STOPPING AND RESUMING: HOW AND WHY DO PEOPLE SEARCH ACROSS SESSIONS FOR COMPLEX TASKS?

Yuan Li

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the School of Information and Library Sciences.

Chapel Hill
2023

Approved by:

Robert Capra

Jaime Arguello

David Gotz

Marijel (Maggie) Melo

Dan Wu

ABSTRACT

Yuan Li: Stopping and Resuming: How and why do people search across sessions for complex tasks?
(Under the direction of Robert Capra)

Cross-session searches (XSS) occur when people look for information online for multiple sessions to complete complex task goals over time. Previous studies explored aspects of XSS, including the reasons that lead to it, like the Multiple Information Seeking Episode (MISE) model, which highlights eight causes. However, less is known about how these reasons manifest in real-life XSS and their relationship with task characteristics.

I conducted a diary study with 25 participants engaging in XSS for real-life tasks. Participants reported on at least three search sessions spanning at least two days, and 15 participants attended an interview after they completed the diary study. We used qualitative methods to explore motivations for expected XSS, goal complexity, session resuming and stopping reasons, types of found information, cognitive activities, and the non-search task activities that happened during the XSS process. Our results validated and refined the MISE session resuming and stopping reasons and distinguished subcategories and reasons unique to real-life XSS tasks. We discerned task-oriented and cognition-oriented motivations for XSS. We identified seven types of non-search task activities and three popular modes describing how people intertwine search activities with non-search activities during XSS.

We assessed relationships among factors including session goal complexity, information types, cognitive activities, session resuming, and stopping reasons using quantitative methods. Our results show significant associations between information types, cognitive activities, session goal complexity, and session resuming and stopping reasons. Furthermore, task stages significantly correlate with perceived overall task difficulty and the difficulty to find enough information. We also identified five XSS-specific challenges.

Our results have implications for tailoring future search engines to customize search results according to session resuming reasons and designing tools to assist task management and preparation for session stops. Methodologically, our results have insights into designing tasks and subtasks and controlling the reasons that can lead to successive searches for tasks with varying complexity.

This work is dedicated to my mother, Gailian, and my daughter, Caroline, whose resilience and fortitude have instilled in me an invaluable appreciation for life.

ACKNOWLEDGEMENTS

Reflecting on my journey, I realize that obtaining a Ph.D. was far from my mind when I first set foot on American soil in October 2011. This achievement, however, is not mine alone. It was made possible by the boundless love, unwavering support, abundant inspiration, and disciplined guidance gifted to me by numerous individuals to whom I owe my deepest gratitude.

First and foremost, I must express my profound gratitude to Dr. Rob Carpa. Thank you, Rob, for taking me under your wing when I was full of doubt and hesitations. Thank you for seeing potential in me, even when I was unsure of myself. You accepted me as your student, guided me every step of the way, and taught me the true essence of being a researcher. Your inspiration, availability, and endless support have been the pillars I leaned on throughout my academic journey. The role model you represent is exactly the kind of person I strive to be. Thank you for everything, truly.

I sincerely appreciate my committee members: Dr. Jaime Arguello, Dr. David Gotz, Dr. Maggie Melo, and Dr. Dan Wu. Your insightful suggestions, constructive feedback, and encouragement have been instrumental in my achievement of the Ph.D. Jaime, your invaluable advice, collaboration, and mentorship have enriched my experiences in the lab and beyond. Maggie, your open door and attentive ear provided the comfort I needed during both research discussions and professional life events. David and Dan, your meticulous attention and thoughtful considerations have been instrumental in shaping my dissertation study.

I want to extend my warm thanks to Dr. Barbara Wildemuth and Dr. Jingjing Liu. Barbara, thank you for your faith in my capabilities and constant concern for my well-being, which have been an enduring source of strength since my first day at SILS till now. Jingjing, my Master's advisor, thank you for opening the doors of Information Science to me at the University of South Carolina. Your guidance then was the stepping stone to my later exploration of this intricate field.

I am also grateful to Dr. Gary Marchionini, Dr. Ron Bergquist, Dr. Brian Sturm, Dr. Ryan Shaw, and Dr. Yue (Ray) Wang. Gary, thank you for your availability and attention during the research discussions and presentations in our lab. Ron, thank you for the one-on-one teaching class,

which was instrumental in shaping my pedagogical approach. Brian, thank you for your indispensable assistance during the difficulties brought about by the pandemic. Ryan, your willingness to allow me to observe your teaching sessions enriched my skillset. Ray, your enthusiastic support made our Chinese New Year celebration unforgettable.

I want to thank my colleagues at SILS: Anita Crescenzi, Yinglong Zhang, Austin Ward, Sandeep Avula, Bogeum Choi, Shenmeng Xu, Kelsey Urgo, Colin Post, Yuanye Ma, Wenyuan Wang, and Yuyu Yang. Thank you for the stimulating discussions, collaborative work, and emotional support. You are my academic family and have been invaluable to my time at UNC. A special mention to Anita, Yinglong, Austin, Bogeum, and Kelsey. Thank you for all the support and love you have given to me. You have made this journey an extraordinary experience.

My best friend, Mengting Dai, warrants special recognition. You have provided a refuge for me when I needed a break from research to appreciate the simpler aspects of life. Our shared experiences - from leisurely brunches and laughter to fun-filled trips - have brought joy to my life.

My heartfelt gratitude extends to my family, who are the cornerstone of my life. I offer my deepest gratitude to my husband, Jia. You've been my pillar of strength throughout this journey. Your strong support, determination, and motivation are the inspiration behind my success. To my daughter, Caroline, my precious treasure, you've brought a beautiful new purpose to my life. I relish watching you grow every day. And to my parents-in-law, your enormous support and understanding are invaluable and have made this journey smoother.

Lastly, but most importantly, I owe a debt of gratitude to my parents. Thank you for the unconditional love and endless support no matter the miles and time that separate us. And my brother, Yuzhong, you have been a rock in my life, always check on me. I am deeply grateful for your love.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

Cross-session search (XSS), also known as multi-session search, describes situations where people conduct a series of information search activities across multiple different sessions (e.g., time periods of days) to achieve a single goal. XSS is ubiquitous. Sellen et al. (2002) showed that 40% of knowledge workers' information-gathering tasks continued over a long period. Morris et al. (2008a) reported that 83% of surveyed information workers had engaged in XSS. Nearly all knowledge workers (n=20) in Backhausen, Klas, and Hemmje (2015) reported that they regularly work on tasks across days.

Prior research has mainly focused on three sub-areas of cross-session search: (1) identifying the characteristics of cross-session search, e.g., task topics, search task types, and time gaps (Spink, Bateman, & Jansen, 1999; MacKay & Watters, 2008b; Morris et al., 2008a); (2) predicting task continuity based on the analysis of large-scale search log data (Agichtein, White, Dumais, & Bennet, 2012; Czerwinski, Horvitz, & Wilhite, 2004; Kotov, Bennett, White, Dumais, & Teevan, 2011); (3) developing search assistant tools to help users saving search history and manage their tasks (MacKay & Watters, 2008b; Morris et al., 2008a; Jhaveri & Räihä, 2005; Gomes & Hoeber, 2021).

As in many cases, it is also described that people search for multiple sessions because their tasks are complex (J. Liu & Belkin, 2010; Czerwinski et al., 2004; Agichtein et al., 2012; Backhausen et al., 2015). Pioneering researchers have explored some reasons that lead to cross-session search. Spink, Bateman, and Jansen (1999) found that academic users often start a successive search to enhance or refine what they saw earlier or find additional information from library databases. MacKay and Watters (2008b) found that college students often stopped a session because they completed their tasks or wanted to do something else. Lin (2001) developed a theoretical model: Multiple Information Seeking Episode (MISE), which conceptualizes eight renewal reasons why people search across sessions and two types of interruptions that cause them to stop a session (i.e., external interruptions and internal interruptions). In their later study, Lin and Belkin (2005) validated this model in an experimental study with simulated tasks. Backhausen et al. (2015) found that people often reported frequent goal-changing or task-changing during XSS via a survey study.

However, there is a paucity of studies to systematically investigate real-life cross-session search task complexity in their original context. Little attention has been paid to the relationships between the stopping and resuming reasons leading people to search multiple sessions, and few studies focus on the effects of task factors on those reasons or users' search experience across sessions. These motivate my study in this area.

To gain a broad understanding of real-world XSS characteristics and explore whether the previous reasons can be generalized to real-life examples, I conducted a preliminary survey study using Amazon Mechanical Turk (AMT). As a crowdsourcing platform, AMT allows researchers to recruit diverse participants (Dupuis, Endicott-Popovsky, & Crossler, 2013). I used a modified Critical Incident Technique (CIT) strategy to design the survey questions in the survey. Using CIT allows me to systematically identify important aspects of an event experienced by a person (Radford, 2006). The questionnaire contained four sections: (1) basic demographic information (e.g., age, gender, educational background), (2) questions about a recent task or project involving information search for multiple sessions, (3) questions about the most recent search session they conducted for the selected task, (4) questions about the methods they used for keeping and transferring the information found between sessions. My question formats included open-response questions, multiple-choice questions, and seven-point Likert-type questions. I piloted the survey several rounds for revising. The final version took about 15-20 minutes to complete. Then I distributed the survey across two weeks and sent multiple batches of the survey at different times of the day (e.g., mornings, afternoon, evening) and on different days (i.e., weekdays and weekends) on AMT, intended to attract the attention of workers from diverse user groups.

As a result, my survey study validates the XSS stopping and resuming reasons conceptualized in the MISE model by showing that all these reasons can be found in real-life XSS processes. Furthermore, my findings also reveal that the frequency and appearance of the stopping and resuming reasons are related to and affected by other factors (i.e., information types, search sessions, task stages) in real-life XSS examples.

Meanwhile, we found that the reasons causing people to stop or resume their searches for the same task over time are complicated. Although many previous studies (e.g., Morris et al. (2008a), Jhaveri (2004), Bharat (2000)) describe that re-finding previously found information is crucial for complex search over time, we find that there are only a small portion (11%) of our participants reported to resume their search for re-finding. Interestingly, 36% of our participants said they did not need the information they saw earlier since it was less useful. This point is especially important since many prior tools designed to help users with XSS have focused on helping users “get back” to previously found information/searches. Instead, looking for updated information, continuing to find more information, and finding information related to sub-topics are the most popular reasons (22%, 17%, 15%, respectively) leading people to resume their search. Other resuming reasons include: *rolling back*, meaning that the old information did not work and they need to find new solutions; *transmuting*, meaning that they need to find information to help understand the task requirement; *anticipated*, meaning that they want to find information that is helpful in the future.

Additionally, participants reported a variety of reasons why they stopped a search session with different frequencies. As the *found all needed information* is the top session stopping reason, the other three most popular reasons that lead them to stop temporarily are *need to process, validate the found information*, and *consult other sources*. *Cannot find the needed information* and other non-task-related interruptions also caused stops during cross-session searches. These findings suggest that people have various needs when they search across sessions besides reviewing or re-finding the information they saw in past search sessions.

Based on the collected data, we also analyzed aspects related to cross-session search, such as cross-device search behavior, the information types users needed for their tasks, and the cognitive behaviors required for completing cross-session tasks. We also explored the relation trends between different session resuming reasons and stopping reasons and how they are influenced by task factors extracted from the survey data (i.e., needed information types, search sessions, and task performance stages). Our analysis found that these reasons change across different task factors and showed several general patterns. For instance, (1) *found all needed information* and *need to stop to process the found information* are the two most popular reasons that make people stop a search session; (2) People usually need more specific information than general information, but general information (e.g., background information) is very helpful when they resumed the session for looking for sub-topics, or when it is hard to find the answer by previous sessions or when they just check for updated information; (3) the early sessions (e.g., first two sessions) are often started for looking for sub-problems, but as people search for more sessions, the diversity of resuming reasons increases; (4) as people search for more sessions, they stopped more often for validating the found information.

However, due to the limitations of the survey study, we are unable to collect sufficient details about participants' information activities and the corresponding reasons for each session associated with the entire task. In addition, participants' responses to the survey may be biased by recall accuracy. To further investigate users' XSS behavior, the motivations, and triggers that lead them to search across sessions, and provide a more comprehensive, updated view of XSS in today's work and life, I designed a diary study, together with other methods (i.e., questionnaire, interview) in my dissertation. In particular, my dissertation aims to answer the following five research questions:

- **RQ1:** What are the stopping and resuming reasons leading people to search for multiple sessions? Are the reasons related and how?

- **RQ2:** What are the effects of session-level search goal complexity on the types of information sought, the cognitive activities involved during the session, and the resuming/stopping reasons of the session?

  RQ2a: What are the effects of session-level goal complexity on the types of information sought during the session?

