment (only using the searchbot) resulted in fewer queries with the same intent (i.e., less query refinement) and possibly sub-optimal search results and task outcomes.

Outside of IR, the CSCW community has also identified collaborative agents (e.g., personal assistants, bots, and embodied robots) as an important area for future research [7, 16, 26]. Our findings may help inform the design of search-based agents that support remote or co-located synchronous collaborations.

7 ACKNOWLEDGEMENTS

This work was supported in part by NSF grants IIS-1451668 and IIS-1552587. Any opinions, findings, conclusions, and recommendations expressed in this paper are the authors' and do not necessarily reflect those of the NSF.

REFERENCES

[1] National Aeronautics and Space Administration Human Systems Integration Division. 2016. TLX NASA Ames. (2016).

[2] Saleema Amershi and Meredith Ringel Morris. 2008. CoSearch: a system for co-located collaborative web search. In *CHI*. ACM, 1647–1656.

[3] Sandeep Avula, Gordon Chadwick, Jaime Arguello, and Robert Capra. 2018. SearchBots: User Engagement with ChatBots During Collaborative Search. In *CHIIR*. ACM, 52–61.

[4] Robert Capra, Annie Chen, Katie Hawthorne, Jaime Arguello, Lee Shaw, and Gary Marchionini. 2012. Design and Evaluation of a System to Support Collaborative Search. In *ASIST*. Wiley.

[5] Robert Capra, Gary Marchionini, Javier Velasco-Martin, and Katrina Muller. 2010. Tools-at-hand and Learning in Multi-session, Collaborative Search. In *CHI*. ACM, 951–960.

[6] Mary Elwart-Keys, David Halonen, Marjorie Horton, Robert Kass, and Paul Scott. 1990. User interface requirements for face to face groupware. In *CHI*. ACM, 295–301.

[7] Lia Emanuel, Joel Fischer, Wendy Ju, and Saiph Savage. 2016. Innovations in Autonomous Systems: Challenges and Opportunities for Human-agent Collaboration. In *CSCW*. ACM, 193–196.

[8] Gene Golovchinsky, Jeremy Pickens, and Abdigani Diriye. 2011. Designing for Collaboration in Information Seeking. In *HCIR*.

[9] Roberto Gonzales-Ibanez, Muge Haseki, and Chirag Shah. 2013. Let's search together, but not too close! An analysis of communication and performance in collaborative information seeking. *Information Processing & Management* 49, 5 (2013), 1165 – 1179.

[10] Ralitza Gueorguieva and John H Krystal. 2004. Move over anova: progress in analyzing repeated-measures data andits reflection in papers published in the archives of general psychiatry. *Archives of general psychiatry* 61, 3 (2004), 310–317.

[11] Marti A. Hearst. 2014. What's Missing from Collaborative Search? *Computer* 47, 3 (2014).

[12] B. Hecht, J. Teevan, M.R. Morris, and D. Liebling. 2012. SearchBuddies: Bringing Search Engines into the Conversation. In *ICWSM*. AAAI, 138–145.

[13] Sture Holm. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics* (1979), 65–70.

[14] Nyi Nyi Htun, Martin Halvey, and Lynne Baillie. 2017. How Can We Better Support Users with Non-Uniform Information Access in Collaborative Information Retrieval?. In *CHIIR*. ACM, 235–244.

[15] Shahram Izadi, Harry Brignull, Tom Rodden, Yvonne Rogers, and Mia Underwood. 2003. Dynamo: a public interactive surface supporting the cooperative sharing and exchange of media. In *UIST*. ACM, 159–168.

[16] Malte F. Jung, Selma Šabanović, Friederike Eyssel, and Marlena Fraune. 2017. Robots in Groups and Teams. In *CSCW*. ACM, 401–407.

[17] J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics* (1977), 159–174.

[18] Andrés Lucero, Jussi Holopainen, and Tero Jokela. 2012. MobiComics: collaborative use of mobile phones and large displays for public expression. In *MobileHCI*. ACM, 383–392.