  RQ2b: What are the effects of session-level goal complexity on the cognitive activities involved during the session?

  RQ2c: What are the effects of session-level goal complexity on the session's resuming/stopping reasons?

- **RQ3:** What are the effects of the information types used and cognitive activities involved during a session on the session resuming/stopping reasons?

  RQ3a: What are the effects of the information types used on the stopping/ resuming reasons at the session level?

  RQ3b: What are the effects of the cognitive activities involved on the resuming/stopping reasons at the session level?

- **RQ4:** What are the effects of task stages on participants' perceptions of pre-task, post-task, and post-session difficulties?

RQ4a: What are the effects of task stages on participants' perception of the overall difficulty of their tasks and search sessions?

RQ4b: What are the effects of task stages on participants' perception of the difficulty to search for information?

RQ4c: What are the effects of task stages on participants' perception of difficulty to find enough information?

RQ4d: What are the effects of task stages on participants' perception of difficulty to integrate the found information?

RQ4e: What are the relationships between participants' perception of difficulty and the session resuming/stopping reasons?

- **RQ5:** What challenges did participants face due to conducting searches across multiple sessions?

This dissertation will have a positive impact on the following three subareas in the fields of Human-Computer Interaction (HCI), Information Science (IS), and Information Retrieval (IR).

Our findings can help differentiate users' information needs and search intentions for a task that spans multiple search sessions. In our work, we collect data from survey and diary studies of real-life tasks for investigating the stopping and resuming reasons in-situ. Our preliminary findings from the survey study show that re-finding information and the need for reacquainting with the found information only take a small portion of the reasons for resuming a search session. Our results validate Lin and Belkin's (Lin & Belkin, 2005) theoretical session renewal modes and indicate that these reasons can be generalized to everyday work and life cross-session tasks. Furthermore, our initial analysis also shows that these reasons are not isolated. Instead, they are related and affected by other task factors, such as needed information types, completed search sessions, and users' task performance stages. As users found more information, their information needs evolved, leading to their search activities. A better understanding of users' information intents over time can inform the future design of search systems. Search results and recommendations can be more relevant and useful if a search system can better understand users' needs and predict their future demands based on different reasons for searching multiple sessions.

Our findings have implications for re-designing search engine pages to support learning and long-term projects. Our findings can help recommend potential useful search results and follow-on actions to facilitate task completion. Based on our understanding of users' evolving information needs and the potential reasons that trigger them to continue to search for the same task, search systems can consider customizing the search results not only by the relevance to the topic but also by the needed information types (e.g., general background information, factual information, reviews) and support different types of cognitive activities involved in different search sessions (e.g., compare and evaluate search results, organize and synthesize information from various sources). For example, for a person who self-studies a new language, a search system can detect the learner's purposes based on the queries used during the search, and the clicked search results, then make appropriate recommendations. For a student who is writing a research paper, a search system can suggest outlines and strategies, as it detects the writer's task needs. Recent studies found that people learn during search and that many different types of cognition are involved, such as analyzing, evaluating, and creating. Future search engines can consider providing different types of information (e.g., conceptual, opinions, procedural knowledge) to facilitate users' cognitive activities during the learning process, especially for tasks that spread over time and involve complicated components. Besides, we could also explore ways to put users in the loop, for instance, designing new features to allow them to input more information about their tasks. Consequently, search engines can use this contextual information to customize the search results further.

My research envisions that future search systems can help users accomplish more, rather than just finding/presenting more and leaving the rest to users. As we said earlier, the search isn't an end per se. Instead, in most cases, it is probably just the beginning of a task where people start to explore the topic area. The task could be as important as making a real estate purchase, conducting a research study, or just looking for information for curiosity. Throughout the search process, users' knowledge of the topic, their perception of the task requirement, and their understanding of what they need to keep evolving. Future search engines should not only focus on providing search results *relevant* to the input queries for single session search but also support participants' evolving cross-session information needs in completing broader work tasks. From this perspective, we believe our study will provide insights into the development of future-generation search systems.

Following this introduction, there are six chapters in this dissertation: In Chapter 2, I provide an overview of previous studies on cross-session search and a review of relevant task factors, especially task complexity and task difficulty, I also reviewed the work that investigated task process and task analysis that researchers commonly use in Human-computer interaction (HCI) for studying task stages and decomposing task components, which can help us untangle the complex cross-session search tasks. In Chapter 3, I elaborated more details about the five main research questions and sub-research questions of this dissertation. I designed a *cross-session search study framework* that presents the context and crucial components of the cross-session search for my study. Then I provided explanations for each research question and sub-research question, the key factors and elements needed to answer these questions, and how I would collect appropriate data to achieve my goals. In Chapter 4, I demonstrated the design of the diary study step by step, including participant recruiting and screening, introduction meetings with participants, semi-structured diaries and questionnaires that require participants to record during the study period, and retrospective interview questions. I also introduce data analysis methods and fundamental theories for my analysis. In Chapter 5 and Chapter 6, I presented our observations and findings based on quantitative and qualitative analyses and discussed the connections and differences between this study with previous research related to cross-session search behaviors. In Chapter 7, I briefly reviewed our findings and the limitations of the dissertation study.

CHAPTER 2

Background

This dissertation study is built upon previous research on cross session search (XSS) and task-based information interaction. Prior researchers have investigated XSS from various perspectives: task topics, search session features, reasons leading people to search for multiple sessions, and investigation of search activities, etc., which provide us with rich descriptions of different aspects of the XSS behavior. Meanwhile, task-based information interaction research focuses on the impact of tasks (e.g., simulated work tasks, search tasks) on users' information behaviors and cognitive activities performed during search processes. In the following sections, I provide an overview of the XSS and literature review on task factors, including task complexity, task difficulty, task process, and task analysis.

2.1 Cross-session search (XSS)

People search for information to fulfill various needs in daily life and work. Some of these searches are simple and can be quickly completed in a single search session. However, in other cases, people may need to search multiple times to complete a task or solve a problem. A series of search activities to achieve a single goal, but spread across multiple sessions, is often referred to as *multi-session search*, or *cross-session search* (XSS) (Lin, 2001). In a single-session search, the task's search activities are performed all within one well-defined time period. However, XSS activities are spread across multiple different sessions (e.g., time periods or days). And the search activities may interleave with other actions relevant to the task.

XSS is ubiquitous. Since the interactive information retrieval system was still in the infancy stage, researchers already noticed that end-users do not limit their interaction with an information system to a single search session when seeking information related to a particular information problem. For instance, an early study on mediated online searching with 40 academic users found that 18 of them frequently interacted with search intermediaries to search for information on the same topic (Saracevic, 1991). Several other studies also found similar trends among academic library users: M. Huang (1992) found that students and faculty members conducted semester-long projects that involve multi-session search. Bates, Wilde, and Siegfried (1993) investigated the patterns of online searching by humanities scholars and found that to identify the relevant literature, humanities scholars usually conducted more search at the early stages of their research projects. Spink (1996) interviewed 200 academic library users and found that more than half (56%) of them had the experience of multi-session search using OPAC or CD-ROM databases. Sellen et al. (2002) conducted a survey study with 24 knowledge workers and found that 40% of their information gathering (defined as "collecting information around a particular topic") [p.231] tasks were not completed in a single sitting search but spread for days or even longer. Morris et al. (2008a) surveyed information workers in a large technology company. They found that 83% of the population said they had cross-session search experience when their information-seeking tasks lasted for multiple sessions, and the time gaps between sessions ranged from minutes to years. More recently, Agichtein et al. (2012) analyzed a large scale of search log data of more than 1000 users and identified a variety of intents, topics, and search behavior patterns associated with cross-session search. Researchers found that users' search activities, along with the high-speed development of technology, are very complicated in a natural context. For instance, Vuong, Saastamoinen, Jacucci, and Ruotsalo (2019) analyzed 24/7 continuous recordings of participants' computer screens and written diaries. They found that users frequently engage in multi-tasking search activities, meaning they search for different topics in parallel. Their search process often across a variety of sources and applications.

2.1.1 Task types in cross-session search

There are many different ways to categorize task types in information seeking, such as the three-level nested task classification (Byström & Järvelin, 1995), other classifications based on task topics or domains, task goals, outcome types, the level of difficulty, time requirements, etc. (see a comprehensive review in (Li & Belkin, 2008)). Previous researchers often describe tasks involving cross-session search as complex due to the complexity of the task requirements or that users may not be clear about their goals and are not familiar with the task domain (J. Liu & Belkin, 2010; Czerwinski et al., 2004; MacKay & Watters, 2008a; Kotov et al., 2011).

Sellen et al. (2002) identified six types of web task types conducted by knowledge workers (i.e., browsing, communicating, finding, housekeeping, information gathering, and transacting). Among the tasks, 40% of information-gathering tasks were spread across multiple days to weeks. And 57% of browsing activities were also conducted on a routine basis (e.g., daily or weekly). A critical reason causing information-gathering tasks to spread over time is that users are often not clear about the amount of information they need in advance. The task requirements are often less specific. Therefore, users need to look for information from various information sources, which often involve comparing, contrasting, and synthesizing the found information (Sellen et al., 2002). Meanwhile, browsing tasks often involve multiple search sessions. Similar to information-gathering tasks, users generally do not have a specific search goal. But unlike information-gathering cross-session search tasks, browsing tasks usually involve similar purposes. For instance, many routine tasks occur regularly, such as checking emails, keeping up with events, and tasks related to personal hobbies, such as reading news or watching sports games.

Although both information-gathering tasks and browsing tasks include cross-session search behaviors for similar information needs, they are different from each other. Information-gathering tasks are often driven by a specific project with a clear start and end, while many browsing tasks did not have a particular goal for leading the browsing activities (Sellen et al., 2002). This difference is also reflected by Sellen et al.'s (2002) survey in which users provided a higher score of importance for information-gathering tasks than for that for browsing tasks (6.8 vs. 5.4) (Table 2.1).

Table 2.1: Knowledge workers' web use activities by Sellen et al. (2002)

Category	Pct.	Mean duration(min)	Mean importance
Info gathering	35%	22.8	6.8
Finding	24%	7.5	6.9
Browsing	27%	7.9	5.4
Transacting	5%	5.5	7.3
Communicating	4%	12.3	5.8
Housekeeping	5%	40.4	7.6

Czerwinski et al. (2004) specifically investigated users' task switching and interruption in workplaces and provide some insights about tasks involving XSS. They categorized eight typical task types that comprised the majority of the participants (i.e., routine task (27%), email (23%), project (18%), task tracking (13%), telephone call (8%), meeting (6%), personal (5%), and downtime (0.3%)), and found that on average, each task would be interrupted once. Although not all tasks require people to return back after the interruption, they found that: tasks that require people to switch back/resume after interruption are often more difficult and complex, take longer time, and need more documents than those does not (Czerwinski et al., 2004). Participants spent a significant amount of time tracking the task's progress if they needed to resume it. Because these tasks often need more documents that cost longer time to find, they are more easily interrupted. Vice versa, the more frequently they are interrupted, the fewer users are productive, and the more sessions people need to work on the tasks to complete them. Similar to Sellen et al.'s (2002) findings, most of these tasks are project-driven and involve routine tasks as well.

MacKay and Watters (2008a) collected data from college students and classified 235 work tasks that involved cross-session search into eight different topic types: school work, general topic search, research, travel/tourism, projects, action-based, shopping, and status checking. Furthermore, they adopted the similar task categories developed by Sellen et al. (2002) and Kellar, Watters, and Shepherd (2006) for categorizing the type of sub-tasks (information search task) and found that: Among all the 428 sub-tasks (of the total 235 work tasks), information gathering tasks are the most frequent sub-task type (53.27%), followed by fact-finding (27.8%), transactions (10.75%), Communication (4.91%), and maintenance (3.27%).

Agichtein et al. (2012) analyzed millions of search log data generated by 1,000 users within one week and found that tasks that are more likely to be continued in a future search session involve topics associated with adults; kids, and teens; and news are more often. Information maintenance (e.g., monitoring for updating information or an in-process topic) is the most likely to be a continued task after breaks, followed by browsing tasks. Additionally, they also found tasks that are less goal-directed are more often resumed later, and so does tasks that are less time-sensitive. Because people might prefer to complete urgent tasks first. Since the analysis of the *tasks* is based on the search log data (timestamp, queries) without knowledge of users' context information, the task defined in the study is at the search task level rather than at the work task level.

Among these studies, MacKay and Watters first explicitly separates cross-session work tasks from information search tasks. They treated the search tasks as sub-tasks that comprise the work tasks. Their findings are in line with the previous research that information-gathering tasks take up a large portion of tasks related to cross-session search behavior. They also show that even for the same task (at the work task level), one cross-session work task can contain multiple types of sub-tasks (information search tasks) more than gathering information. Together with other studies listed above, we can conclude that: (1) overall, cross-session task types cover many different areas and topics; (2) the purpose/goal of the tasks and users' knowledge of the tasks may affect their cross-session search behavior; (3) the process of cross-session search can involve other task steps may (or may not) involve information seeking activities; (4) both self-initiated and external interruptions have an impact on the processes of work tasks and cause cross-session task activities.

In addition to the differences between methods for categorizing different types of tasks, we can also find that not all studies mentioned above agree upon the definition for *tasks* that involve cross-session search. In realization of the lack of definition, MacKay and Watters (2012) defined multisession tasks as "require more than one web session to complete, to have subtasks, and to have a definable endpoint when the task is over." [p.1183] This definition clearly distinguishes work tasks related to cross-session search from routine tasks that people conduct on a regular basis but are less goal-oriented. The definition helps us identify cross-session search activities related to one specific project from general browsing activities. Thus, we can narrow down the scope of cross-session work tasks to a more manageable scale.

2.1.2 Sessions in cross-session search

Another important element of cross-session search is *session*. However, the definitions of search sessions are still fuzzy and used differently depending on research purposes and scopes. For instance, some researchers focus on a single user's search behavior within a defined length of time period (Silverstein, Marais, Henzinger, & Moricz, 1999), others focus on auto-detecting/cut-off using algorithms (Jansen, Spink, Blakely, & Koshman, 2007), and some others suggest adapting users' perspective to set the boundaries between sessions (Ye, Wilson, & Rodden, 2014). For instance, Ye et al. (2014) found that users' standards of time gap for dividing sessions vary, and their tolerance for the time gap between two sessions is higher for tasks that involve more search queries and more page reviews. Thus users may combine two sessions that seemingly separate together. They also found that people are more tolerant of real-life interruptions (e.g., cooking, sleep) and often resume previous interrupted sessions. As a result, they pointed out that detecting sessions only based on query number and page views or time length may lead to misjudgment of search session length in reality.

Learning from these different perspectives, researchers of XSS studies take both the system-side feature of continually interacting with search systems and users' self-identify session into consideration without restricting the time length or the number of queries issued or retrieved documents for one session. For instance, Spink (1996) asked participants to report self-counted search sessions related to a specific task when they used library databases. Researchers also divide search sessions based on goals/sub-task goals that require participants to complete during multiple lab-based search sessions in experimental studies. For example, MacKay and Watters (2008a) required their participants to record work tasks and the start and end time for every associated search session using an add-on browser application. Both Lin and Belkin (2005) and J. Liu and Belkin (2010) designed complex tasks with multiple subtask goals that require participants to come to the lab three times to conduct a search over time. Cross-device search studies often divide different sessions based on control of time and devices used by participants (D. Wu, Liang, & Bi, 2018; Han, He, & Chi, 2017).

Previous researchers have also explored some aspects of XSS sessions. Based on the results from the diary study and field study. MacKay and Watters (2008a) found that college students work on four tasks involving cross-session search per week, and about six tasks per month. On average, they conducted about 2.5 sessions per task (Table 2.2).

In Table 2.2, we can find that the number of recorded XSS tasks per person and sessions per task for each person of *field study* are higher than those of *the diary study*. Different research methods might cause this. In the field study, the researchers installed search log applications that helped log all search information, and participants recorded their sessions based on the logged data. In contrast, in the diary study, participants independently logged their search information through a web diary form. They might not record everything manually so that the number would be less, compared to the automatic log data.

Table 2.2: Tasks and average session numbers by MacKay and Watters (2008a)

Study type	**Participant**	**Total tasks**	**Total visits**	**Task/person**	**Session No./task**
Diary study	22	85	204	3.86(per week)	2.40
Field study	24	150	391	6.25(per week)	2.61

Additionally, Morris et al. (2008a) surveyed 170 participants and found that the time gap between search sessions for one task ranged from minutes to months and years. The most frequent time gaps between sessions are hours and days, followed by weeks and then months (Table 2.3). The results are in line with an earlier study conducted by Obendorf, Weinreich, Herder, and Mayer (2007). Researchers found medium-term and long-term revisitation (revisiting between a day to weeks or longer) are very often in that study.

Table 2.3: Frequency percentage of time gaps between different sessions by Morris et al. (2008a)

Minutes	**Hours**	**Days**	**Weeks**	**Months**	**Years**
1.2%	24%	51%	13%	10%	1.2%

Similar trends in session length and frequency of search sessions are also reflected by experimental XSS study designs and projects involving multi-session search behaviors. For instance, Lin and Belkin (2005) designed a cross-session search study to validate the MISE mode. They required participants to come three times during two weeks to work on a traveling plan project, and each of their search sessions lasted between 20 to 30 minutes. J. Liu and Belkin (2010) adapted Lin's methods and designed a three-session cross-session search experiment and allowed their participants to search up to 40 minutes for each session. Both Capra (2006) and Jhaveri and Räihä (2005) designed lab studies that invited participants to come on two different days and search for 30 to up to 45 minutes (for multiple web pages and multiple tasks) to investigate their information re-finding and revisiting behavior. Kuhlthau (1991) and Vakkari (2001) conducted a series of longitudinal studies to study participants' search behavior in the context of working on semester-long research projects. They collected users' search data at three different time points when participants were required to conduct relevant search behaviors for the project they worked on.

2.1.3 Causes of cross-session search

General motivations for starting/stopping information search. Researchers have long sought to understand users' motivations and intentions for their information searching behavior. Wilson (1981) suggested that people perform information searching to fulfill their physiological, affective, and cognitive needs. Dervin (1983)'s sense-making model shows that users need to search for information to bridge the knowledge gap in order to make sense of the whole context. Ingwersen (2000) distinguished different types of information needs based on users' information search intents, for instance, "verificative needs" describes that searchers need to verify information objects with structured data, whereas in "ill-defined" or "muddled topic needs" situation, searchers need to explore new concepts and relations outside of their known subject area or domain. Based on analysis of web search queries, Broder (2002) classified users' web search intentions into three categories: navigational intent to reach specific websites that users have in mind, informational intent for gaining relevant information that searchers assumed could be found on the web, while transactional intent for finding websites for further interaction like shopping, finding various services, etc. Kellar, Watters, and Shepherd (2007) built a more detailed typology of users' web task types and intentions: fact-finding, information gathering, browsing, and transactions. By asking participants to identify their intentions through empirical studies, H. Xie (2002) classified users' information seeking intentions into seven categories (Figure: 2.1). They found that participants often reported multiple intentions for a single search query segment.

Categories	Information Seeking Intentions
Keep record	Keep record of a link (KR)
Identify search information	Identifying something to start (IS); Identify something more to search (IM)
Learn	Learn domain knowledge (LK); Learn database content (LD)
Find	Find known item(s) (FK); Find specific information (FS); Find items sharing a named feature (FN); Find items without predefined criteria (FW)
Access item(s)	Access a specific item (AS); Access items with common characteristics (AC); Access a website/homepage or similar (AW)
Evaluate	Evaluate correctness of an item (EC); Evaluate usefulness of an item (EU); Pick best items from all the useful ones (EB); Evaluate specificity of an item (ES); Evaluate duplication of an item (ED) (i.e., determine whether the information in one item is the same as in others)
Obtain	Obtain specific information to highlight or copy (OS); Obtain part of an item (OP); Obtain a whole item(s) (OW)

Figure 2.1: Information seeking intentions and the associated acronyms H. Xie (2002)

Meanwhile, the reasons that cause people to stop searching activities have also been investigated well in the IR field. Researchers identified different types of search stopping rules and strategies users apply to decide when to stop searching (Table 2.4).

Table 2.4: Examples of search stopping rules/strategies identified by previous studies

Stopping rules/strategies	**Meaning**	**References**
Frustration/satisfaction-point	Stop searching when users discard/found a certain number of useless/useful documents	Cooper (1968)
Fallacy stop	Stop searching when users reach a predetermined size of desired search results.	Bates (1984)
Enough information/satisficing	Stop searching when users feel satisfied with the amount of information they found.	March (1994) Zach (2005) Marchionini (1997)
Cognitive rules for stopping	Stop searching based on users' cognitive process of the information they collected	Nickles, Curley, and Benson (1995)

These early theoretical classifications only identified broad intention categories associated with a single search query or an entire search task. Some researchers noticed that users' web searching behaviors are dynamic, often discountinuous (Savolainen, 2012). From this perspective, the nuances changes of users' search intentions and reasons that cause them to stop have been ignored by these categories, especially when their search tasks spread over time.

Motivations for cross-session search. Users' motivations for cross-session searches could be more complicated to understand. Early researchers noticed that people searched multiple times over an extended period for exploring a topic, evolving information problems at hand, etc. (Saracevic, 1991; Robertson & Hancock-Beaulieu, 1992). Much of previous research on the reasons that lead people to search for multiple sessions can be classified into two groups: (1) studies states that people search across sessions due to the complexity of the task, or difficulties to find all needed information at once. Most of these descriptions are based on assumptions, retrospective reviews,s or analyses of users' XSS activities or search log data (e.g., Sellen et al. (2002), Morris et al. (2008a), Agichtein et al. (2012)). Following this trend, it has been commonly accepted by cross-session/cross-device experimental study researchers to design XSS tasks that require people to look for multiple information pieces or tasks that include multiple steps to complete (e.g., J. Liu and Belkin (2010), D. Wu (2018)). (2) studies focus on motivations about why people need to conduct successive search sessions after their initial search session (e.g., Spink, Griesdorf, and Bateman (1999), MacKay and Watters (2008a), Li, Capra, and Zhang (2020a)). In this second group, Spink and her colleagues' work could be an example. They conducted a series of studies to learn how users search multiple times through the web (e.g., EXCITE, Vivisimo) (Spink, Bateman, & Jansen, 1999; Jansen, Koshman, & Spink, 2005). Through experimental mediated successive search studies on the Dialog information service system, Spink and her colleagues (Spink, Bateman, & Jansen, 1999; Spink, Wilson, Ford, Foster, & Ellis, 2002) found that the main purpose of successive search (a later search for continuing a previous incomplete search) is to extend, expand, or refine the search results of previous searchers, based on the searchers' evaluation of the results of the previous search or due to changes in their information problem (Figure 2.2).

Reason for successive searching	No. of searches	% of searches
Refine and enhance the search, using results from the previous search, e.g. new terms	7	25.9%
Information seeker requested more information	6	22.2%
Search different databases	4	14.8%
Refine the search, as too much data was retrieved in the previous search	3	11.2%
Refine the search, because of the increased complexity of the information problem since the previous search	2	7.4%
Refine the search to print abstracts	1	3.7%
Intermediary suggested another search	1	3.7%
Lost data from previous search	1	3.7%
Secure more valuable information	1	3.7%
First search was only exploratory	1	3.7%
Total	27	100.0%

Figure 2.2: Reasons for mediated successive search by Spink, Bateman, and Jansen (1999)

Unlike the others mentioned above, Lin and Belkin (Lin & Belkin, 2000, 2005) proposed a theoretical model of Multiple Information Seeking Episodes (MISE), which includes eight different reasons why people resume searches for the same information problem (Figure 2.3):

> "(1) transmuting – the problem gets elaborated and changes from its original form to a transmuted form; (2) spawning – the problem spawns sub-problems; (3) transiting – the original problem transits to another, different problem; (4) rolling back – something that was thought to have been solved by a previous search turns out to be unresolved; (5) lost-treatment – the information... once found, is not available in the treatment application stage; (6) unanswered – the problem was unanswered by previous searches; (7) cultivated – occurs when a searcher is trying to stay abreast of an area of interest; (8) anticipated – the information problem has not occurred yet, but is anticipated based on the current information" (Lin & Belkin, 2005, p.396).