[19] Andrés Lucero, Jaakko Keränen, and Hannu Korhonen. 2010. Collaborative use of mobile phones for brainstorming. In *MobileHCI*. ACM, 337–340.

[20] Munir Mandviwalla and Lorne Olfman. 1994. What do groups need? A proposed set of generic groupware requirements. *TOCHI* 1, 3 (1994), 245–268.

[21] Meredith Ringel Morris. 2008. A Survey of Collaborative Web Search Practices. In *CHI*. ACM, 1657–1660.

[22] Meredith Ringel Morris. 2013. Collaborative Search Revisited. In *CSCW*. ACM, 1181–1192.

[23] Meredith Ringel Morris and Eric Horvitz. 2007. SearchTogether: An Interface for Collaborative Web Search. In *UIST*. ACM, 3–12.

[24] Meredith Ringel Morris, Andreas Paepcke, and Terry Winograd. 2006. Teamsearch: Comparing techniques for co-present collaborative search of digital media. In *TABLETOP*. IEEE.

[25] Sharoda A. Paul and Meredith Ringel Morris. 2009. CoSense: Enhancing Sensemaking for Collaborative Web Search. In *CHI*. ACM, 2007–2012.

[26] Martin Porcheron, Joel E. Fischer, Moira McGregor, Barry Brown, Ewa Luger, Heloisa Candello, and Kenton O'Hara. 2017. Talking with Conversational Agents in Collaborative Action. In *CSCW*. ACM, 431–436.

[27] Martin Porcheron, Joel E. Fischer, and Sarah Sharples. 2017. "Do Animals Have Accents?": Talking with Agents in Multi-Party Conversation. In *CSCW*. ACM, 207–219.

[28] Sindunuraga Rikarno Putra, Felipe Moraes, and Claudia Hauff. 2018. SearchX: Empowering Collaborative Search Research. In *SIGIR*. ACM, 1265–1268.

[29] Daniel M Russell, Clemens Drews, and Alison Sue. 2002. Social aspects of using large public interactive displays for collaboration. In *UbiComp*. Springer, 229–236.

[30] Chirag Shah. 2010. Coagmento-a collaborative information seeking, synthesis and sense-making framework. In *CSCW*. ACM, 6–11.

[31] Chirag Shah. 2012. *Collaborative Information Seeking: The Art and Science of Making the Whole Greater Than the Sum of All*. Springer.

[32] Chirag Shah. 2013. Effects of awareness on coordination in collaborative information seeking. *JASIST* 64, 6 (2013).

[33] Zhen Yue, Shuguang Han, and Daqing He. 2012. An investigation of search processes in collaborative exploratory web search. In *ASIST*. Wiley.

8 APPENDIX

Table 2: Post-task questions about collaborative experience.

Theme	Tag	Description
Awareness	aware_partner	During the task, I had a pretty good idea about what my partner was doing.
	aware_self	During the task, I think my partner had a pretty good idea about what I was doing.
	aware_part_info	During the task, I was confident that my partner was looking at information I would find valuable for completing the task.
	aware_self_info	During the task, I was confident that I was looking at information my partner would find valuable for completing the task.
Effort	ease_share	It was easy to share information with my partner during the task.
	ease_coord	It was easy for my partner and I to coordinate our efforts during this task.
	ease_comm	It was easy to communicate with my partner during this task.
	ease_cons	It was easy for my partner and I to reach consensus during this task.
Enjoyment	enjoy_me	I enjoyed completing this task.
	enjoy_part	I think my partner enjoyed completing this task.

Table 3: Post-task questions about the searchbot.

Theme	Tag	Description
Gains	bot_useful_info	The *Searchbot* provided us with useful information.
	bot_saved_time	The *Searchbot* saved me and my partner some time.
Coordination	coordinate_bot_use	It was easy for me and my partner to coordinate our use of the *Searchbot* during this task.
	bot_part_distract	I found my partner's interactions with the *Searchbot* to be distracting.
	bot_part_intent	I always knew why my partner was interacting with the *Searchbot*.
	bot_part_info	I always knew what information my partner was requesting from the *Searchbot*.