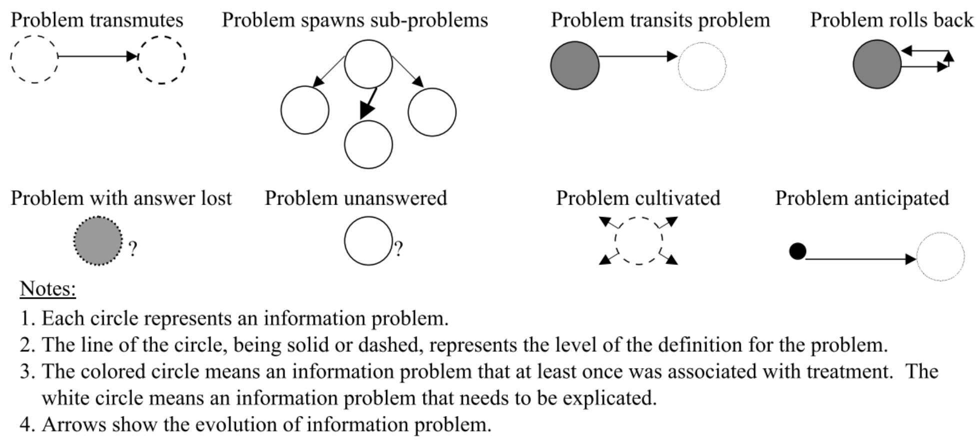


Figure 2.3: Modes of successive information seeking episodes adapted from Lin and Belkin (2005)

The MISE reasons are based on the conceptualized problem stages of the underlying work task that motivate participants to search. Lin and Belkin conducted a lab experimental study with simulated work tasks and proved the existence of these reasons (Lin & Belkin, 2005), our recent survey study with Amazon Mechanical Turk workers (Li et al., 2020a) validates that these reasons can also be found in real work and life cross-session searches.

Meanwhile, the MISE model (Lin & Belkin, 2005) also distinguished two types of reasons that cause people to stop during a multisession search: External reasons are imposed by the environment, such as interruptions. Internal reasons are self-motivated and relate to users' information needs and behavior, such as lack of domain knowledge, the need to stop and process (e.g., synthesize, validate, evaluate) the found information, or to consult other sources. MacKay and Watters (2008b) observed that college students usually stopped their search sessions because they found what they needed or finished their task; they also stopped if they were tired, or needed to do something else. Unfortunately, to the best of our knowledge, relatively little research has been conducted to investigate users' motivations for starting/resuming a specific search session and why they stop for each session, how these reasons interact with each other, and more importantly, how these reasons can affect their search behaviors across the different sessions during XSS.

2.1.4 Cross-session information search activity

Information search behavior has been studied intensively in recent years. Researchers investigated information search behavior under various contexts. The availability of large-scale search data from search engines has proven to be extremely valuable in studying how people search in naturalistic settings with a wide variety of search intents. Other study methods (e.g., survey, interview, field study, experimental study) also provide rich details related to users' interactions with different types of information systems and applications. In this section, I will review previous studies' results about people's information search activities in the XSS process.

Information search/re-search. Information search activities are about users' interaction with the search systems. Similar to studies of single session search activity, query terms, URLs, clicked documents, and time spent on a specific web page are important elements for measuring users' information search activities during XSS. Previous XSS studies often obtain these data from search engine companies' databases(e.g., Agichtein et al. (2012)), or by using customized software like browser extensions, or add-on software installed into users' personal computers or devices (e.g., MacKay and Watters (2009)), or require their participants to use specific websites or browsers (e.g., D. Wu et al. (2018), Zhang, Capra, and Li (2020)). When analyzing the search log data without users' input of contextual information, researchers often use timestamps to distinguish different sessions (e.g., start from issuing a query to the end of 30 minutes inactive). URLs and queries are often used as indicators for underlying information needs. Similar queries or the same query often indicate related or same information needs, so do clicks for the same URLs or documents, which are often used as complements to queries (e.g., Agichtein et al. (2012), D. Wu (2018), Wang, Huang, and White (2013)).

By investigating queries (or query chains) from different sessions, researchers could identify whether the two sessions are associated with the same information needs and predict whether a searcher will continue a task later. For instance, Teevan, Adar, Jones, and Potts (2007) analyzed users' query logs and found that nearly 40% of queries were attempted to re-find previously encountered results. R. Jones and Klinkner (2008) proposed methods to partition a query stream to identify research missions (work tasks) and goals (sub-tasks) that spread across days. Each mission corresponds to a set of related information needs and may include multiple search goals. They suggested that using this hierarchical model for the grouping of user queries can help predict the type of users' task and measure the length of time or number of queries needed to complete a task. Kotov et al. (2011) and Agichtein et al. (2012) labeled millions of queries from thousands of users and developed algorithms that can detect task continuation effectively. J. Liu and Belkin(2010) examined whether different types of time (dwell time, display time, and decision time) a user spends on a document can predict the usefulness of the document and found that the *total dwell time* is a reliable predictor for document usefulness for cross-session tasks. The longer the time a participant spent on a document across multiple search sessions, the more useful it is. Probably because participants need to come back and forth to the same document as they found it useful. The similarity analysis of queries between sessions is often used as a feature to developing computational models to predict cross-session or cross-device searches, which is out of the scope of this review (e.g., Wang et al. (2013), Kotov et al.; D. Wu (2011)).

As many XSS studies address the importance of the similarity between search queries across sessions, the differences among the queries during XSS have less been discussed. A few studies have noticed the differences among queries. For instance, D. Wu (2018) noticed that besides using the same queries to resume search across devices (and sessions), users often resume the task by refining their queries, which may be because of the impact of the previous search experience on last device (in the previous session). Zhang et al. (2020) analyzed users' search intentions when they work on long-term creative tasks and found that users often modify their search queries during different sessions in order to switch their search directions at different creative stages.

Due to the complexity of cross-session tasks, users' information needs may vary at a higher level that is not directly reflected in single queries or query pairs (Kotov et al., 2011). Queries corresponding to the same task may not have any terms in common. Based on an analysis of users' search log, Agichtein et al. (2012) found that information-gathering tasks are less likely to be continued because they are "fairly simple" [p.4]. Their result contrasts with other study findings based on interviews and field studies of cross-session information search tasks (e.g., Kellar et al. (2007), Sellen et al. (2002)). This contrast may indicate that the investigation of XSS search behavior should take more contextual factors into account.

Cross-device search. Along with the development of high technology and a speedy internet connection, users now can use many different types of devices for searching for information at any time and in many different places. Users information interactions are no longer limited to desktop computers. Search across devices become common in our daily life. The common device shifting modes are shifting between PC (desktop/laptop) and mobile devices (e.g., smartphone, tablet) (Wang et al., 2013; Montanez, White, & Huang, 2014). On the one hand, not all cross-device searches are cross-session searches. A previous study of device switching search found that about 15% of them involve task continuation while others are not (Wang et al., 2013). Using different devices at different times and location is ubiquitous. The choice of devices for searching depends on many aspects, such as accessibility, restrictions (e.g., work computers mainly for work tasks, personal computers for private tasks), or other user habits. On the other hand, cross-session search involving using multiple devices usually is categorized into the scope of cross-device search. Many experimental studies of cross-device search often design their tasks as cross-session searches for the same task. As a result, studies focused on cross-device search can provide a unique perspective for research about cross-session search.

Users' search behavior changes as they shift between different devices, including search topics, query patterns, search strategies, etc. Church, Smyth, Bradley, and Cotter (2008) studied early day's mobile users' information search behavior and compared their results with the results of traditional web search. They analyzed 6 million search queries from 260, 000 users over a 7-day period in 2006. They found that mobile searchers tend to use similar strategies on mobile devices as they search on the web: They tend to use short queries and mainly focused on the top-ranked results. On the other hand, mobile searchers searched through more result pages to locate their target results, but people modify their search queries less often. The adult-related topic was the mainstream search topic for mobile search queries. While the search for social media content, and navigational queries also took a high proportion overall.

Through a survey study of 929 mobile searchers, Teevan, Karlson, Amini, Brush, and Krumm (2011) found that their mobile search experience varied a lot compared to their desktop experiences: Mobile search is primarily affected by geographic and temporal, and social aspects, which shaped the context of mobile search behavior. For example, their participants reported that they often looked for location information near their current location or when they were in transit. Sixty-three (63%) percent searchers reported that their most recent mobile search happened when they were communicating with somebody else or collaborating with others to search something together, which indicates that a large portion of mobile search has social meaning (Teevan et al., 2011).

In another study, Song, Ma, Wang, and Wang (2013) analyzed over 1 million users' three-month search engine logs from both desktop (including tablet) and mobile devices. Their results show that: 1) The query length on phone and table were longer than those on desktop; 2) Tablet users were more likely to issue adult, celebrity, and image queries; 3) Desktop users performed search mostly during working hours while mobile and tablet usage peaked during evenings; 4) Mobile and tablet users tended to click more on knowledge-base sites like Wikipedia. They also found that transferring information from search activities on desktop computers could improve users' search performance on continued mobile devices such as tablets and smartphones.

These studies show that users' search behavior patterns change when searching across devices. Their search strategies and patterns can be affected by internal factors (e.g., personal habits, lifestyle, daily activities) and external factors (e.g., locations, time, social context). Additionally, these patterns are evolving with the development of technology. Thus, it is essential to take these features into consideration as we investigate cross-session search behavior with multiple devices.

Researchers consider cross-device search for the same task/topic as a specific type of cross-session search. Since it describes the shift of search behavior from one device to another device (usually to continue a suspended search at an earlier time) on the same task goal (Wang et al., 2013; D. Wu et al., 2018; Han, 2018). Many studies in this area focus on understanding users' search behavior across multiple devices to build models to predict users' device-switching behaviors. They explored the popularity of device switching patterns, topic trends, delay times between the search on different devices, and other contextual factors (location, time of day, etc.).

Wang et al. (2013) investigated users' search behavior for *contiguous* cross-device search tasks, for which the time limit between the last query before switching on one device and the resumption of the search on another device is less or equal to six hours. Similar to general cross-device search behavior investigated by previous studies, they found that task topic distribution and location changes are often associated with device change. For instance, image and navigational tasks are more likely to be resumed on a different device. They also found that the direction of device switching (from desktop to mobile phone, and from mobile phone to desktop) shows different characteristics which may be caused by the underlying reason: It usually took less time for users to switch from mobile to desktop compared to the time when they switch from desktop to mobile. This probably is because searchers were not satisfied with the results found on mobile phones and re-conducted the search on a computer. On the contrary, people switch from desktops to mobile phones often because they were interrupted, or they stopped searching to leave the workplace/desktop computer, etc.

Montanez et al. (2014) studied users' cross-device search based on more than 33 million unique users' search log data involving four types of devices (i.e., PC, smartphone, tablet, and gaming console). They found that cross-device users are often active searchers who contribute a significant amount of search volume of the overall search records compared to other non-active searchers. Furthermore, they found that most transitions are *self-transitions*, meaning that users are likely to continue using the same device after transit to a later search session. Self-transition also took less time than transiting between different devices. Meanwhile, most transition mass leads to PC devices, indicating the PC device still plays a central role in information search.

D. Wu et al. (2018) investigated six months of users' cross-device search behavioral data logged when they were interacting with the OPAC (Online Public Access Catalog) system. They identified unique users' search log data by matching the log activities of the same USER ID. They found that most users searched across sessions using the same device (mostly PC to PC), especially when their tasks did not change (the two adjacent queries are the same. This is in line with Montanez et al.'s (2014) finding of *self-transition*. D. Wu et al. (2018) also found that the self-transition strongly indicated that the users were searching for information for the same task. On the contrary, the cross-task pattern is more significant when people search using different devices. The cross-device search pattern is also associated with topics of varying domains. For instance, searches on industry and the technology-related topic is more often found with device transition. Additionally, their study also identified many other activities users conducted as they interacted with the OPAC system with different types of devices, such as browsing, clicking, managing personal accounts, and using other services provided by the OPAC system (D. Wu et al., 2018). These findings have insights into users' cross-session search behavior that to what extent users would like to switch devices as they work on the same task, and for what purposes they switch the device as they are working on the same task. Understanding the reasons leading users to switch devices within/between sessions can help us better understand their cross-session search behavior.

Many cross-device search studies focus on predicting task continuation and the types of next devices used for the post-switching session. Some factors that may cause device switching have been discussed (e.g., interruptions, need to leave work), but few address the impact of task features or users' search progress. Another limitation of cross-device search research is that their investigation is built upon large-scale search log data analysis. Like other search-log studies, it is hard to tell whether users' context for the queries they used. Because clicking on the same results, using similar queries may or may not indicate the same work task or project. Vice versa, queries without overlaps may be still for the same task but from different aspects.

Studying cross-session/device under a more controlled environment has started to attract more and more attention. To fill the gap, many researchers conducted a series of experimental studies to examine cross-device search behavior. Wu and her colleagues distinguished the sessions before and after the transition between devices as *pre-switch* and *post-switch* sessions (D. Wu & Yuan, 2018). Furthermore, they defined users' search activities during the pre-switch session as *task preparation*, and the search activities for the same topic/task during the post-switch session as *task resumption* (D. Wu, Dong, Tang, & Capra, 2020). Via controlled laboratory studies. Wu and her colleagues investigated users' search performance differences between the pre-switch and post-switch sessions and their search behavior features during the task preparation and resumption stages. They found that users' information need remains blurred in the task preparation/pre-switch session and becomes more focused in the task resumption/post-switch session. However, users' search performance decreased in the post-switch/task resumption session, which might be caused by the difficulty of gaining new information in that session (D. Wu et al., 2020). They also adopted different features to model user search behavior for different sessions and found that query features, click on search results, time spent on the document and users' perceptions (satisfaction, familiarity) play an important role in distinguishing different sessions.

Han and his colleagues conducted a series of experimental cross-device studies (specifically, cross-device search between desktop computers and mobile phones) (Han, Yue, & He, 2015; Han et al., 2017; Han, 2018). Their studies found that users' information behavior patterns for exploration and sense-making are different between switching from mobile-to-desktop search and the switch for desktop-to-desktop. Specifically, they found that users performed more sense-making actions (e.g., reviewing the current search status, and figuring out further search directions) during the first sessions when searching on a computer desktop (Han et al., 2015). Users issued more search queries, visited, and saved more pages as they search across multiple sessions with desktop computers compared to the search after switching to mobile devices. On the other hand, the dwell time (e.g., inactive time) users spent on the visited page on a mobile device is significantly longer than that they spent on a desktop computer (Han et al., 2015, 2017). These findings indicate that devices impact users' information search behavior substantially in cross-device & session information search. Their findings also demonstrate that mobile touch interactions (MIT) (e.g., tap, drag, pitch-in, pitch-out) with the Search Engine Results Pages (SERPs) on a touch-based device is useful. Because MIT can help identify relevant text chunks at the subdocument level to help achieve better search performance and facilitate search system design for supporting cross-device search (Han et al., 2017).

Summary Both the search log data analysis and experimental studies provide rich details regarding users' cross-device and cross-session search behavior. The researchers reveal the differences and connections between the search features and users' performance. They proposed strategies and algorithms for modeling and predicting users' successive search behavior over time. These findings are beneficial for gaining an overview of cross-session search behavior. On the other hand, there are some limitations of the current studies, making them less appropriate to describe cross-session search. As we said earlier, most of the search log data cover a significant length of time (ranging from weeks to months). Their focuses were to detect the relations between search logs to predict task continuation. Although queries can reflect users' information needs, it is impossible to know the exact context when users issued a specific query. Even we can infer users' underlying information needs, the inference of scattered queries can hardly cover all information needs that are related to a specific work task. Thus, the strength of the inferred connections among queries or search histories is significantly weakened. Second, many of the cross-device search focus on only two sessions: the pre-switch session and the post-switch session. In real-life cases, users may search and switch more than two times to complete one task. Their search behaviors and other task-related activities could interleave intricately but in an accumulated manner. From this perspective, we need to explore further how multiple ($>= 3$) sessions transition from one to another and finally integrate to achieve their task goal. We are essentially curious about the changes in the underlying information needs that trigger the starts and stops of each search session, and the transitions among these sessions related to a specific work task.

2.1.5 Challenges and support in cross-session search

Searching across sessions and devices can cause many different challenges for users. A frequently discussed issue in XSS is about keeping information across sessions (and devices).

Information keeping plays an important role in XSS: it can help searchers remember what they have found, save useful information for later use, and monitor their task progress. While their task and search activities spread over time, people need to keep many different types of information through the process, including their task progress, the information they found online, and other non-search information activities, such as processing the found information, sharing with others, keeping information for later consulting, etc. For instance, during a survey of people's long-term projects (ranging from 2 weeks to 2 years), W. Jones, Bruce, Foxley, and Munat (2006) found that participants all reported that they experienced problems with information fragmentation across their projects and that maintaining (and maintaining consistency between) different organizations was a major problem. Lin and Belkin (2005) found that under a controlled environment when participants search for multiple sessions using an experimental search system, users encounter difficulties regarding articulating information problems, navigating to appropriate information sources, and monitoring the status of their achievement.

To solve some of these problems, researchers found that users often leverage the functions provided by browsers to re-find information they saw earlier, such as the Back button to support in-session revisitation, Bookmarks, History tool, etc. (Jhaveri & Räihä, 2005). Besides, by analyzing survey responses about strategies people used for resuming a suspended search, Morris et al. (2008a) also identified other methods people use for keeping information, including written or typed notes, saving pages in local drive, emailing to themselves, printing out, leaving the browser open, using online query history or autocomplete, or just based on their memories. Many of these are still the main approaches for today's users to keep the information during XSS, although the technology has significantly developed during the past two decades.

In the recognition of cross-session search tasks' complexity and their unique requirements that lead users to conduct cross-session search, researchers and developers have devoted lots of effort to develop novel tools to assist users' cross-session search process. Many of these tools try to reduce users' workload and overcome the inconveniences people encounter during the cross-session search.

Some of the existing features of current commercial search engines have been proven useful for assisting cross-session search. For instance, previous research found that bookmarks, search history, reading lists, etc. are often used by searchers to keep what they saw and go back to the information they visited during previous sessions (Morris et al., 2008a). Some researchers developed novel prototypes for helping users better organize the visited pages, annotations, and their search paths, though not all of them were initially intended to help search across sessions.

SearchPad– Bharat (2000) developed *SearchPad* to help users keep track of the search context, especially when they search using multiple search engines or across multiple search sessions. The *SearchPad* can log the context (e.g., the actions taken by the users, the order of visited pages, and the time spent on a landing page) of each query and record the links that were considered relevant/visited by the user in the context of the query. Their findings of real users' experience with the *SearchPad* found that users looked at more search results than they typically did without using the tool, which indicates that providing the search context may have a positive impact on exploring search results more deeply and compensating for some non-relevant pages in the ranking.

Session Highlights– Jhaveri (2004) developed *Session Highlights* to help users keep their search histories (for complex tasks that spread over time) by saving thumbnail images of the web page in a *snake layout* (images are connected by a thin line in a zig-zag way indicating the chronological relationship. When users are working on their tasks, they can review and compare the content of the collected pages, save, edit, remove the collections or share with others, and save the visited web pages as a session summary so that they can pick it up at a later time. They can also delete to the collection after the task is complete or move web pages they favor of to long-term storage (e.g., bookmarks).

SearchBar–Based on a survey of users' cross-session search, specifically, task resumption and their information re-finding strategies, Morris et al. (2008a) developed *SearchBar* to improve the user experience of resuming suspended investigations and re-finding previously encountered online content. The *SearchBar* can capture all queries issued and URLs visited to each query, store the browsing history (in reverse-chronological order), and allow users to organize this information based on topics. Users can also annotate the saved web pages and use the thumb-up function to indicate the relevance of a topic. Compare to the earlier prototypes, *SearchBar* provides users with more control over the stored information and allows users to keep track of the context (e.g., create topics, add annotations) of previously visited information. Participants can choose to use (or *not*) these extra functions depending on how they want to interact with the tool.

Search Pad– Donato, Bonchi, Chi, and Maarek (2010a) at Yahoo! developed and launched a novel application, *Search Pad* (Donato et al., 2010a) intended to help users take notes as they work on complex *research missions* (e.g., tasks that span over time and require users to gather information from multiple web pages). Unlike other proposed tools, *Search Pad* runs based on algorithms that can automatically monitor users' search behavior (lists of queries) and detect research missions. Once a research mission is detected, the tool will be triggered. Users can arrange and take notes as they visit a web page, and their notes will be automatically saved to the detected mission.

Querium– Golovchinsky, Diriye, and Dunnigan (2012) designed a search interface called Querium to help exploratory searches that often require people to retrieve information from multiple sources, collaborate with others, and evolve over time. Focusing on helping users express their information needs and assess their search results, *Querium* allows users to use three different types of queries: standard keyword queries, relevance feedback (like function), and document fusion that can combine the results of all queries for the task and order the documents based on their metadata. *Querium* can also record the query history, present retrieval histograms, allow faceted filters, and a summary view of process metadata (e.g., users' interaction history with the retrieved documents).

Task management tool– Based on their diary and field study of college students' cross-session search, MacKay and Watters (2009) developed three guidelines for designing tools to support cross-session search: (1) the tools should keep a list of the multi-session tasks that users are working on; (2) the tools should be able to remind users to keep on track on the tasks they work on, especially when they are multi-tasking; (3) the tools should allow users to manage their cross-session tasks between the search sessions. Following these guidelines, the authors developed three prototype toolbars (from simplest to more complex ones by adding extra features) built into the Firefox browser. Using these toolbars, users can: create tasks that they want to work on, click on "start" or "end" to complete a session for a task, return to a task by the "resume task" tab, view their browser interaction logs during a session, manage the URLs, add task descriptions, and so on. Backhausen et al. (2015) developed a prototype called *TasksTo* that allows users to set up a number of goals at different levels, keep track of their search activities (browsing history, queries, etc.) and adapted task recommendations according to users' bookmarks and their recent search queries.

In addition, many studies have investigated modeling cross-session or cross-device searches in order for predicting users' task continuation behavior, which is out of the scope of this study since few of them have proposed extra support for cross-session or cross device searches regarding users' task activities. Recent development in technology support users' cross-session, cross-device search by providing the history of search queries across sessions and devices when users use the same account when searching for information. But such search systems so far cannot distinguish queries that belong to different tasks.

2.2 Tasks in information retrieval

2.2.1 Tasks and task facets

Tasks exist almost everywhere in our daily work and lives. Many tasks require people to conduct various activities under different circumstances. Different tasks take different amounts of time and energy to complete. Tasks play a central role in information seeking and retrieval fields. Tasks motivate people to conduct information activities to fulfill their goals and can influence searchers' information behaviors (Byström & Hansen, 2005; Kelly, Arguello, Edwards, & Wu, 2015). Tasks are not only associated with users' information seeking strategies, but they are also the foundation for users assessing the information they found through the process (Berryman, 2006; Järvelin et al., 2015).

In Interactive Information Retrieval (IIR) areas, tasks are often considered at three different, but nested levels: work tasks, information-seeking tasks, and search tasks (Byström & Järvelin, 1995; Vakkari, 2003; Ingwersen & Järvelin, 2005). Researchers have identified many different aspects of tasks. For instance, based on users' information requests, tasks can be classified as "known item search tasks" and "exploratory search tasks" (K.-S. Kim, 2001; Marchionini, 2006). I. Xie (2009) classified the dimensions of work tasks into three categories: (1) the nature of the task, (2) the stages of the task, and (3) the time frame of the tasks. Li and Belkin (2008) provides a systematical review of a variety of task classification schemes and proposes a faceted classification of task that include: source of tasks, task doer, time, product, process, goal (as a generic facet of task), task characteristics (i.e., objective task complexity, interdependence), and users' perception (i.e., the salience of a task, urgency, difficulty, subjective task complexity, knowledge of task topic). This faceted classification scheme also provides a convenient way to analyze tasks at different levels and to investigate the relationships between them along one or several facets.

It has also been widely recognized that task facets have significant impacts on users' information search behavior from different perspectives. For example, much of research in task-based information interaction (TBII) (Järvelin et al., 2015) found that task types can affect users' information seeking behavior on the web such as the types of information people look for, different databases they consult, and various search strategies, etc. For instance, search task type is a strong indicator of the information types people need. Users often search for specific precise data for factual tasks and look for more general information for exploratory tasks (J. Kim, 2006a). The sources of the task (i.e., self-initiated/internally generated v.s. externally assigned tasks) can affect users' motivations and to what extent they may feel satisfied with the found information and when they decide to stop (Trenck, Neben, & Heinzl, 2014; Reid, 2000). Marchionini (1997) distinguished different types of information search behaviors between look-up information tasks and exploratory tasks because exploratory tasks need participants to explore information spaces they are unfamiliar with and need to learn new knowledge as they search for information on the web. Thanks to the rapid development of internet service and intelligent devices (e.g., laptop computers, smartphones, tablets, and Intelligent Assistant devices), people can conduct search activities almost anywhere at any time, many more recent studies found that locations, surroundings, time (i.e., at work or after work) play important roles as users interact with information and information systems (e.g., "Exploring social context in mobile information behavior" (2014), Chen and Qi (2010), Bouidghaghen, Tamine, and Boughanem (2011), Vuong et al. (2019)).

In addition, many experimental studies investigated the influence of different task factors on users' information search behavior under controlled lab environments. For instance, C. Liu and Song (2018) found that different task types affect users' preferences of information sources, and future affect their learning outcomes. Capra, Arguello, O'Brien, Li, and Choi (2018) found that tasks with different levels of priori determinability (manipulated) can affect users' perception of task difficulty and their search strategies. J. Liu and Belkin (2010) found that different task structures can also affect users' search behaviors under cross-session search conditions. Aula, Khan, and Guan (2010) found that as users tend to formulate more diverse queries, use advanced operators, and spend more time on SERPs as they search for more difficult tasks compared to the easier ones.

Given that there are many other factors that can affect users' information search behavior and we cannot review all of them here, below we will particularly focus on two important task aspects which would also be the focused features of XSS tasks in the dissertation study: task complexity, task difficulty. Among the various task facets and attributes, task complexity and difficulty are two popular constructs that are often used by IIR researchers to describe the effects of tasks on users' information search behavior and their perceptions of the tasks as well as their information search processes, especially for work studying task effects in controlled experimental environments. Previous research often describes XSS tasks as *complex* tasks, like information gathering tasks or long-term tasks (Morris et al., 2008a; J. Liu & Belkin, 2010). Complex search tasks are defined as tasks that are multi-aspect or need multistep information that consists of a set of related subtasks, each of which might recursively be complex (Mehrotra & Yilmaz, 2017). However, to the best of our knowledge, little research has systematically investigated the complexity of real-life XSS tasks. Although few researchers did explore the complexity feature of XSS task features such as task goals and task structures (based on designed, experimental study), many of those are based on search logs without contextual information from user ends or simulated work tasks in lab studies (e.g., Agichtein et al. (2012), J. Liu and Belkin (2010)). Few of them use similar complexity/difficulty dimensions to study XSS tasks at finer levels than they do for single-session search tasks.

Task difficulty describes participants' perceptions of the amount of effort required to carry out the task (Campbell, 1988). Task difficulty is closely related to task complexity whereas task complexity is an inherent property of the task, independent of the task doer (Arguello, 2014; Li & Belkin, 2008), task difficulty is more subjective and related to "the interactions between the task attributes and the attributes of the searcher and the searcher's situation" (Wildemuth, Freund, & Toms, 2014, p.1129). In other words, task difficulty can be affected by task complexity (Arguello, 2014; Bell & Ruthven, 2004); meanwhile, for tasks with similar levels of complexity, users from different backgrounds with different expectations and assessment of the task difficulty, consequentially, their search behavior might be different. Thus, to investigate XSS task features and users' efforts to conduct XSS and complete their tasks, both task complexity and task difficulty are important for our study.

2.2.2 Task complexity

Complexity is an important task characteristic that can affect users' information search behavior and their task performance (Byström & Järvelin, 1995). Previous work has studied task complexity from many different perspectives. However, since task complexity cannot be directly observed, researchers have leveraged a variety of measurable elements to describe it (Wildemuth et al., 2014).

Campbell (1988) defined task complexity as "any objective task characteristic that implies an increase in information load, information diversity, or rate of information change can be considered a contributor to complexity" [p.43]. Based on this objective perspective definition, the author further identified four criteria of a *complex* task: (1) the task should have multiple potential paths allowing people to achieve the desired goals; (2) the task should have multiple desired outcomes; (3) the paths to different outcomes are conflicting and interdependent; (4) the connections between potential paths and the desired outcomes are uncertain. Any increase in either one of these four features can increase the complexity of a task (Campbell, 1988).

Byström and Järvelin (1995) perceive that task complexity is determined by the degree of the uncertainty, also called *a priori determinability*, of task inputs, process, and outcomes, the higher level the uncertainty is, the more complex a task will be. Accordingly, they classify tasks into five complexity levels, including (1) *automatic information processing task* as the easiest one because it is completely determinable and can be automated, (2) *normal information-processing tasks*, (3) *Normal decision tasks*, (4) *known, genuine decision tasks*, to *genuine decision task* as the most complicated one because it is completely unstructured in terms of task requirements, forms of solutions, and processes involved (Byström & Järvelin, 1995).

Byström and Järvelin (1995) also investigated different information types needed for tasks with different complexity levels. They identified three types of information in the expert system: (1) *Problem information* for helping understand tasks, that could be found in the problem environment or documents that can help people understand the task structure, properties, and requirements; (2) *Domain information* related to the problem domain like facts, concepts, etc.; and (3) *Problem-solving information* referring to methods, and strategies on how to approach the problem. They found that complex tasks often require all three types of information while simple tasks usually only require problem information.

P. Liu and Li (2012) systematically examined previous literature on task complexity and proposed a general task complexity model that consists of six components as the resources of task complexity (i.e., goal, input, process, output, time, and presentation). They further defined ten dimensions for measuring the complexity induced by each component (i.e., size, variety, ambiguity, relationship, variability, unreliability, novelty, incongruity, action complexity, and temporal demand) (P. Liu & Li, 2012).

Task complexity has also been studied from the perspective of *cognitive complexity*, which is a concept used for describing learning objectives in educational areas. Researchers in educational area Anderson and Krathwohl developed a 2D taxonomy of learning, including (1) cognitive process dimension (i.e., remembering, understanding, applying, analyzing, evaluating, and creating); (2) knowledge dimension (i.e., factual, conceptual, procedural, and metacognitive) (Anderson & Krathwohl, 2001).

Jansen, Booth, and Smith (2009) and later Kelly et al. (2015) adapted the cognitive complexity taxonomy from Anderson and Krathwohl (2001)'s learning objectives for students (i.e., remember, understand, apply, analyze, evaluate, and create, from lower-level to higher-level of complexity) to investigate the effects of task complexity on users' search behaviors. They found that tasks with a higher level of cognitive complexity could trigger more search activities (Kelly et al., 2015; Jansen et al., 2009).

Recently, Urgo, Arguello, and Capra (2020) used both the *cognitive complexity* and *knowledge type* dimensions of Anderson and Krathwohl (2001)'s taxonomy to design search tasks and found that knowledge types had significant effects on participants' perceptions of task difficulty, their performance, and search behaviors. Choi, Capra, and Arguello (2019) classified four types of information from objective to more subjective (i.e., facts, concepts, opinions, and insights) and found that the types of information their participants used are also associated with tasks' cognitive complexity: More cognitively complex tasks require a greater variety of information types and more subjective information than less complex tasks.

While these prior research findings provide insights into the effects of task complexity, less attention has been given to untangling the XSS task complexity, especially for the real tasks involved in people's daily work and life. A few studies tried to describe the complexity of tasks that involve cross-session search. For instance, Agichtein et al. (2012) hypothesized that complex tasks that lead people to search across sessions may contain multiple goals. However, their analysis based on the human annotation of search logs did not show valid evidence to support their hypothesis. Their results did not find relationships between task complexity (assessed by goal numbers) and task continuation. A potential reason for the failure of using task complexity to predict cross-session search could be because of the lack of contextual information from users' input. In other words, human annotations of search log data did not reflect the real task conditions of those search queries. Other experimental studies designed simulated work tasks that contain multiple steps, or multiple sub-tasks to mimic the realistic work scenarios of cross-session search and made conclusions about the effects of task process, task structure, and sub-tasks on users' search behaviors across sessions (e.g., Lin and Belkin (2005), J. Liu and Belkin (2010), D. Wu (2018)).

But some shortcomings of these experimental studies are that the simulated work tasks were designed to be *complex* with multiple steps, subtasks, and participants required to search for multiple sessions. And these studies did not focus on the effects of the task complexity levels on users' search behavior over time. Nevertheless, there is still a need to investigate the complexity of real-life XSS tasks and measure XSS task complexity. More importantly, unlike single-session search tasks, XSS work task often contains multiple sub-tasks across multiple search sessions. It is unclear how the entire task complexity relates to the complexity levels of sub-tasks or a search session in the process.

2.2.3 Task difficulty

Task difficulty is another important task factor in IR studies and is closely related to task complexity. Similar to task complexity, task difficulty can be affected by task types (including task complexity) and task doers (J. Kim, 2006b). On the other hand, unlike task complexity, *task difficulty* is about users' subjective perceptions about the amount of effort to carry out a task (Campbell, 1988; Arguello, 2014). As a subjective construct (Wildemuth et al., 2014), task difficulty can be evaluated based on users' expected efforts before the search as *pre-task difficulty* (i.e., how difficult do you think this search will be), or by their assessment of the effort they spend after the search called *post-task difficulty* (i.e., how difficult was it to complete this search?) (J. Kim, 2006b). Task difficulty has been frequently used to measure the effects of search tasks with different levels of *complexity* on a searcher's search activities under a controlled environment.

Based on the review of previous research, Wildemuth et al. (2014) identifies four major approaches that previous researchers used to study task difficulty, including: "(1) searcher performance, (2) the match between terms in the task description and in the target page; (3) the number of relevant documents in the collection; and (4) the searchers' or experts' perceptions of difficulty." (Wildemuth et al., 2014, p.1129). Instead of applying a single approach, researchers often leverage multiple approaches at the same time to study task difficulty. For instance, many researchers often design tasks by manipulating the terms in the task description and evaluating users' search performance during the search process, such as query formulation, clicks on SERPs, dwell time on specific web pages, etc. For instance, Arguello's (2014) tasks were designed based on different levels of cognitive complexity that are adapted from educational learning complexity (i.e., remember, understand, analyze, evaluate, create) and measure users' perceptions of task difficulty by post-task questionnaires. In Arguello's (2014) study, they used five questions about task difficulty, including (Arguello, 2014, p.5): (1) How difficult was it to search for information for this task? (2) How difficult was it to understand the information the search engine found? (3) How difficult was it to decide if the information the search engine found would be useful in completing the task? (4) How difficult was it to determine when you had enough information to finish? (5) overall, how difficult was the task?

Capra et al. (2018) manipulated the task items and dimensions described in task descriptions and measured users' expected and experienced difficulty by both pre/post-task questionnaires. J. Liu, Liu, Cole, Belkin, and Zhang (2012) classified TREC task topics depending on the retrieval performance by using the topic titles as queries into their experimental search system. These studies found similar trends that compared to easier tasks, users often formulate more diverse queries, frequently use advanced operators (more often than for easier tasks), and spend more time on SERPs as the search task difficulty increases (e.g., Singer, Norbisrath, and Lewandowski; Aula et al.; Arguello; J. Liu et al.. Capra et al. (2018) found that different task versions (depending by the specificity of items and dimensions included in the task description) can have a significant effect on users' perceptions of task difficulty. Users often expect that tasks with specific dimension descriptions are more difficult, and indeed, they also felt a higher level of difficulty during post-task difficulty evaluation. This might be because specified dimensions of tasks narrow down the scope of potential relevant results and require more effort to search (Capra et al., 2018).

In addition to evaluating users' search behavior regarding different levels of task difficulty, many studies (including several of the above ones) also proposed different computational models to predict task difficulty by users' search activities at (1) the whole search session level or, (2) by their search activities at the first round/query search (e.g., J. Liu et al. (2012), Arguello (2014)). These models are out of the scope of this review and we won't discuss them here.

Regarding task difficulty in XSS, Lin and Belkin (2005)'s validation about the challenges users encountered during a controlled environment when search for multiple sessions using an experimental search system could provide some directions for consideration as efforts regarding XSS task difficulty, such as: articulating information problems, navigating to appropriate information sources, and monitoring the status of their achievement. However, it is still not clear whether and how these challenges are reflected when participants interact with normal search engines or in real-life XSS task process. In their experimental study for examining task stage and task type on users search behavior in a cross-session mode, J. Liu and Belkin (2010) also evaluated participants' perceptions about task difficulty after each search session and by the end of the entire task, however, they did not provide a detailed report or any analysis regarding task difficulty. To our knowledge, so far there is little research about users' perception of XSS task difficulty and search session difficulty or any relationships between the different levels of task difficulty, especially for real-life XSS tasks.

2.2.4 Task process

Task process describes a sequence of steps or stages about how it is completed from the start to the end. Previous work has developed a variety of models focusing on *search task process*, from Kuhlthau's (1991) *Information Seeking Process model*, Bates et al.'s (1993) *Berry picking model*, Dervin's (1983) *Sense-making model*, Wilson's (1997) *model of information behavior*, to Marchionini's (1997) *Information-seeking process model.* It has long been realized that broader work tasks have an important effect on users' information seeking behavior, however, the work task processes are often treated as influential factors and are not reflected by the search process models. Although some of them include task context or "information use" within the model (e.g., Wilson (1997), Dervin (1983)). Most others mainly focus on users' information seeking activities and the connections among those activities and the searcher's status (e.g., emotional, cognitive) (e.g., Kuhlthau (1991), Marchionini (1997)).

By saying so, we do not intend to discount the importance of the pioneers' work, however, we need to realize that information seeking/search models are built to conceptualize human information seeking activities in general at the cost of excepting specific details. The models simplify users' information seeking process at a high abstract level without telling much about how the same activities could change under different task circumstances or when the later actions loop back to previous ones. The models are easy to understand when we treat each search task as an independent task, or each search session as an independent session, or if we only focus on users' information behavior (e.g., seeking, exploring, processing, etc.). On the other side, it reminds us that we should be careful when applying these models to specific task situations as more background and details will be introduced which will complex the process. One example is when work tasks involve cross-session searches over time and when a search session or a search task ends, people's work tasks may have not been completed yet.

During XSS, users conduct multiple search sessions, and each search session may contain multiple search tasks and multiple search queries. On the one hand, we can follow the models to describe users' information seeking activities for each search query round or for each search task. On the other hand, we need to notice that, instead of being independent from each other, all these search tasks, queries, search sessions share the same goal: to complete one work task. Therefore, users' information seeking goals, and other information activities during the process of each search task, or a round of information seeking, are connected in this circumstance. The searcher may conduct similar activities, but these activities could be different as the work task is progressing and users' perceptions of their task status evolve. The same query could aim for different information. For instance, the first use of a query could be used to explore the information space, but a later use of the same query could mean to re-find information they saw earlier or to find updated information on the same topic. To better understand users' information seeking behavior during XSS, we should understand the connections and distinctions between their search processes at different task stages. There is some pioneering work. For instance, Vakkari (2001) proposed a general framework of task-based information searching. This framework contains three task performance stages: *pre-focus, focus formulation*, and *post-focus*. This study reflects the connection between searchers' problem states and their information seeking activities. Vakkari (2003) and Vakkari and Hakala (2000) reveal that users' needs for information types, and their relevance criteria are associated with their problem stages as their tasks progress. They found that during the *pre-focus* stage, users often search for broad background information, like theories and models for their research proposal. In the *focus* stage, their look for more specific information that is expressed in detail for the crucial concepts of the task. They also found that the information users searched for different aspects of the topic and the way users evaluate the importance of certain types of information related to their task performance stage. For instance, the importance of general background information is treated as more relevant at the beginning of their task process and then decrease as searchers' task become more structured and focused.

As we can see, in the previous information seeking/search models, tasks are usually treated as the influential factors outside of the information search loop. However, this might not always be the truth. In other examples, researchers noticed that information search is often an important step to complete the task as part of a broader task process. For instance, people need to find information to help them understand a task problem or alternative decision-making solutions. People who work on different tasks would search for information for various purposes and goals. Even for the same task, information search activities at different task stages could be disparate, although they serve the same task goal. Still, searchers may follow a similar information search process across sessions. But few of the above-mentioned information-process models could reflect the relationships between task progress and information search activities. One possible way to untangle the effects of the task process on information search behavior could be understanding how people intertwine their search activities within the operation of the broader work task that motivates the search, especially for multiple-session searches. As a starting point, understanding the work task process of the ways how people complete their work can help us gain some background information. Researchers from other disciplines have investigated general task processes, also called *task lifecycle* or *task flow*.

Task process is a concept that is frequently mentioned in the Business area or project management area. It describes "the various stages of jobs or tasks from its creation to its destruction" (Van Lingen et al., 2010, p.3), in which the physics workflow process of a task is divided into small slices for different stages (lifecycle stage) or task activities. For instance, the IBM team mapped thirteen different task states, including inactive, ready, running, finished, failed, terminated, claimed, terminating, failing, expired, forwarded, skipped, and stopped [1]. In the automatic task execution scenario, a task will progress through a series of states that have been automatically set up. For example, Figure 2.4[2] describes the lifecycle flow for automatic tasks supported by computer systems:

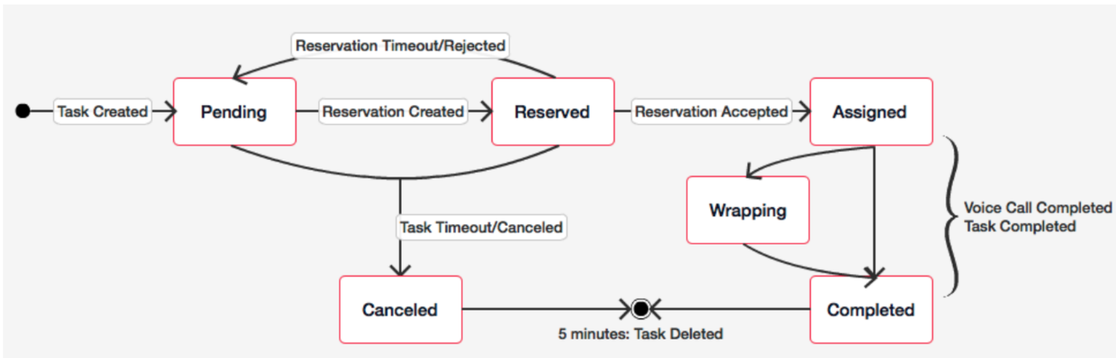


Figure 2.4: An example of automatic task lifecycle by Twilio

Riss et al. (2005) proposed a task management scheme to reflect the process of knowledge-intensive tasks, which is built on the awareness of task-related information activities (e.g., knowledge management, knowledge discovery, knowledge flow analysis) for knowledge workers (Figure 2.5). This model addresses the users' interaction with three types of information: (1) task process patterns which can be used for structuring the task into suitable sub-tasks, (2) task-related information to support the execution of the task, (3) task process relation between different information units and subtasks (Riss et al., 2005).

[1] https://www.ibm.com/support/knowledgecenter/SSFPJS_8.6.0/com.ibm.wbpm.admin.doc/topics/t7taskadminovr.html
[2] https://www.twilio.com/docs/taskrouter/lifecycle-task-state

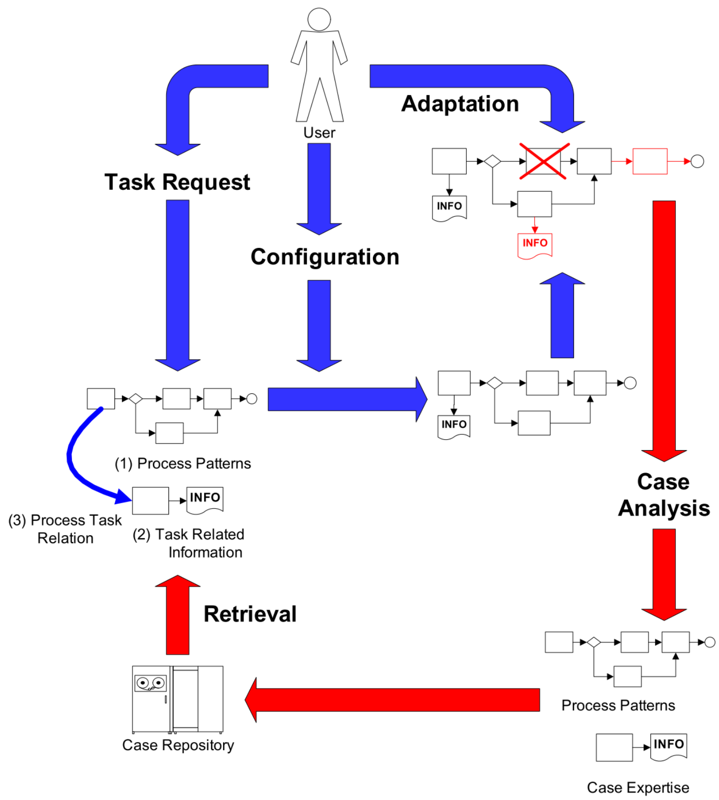


Figure 2.5: An example of the process of how users (knowledge workers) handle tasks and processes by Riss et al. (2005)

As we can see, work task processes are more complex due to the diversity of disciplines or industry backgrounds. The three task lifecycle models mentioned above do not intend to cover all kinds of task models in business and organization areas. Instead, they represent three descriptions of the task process from different perspectives: the IBM model shows a general procedure from the start of a task to its end without specifying the task subject (i.e., human or computer). The Twilio model shows an example of a computer-handling task procedure without human input. In contrast, the model of task process for knowledge workers mainly focuses on how workers deal with different types of information through various task stages. Understanding the general workflow can help us gain an overview of the work task procedure, which is very important for understanding and analyzing task context for cross-session search.

2.2.5 Task analysis

The diversity of task models is also reflected in the HCI area. Research on tasks can be always found in user-centered design areas, such as User Interface (UI) design and User experience (UX) design. For each specific interface design or examination of a website or application's current status, researchers will focus on the type of task that the application/website is designed. Hence, task analysis is an important method that can help researchers investigate relevant task processes and users' activities through the performance processes to develop and evaluate their systems' functions. *Cognitive Task Analysis* is commonly used by HCI researchers.

Cognitive Task Analysis (CTA) is a family of task analysis techniques focusing on uncovering and representing the cognitive activity (e.g., what people know and how they think) that is required by a certain task (Crandall, Klein, Klein, & Hoffman, 2006). The primary phases of CTA include: (1) getting familiar with the domain background, (2) eliciting the tacit knowledge and thought processes of experts, (3) analyzing, assembling, and representing the knowledge via tables, charts, or diagrams (Crandall et al., 2006) (For comprehensive review: Crandall et al. (2006)).

Hierarchical Task Analysis (HTA) and Goals, Operators, Methods, and Selection Rules (GOMS) are frequently used CTA methods. HTA can describe an activity in terms of its specific goals, subgoals, operations, and plans. The purpose of HTA is to produce a task hierarchy of steps that must be completed to complete some high-level task (Annett, 2003). This method is very suitable for analyzing tasks that are well-structured, such as tasks that tend to be performed in similar ways every time (Stanton, 2006).

GOMS is a description of the procedural knowledge that a user must have to carry out tasks on a device or system (Kieras, 1988). With additional techniques, GOMS can provide useful qualitative descriptions of how users interact with systems to perform a task and predict their learning and performance. GOMS only applies to skilled users but not to beginners if they are starting to use a system. Another disadvantage of the GOMS is that all users are assumed to be the same, and their personalities or other physical restrictions are not taken into account, which alters the analysis (Tonn-Eichstädt, 2006).

Inspired by the previous research, in this proposed study, we are aiming to identify the common cognitive activities and stages involved in the XSS work tasks process and analyze the challenges users encounter and their needs for assistance when their cognition is limited during certain points of the process.

CHAPTER 3

Research Questions

Motivated by the fact that previous work has not thoroughly investigated the effects of the complexity feature of cross-session search and the relationships among the various reasons causing cross-session search behavior, my dissertation focuses on the following research questions:

- **RQ1:** What are the stopping and resuming reasons leading people to search for multiple sessions? Are the reasons related and how?

- **RQ2:** What are the effects of session-level search goal complexity on the types of information sought, the cognitive activities actually involved during the session, and the resuming/stopping reasons of the session?

  RQ2a: What are the effects of session-level goal complexity on the types of information sought during the session?

  RQ2b: What are the effects of session-level goal complexity on the cognitive activities involved during the session?

  RQ2c: What are the effects of session-level goal complexity on the sessions' resuming/stopping reasons?

- **RQ3:** What are the effects of the information types used and cognitive activities involved during a session on the session resuming/stopping reasons?

  RQ3a: What are the effects of the information types used on the resuming/stopping reasons at the session level?

  RQ3b: What are the effects of the cognitive activities involved on the resuming/stopping reasons at the session level?

- **RQ4:** What are the effects of task stages on participants' perceptions of pre-task, post-task, and post-session difficulties?

RQ4a: What are the effects of task stages on participants' perception of the overall difficulty of their tasks and search sessions?

RQ4b: What are the effects of task stages on participants' perception of the difficulty to search for information?

RQ4c: What are the effects of task stages on participants' perception of difficulty to find enough information?

RQ4d: What are the effects of task stages on participants' perception of difficulty to integrate the found information?

RQ4e: What are the relationships between participants' perception of difficulty and the session resuming/stopping reasons?

- **RQ5:** What challenges did participants face due to conducting searches across multiple sessions?

Figure 3.1 describes a framework of the context for my research and the primary focuses for each research question: (1) The central part represents an abstract workflow of cross-session search with two consecutive sessions (session N, session N+1), along with two groups of reasons (stopping reasons and resuming reasons) that I studied (RQ1). (2) The left side represents the two XSS task factors I focused on: the task complexity (RQ2) and the (expected) task difficulty (RQ4) at the whole task level; (3) The right side represents users' information behavior and experience: the information they sought for each session, and cognitive activities involved as they process the information to complete their tasks (RQ3), their evaluation of experienced task difficulty at both whole task and search session levels (RQ4). As shown at the bottom of the framework, I analyzed the challenges users face during their XSS processes to provide system design ideas for future studies (RQ5).

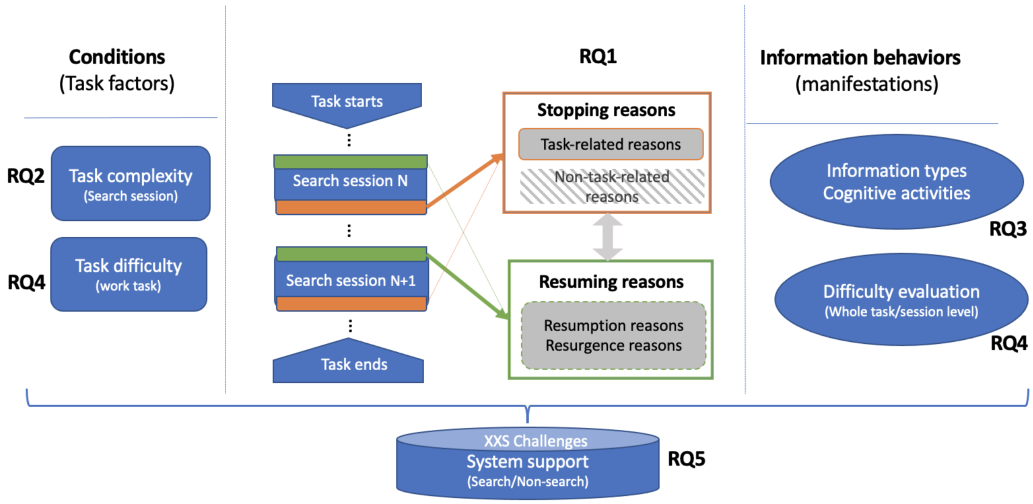


Figure 3.1: Cross-session search research framework and research questions

Below, I explain each research question, the associated sub-questions, and how I collected and analyzed data to answer these questions separately.

RQ1:What are the stopping and resuming reasons leading people to search for multiple sessions? Are the reasons related and how? Previous studies have explored the reasons that lead people to search across sessions. Spink and her colleagues (1996; 2002) conducted a series study about academic library users' cross-session search behavior and identified ten reasons why users need to conduct successive searching (Figure 2.2). MacKay and Watters (2008a, 2008b) collected students' cross-session search data through diaries and search logs and identified a list of reasons that make students stop searching, which include both task-related reasons (e.g., finished sub-task, or task completed) and non-task related reasons (e.g., take a break, frustrated, or run out of time). Lin (2001) developed the MISE model, which conceptualized eight different session resuming reasons and two groups of stopping reasons (i.e., external interruption and internal interruption), and later they designed an experimental study (Lin & Belkin, 2005) and verified the model. Most recent studies on XSS focus on developing computational models to detect, and predict task continuation based on analysis of search log data (e.g., Agichtein et al. (2012), Kotov et al. (2011)) without users' input about the reasons for cross-session search or other task context. However, some limitations to these studies are: many of them were conducted under a controlled environment; some are limited to the study task types and the databases users interacted with; some have no searchers' input. As we gain some knowledge about the XSS reasons, little is known about how these reasons are manifested in real-life tasks from users' perspectives.

Recently, we conducted a survey study on the Amazon Mechanical Turk platform about the reasons that cause people to search for multiple sessions in their daily work and life (Li et al., 2020a). The reasons causing XSS and data sources are summarized in Table 3.1. Our findings show that we can find all the MISE modes in users' daily work and life XSS process. We also found that participants often choose multiple reasons why they started/stopped one search session, and these reasons are affected by the number of search sessions they conducted for the task. Thus, we hypothesized that users have different reasons to start or stop their search sessions during XSS, and the emergency of these reasons contain specific patterns through their XSS process.

Table 3.1: Reasons causing XSS and data sources (RQ1)

Data	**Data source 1**	**Data source 2**
Session sequence	Post-session questionnaires	Search history
Unique resumption reason	Post-session questionnaire	
Unique stopping reason	Post-session questionnaire	
Order of resumption reasons	All ordered post-session questionnaires	
Order of stopping reasons	All ordered post-session questionnaires	

To answer RQ1, I leveraged the following data: session sequence number (e.g., first session, second session, etc.), individual session resuming reasons, individual session stopping reasons, and their orders. The session sequence and reason orders were recorded based on the order when they filled out the search session questionnaire each time they completed a session. For each search session they finished, the post-session questionnaire asked the participant to (1) describe the reasons why they need to start and stop this session separately in their own words; (2) select the most important resuming reason (from the MISE list) for the session; (3) select the most important task-related reason why they stop that session. My analysis (a) summarized the frequency of different stopping/ resuming reasons, (b) summarized the sequence of resuming reasons for each work task, (c) summarized the sequence of stopping reasons for each work task, (d) paired the most important resuming reason with the most important stopping reason by sessions, (e) paired the most important stopping reason of a previous session with the most important resuming reason of the following session, and (e) investigated if there are any patterns among the resuming reasons (by themselves) within a work task, the stopping reasons (by themselves) within a work task, if there were any patterns among the pairs of the resuming reasons and stopping reasons for all sessions of all tasks, and if there were any patterns among the pairs of the stopping reasons for a previous session and the resuming reasons for the following session for all sessions of all work tasks. These analyses helped gain knowledge about the frequencies of different types of stopping and resuming reasons, the patterns of the order of resuming reasons and the stopping reasons within the work task, and the trends between the pairs of the reasons within the same session and across sessions.

RQ2: What are the effects of session-level goal complexity on the types of information sought, the cognitive activities actually involved during the session, and the stopping/ resuming reasons for the session? Task complexity can be measured in numerous ways using different task features. Wildemuth et al.'s (2014) review summarized the approaches of how prior studies define and operationalize task complexity into six categories, including (Wildemuth et al., 2014, p.1122): (1) number of subtasks or steps; (2) number of subtopics or facets; (3) number of query terms and operators required; (4) number of sources or items required; (5) the indeterminate nature of the task; and (6) the cognitive complexity of addressing the information need.

As task complexity is an objective task attribute, the levels of the complexity of a task would not be affected by participants' characteristics or their task activities that are actually involved to complete the task. In other words, task complexity can be determined independently based on the task itself. Prior studies often manipulated task complexity when designing simulated tasks in a controlled environment to study its effect on users' search behavior, especially for search tasks that involve a single search session. Regarding XSS tasks, researchers often describe tasks that involve XSS as *complex tasks* (e.g., J. Liu and Belkin (2010), Agichtein et al. (2012), Czerwinski et al. (2004)). The focuses of these studies include task topics (MacKay & Watters, 2008a), search task types (e.g., information gathering tasks) (Sellen et al., 2002), or collaborative tasks (Capra, Marchionini, Velasco-Martin, & Muller, 2010), etc. A few of them analyzed XSS task complexity and its effects. Czerwinski et al.'s (2004) diary study studied task switching and interruption, and they found that *complex tasks* (e.g., long-term projects, more documents involved) are harder to be resumed compared to more manageable tasks. Based on human annotated search log data, Agichtein et al. (2012) particularly defined search task complexity by the number of task goals (i.e., single goal, multiple goals, undirected) and found that undirected tasks are often continued later. But they did not find the relationship between task goal numbers and task continuation. Byström and Järvelin (1995) investigated the effects of task complexity on users' information needs. Although the study is not about XSS, but many of the collected tasks contain multiple work sessions. They found that the task complexity at the whole task level could positively affect the complexity of users' information needs (e.g., information types, information complexity, information sources).

Many of the abovementioned studies indicate that complex tasks can spread over time and people often work on different parts of their tasks across multiple sessions. Therefore, it is possible that some parts or sessions of their cross-session work could be easier or more complex than others, and their information needs and behaviors during the session could be affected by the level of the session goal complexity at the current stage. RQ2 intends to explore, for an overall complex work task that involves cross-session searches, how the individual session goal complexity could affect the information they search for during the session, their cognitive activities involved during the session, and the resuming and stopping reasons for that session.

Accordingly, RQ2 is broken down into three sub-research questions: *(1) RQ2a: What are the effects of session-level goal complexity on the types of information sought during the session? (2) RQ2b: What are the effects of session-level goal complexity on the cognitive activities actually involved during the session? (3) RQ2c: What are the effects of session-level goal complexity on the session's resuming/stopping reasons?*

Here we defined the session-level goal complexity based on the cognitive complexity (e.g., remember, understand, apply, analyze, evaluate, create, from lower-level complexity to higher-level complexity) of the participant's main goal for the session.

To perform this analysis, I designed a *pre-session questionnaire* (See Appendix F) to collect users' session goal(s) and the cognitive activities that they *anticipate* for the session goals (e.g., I need to find specific factual information vs. I need to learn about important concepts). The *pre-session questionnaire* contains three quick questions to answer before the participant starts a new search session. The three questions include an open-ended question about the user's search goals for this session, a list of options that describe the different types of information and cognitive activities that apply to the participant's own descriptions of their goals, and the same list of options, but requires participants to select one that can best match the participant's main goal for the session. The options list is adapted from Urgo et al. (2020)'s work, where they used the descriptions to validate the task complexity manipulation and to evaluate users' perceptions of the cognitive activities and knowledge types for the designed tasks.

When reading the list for the first question, a participant would select any options from the list that can apply to their goals or sub-goals for the session. For the second question with the same list, the participant must *only* select the option that can best match the main goal of the session. There are two purposes of this design: (1) by asking them to select all that apply their (sub)goals, I wanted to gain a comprehensive understanding of as many as possible purposes the participant wants/plans to achieve during the session, (2) by asking them to select the best option that matches their main goal of the session, I intended to learn the crucial goal of that session—which was be used to define the session goal complexity level. For instance, a participant wanted to achieve multiple goals involving collecting facts and comparing alternatives to develop a solution for the session (all goals). But the essential goal for his/her session might be first to find the factual information and then consider what they can do if the session keeps continuing. In this case, the session goal complexity should be defined by looking for factual information instead of comparing alternatives or coming up with new solutions. From this perspective, we gained knowledge about the participant's session goal and its complexity right before they started the session and also minimized the effects of their actual work experience during the session.

The information types and cognitive activities are concepts adapted from educational areas from Anderson and Krathwohl's (2001) learning taxonomy. Recent studies found that task complexity manipulated from these two perspectives shows significant effects on users' information search behavior during single-session search (e.g., Kelly et al. (2015), Arguello (2014), Urgo et al. (2020)). We adopted these two dimensions to analyze our AMT survey responses. We found that participants (67% out of 110) often search for multiple types of information or multiple pieces of the same types of information and often apply at least two (more than 80% out of 110) cognitive activities (e.g., understand, apply, analyze, evaluate, create) to complete an XSS work task (Li, Ward, & Capra, 2021a). But similar to the session starting and stopping reasons collected by the survey, we could not collect the sought information types or the involved cognitive activities corresponding to each session as participants worked on their XSS tasks.

A session with a lower-level cognitive complexity goal (e.g., remember, understand, etc.) doesn't necessarily mean people would look for less amount of information or information with lower-level complexity. For instance, a participant may need to find a specific recipe they heard from a friend (remember the goal, lowest level of cognitive complexity) during a session, and it turns out the recipe contains a long list of ingredients and complicated instructions (higher-level information/knowledge complexity). Similarly, a session goal's cognitive complexity does not equal the levels of all cognitive activities that people actually conduct during the session. A session with a lower-level goal complexity could actually trigger various types of cognitive activities when the participant interacts with the information they encounter during the session. A session with a higher-level cognitive goal could also involve lower-level cognitive activities such as looking for specific background information, understanding relevant concepts, etc.

In the diary study, participants provided examples for the information they sought, the URLs, and how they think they found information can (not) help them to complete the task in the *post-session questionnaire*. I analyzed the information types based on the examples they provided (or retrieve the information using the URL of the information). And I analyzed the involved cognitive activities based on their descriptions of thoughts of the information they found through the session (e.g., recognize, recall, identify, use, interpret, develop). In addition, the search history log and their descriptions of the session process also provided supplementary information about the information they sought and actual cognitive activities involved in the session. Together with the resuming and stopping reasons collected from the *post-search session questionnaire*, I then investigated the effects of the session-level goal's cognitive complexity on (1) the information types participants sought (RQ2a), (2) the involved cognitive activities (RQ2b), (3) the session resuming and stopping reasons (RQ2c). Table 3.2 shows the data sources related to RQ2 and its sub-questions.

Table 3.2: Data sources of session-level goal complexity, information types, cognitive activities, and resuming/stopping reasons (RQ2)

Factors	**Data source 1**	**Data source 2**
Session-level goal complexity	Pre-session questionnaire	
Involved information types	Post-session questionnaire	Search history
Involved cognitive activities	Post-session questionnaire	
Session resuming reasons	Post-session questionnaire	
Session stopping reasons	Post-session questionnaire	

RQ3: What are the effects of the information types used and cognitive activities involved during a session on the session resuming/stopping reasons? Built upon the understanding of the information types involved in each search session and their cognitive activities about how to deal with the found information in RO2, RQ3 aimed to find the relationships between the information types & cognitive activities and the resuming & stopping reasons for a corresponding session.

The eight session-resuming reasons described in the MISE model indicate several different problem stages. More precisely, these stages reflect users' subjective perception of the dynamic change of the information problems they need to solve to complete their task. In other words, the resuming reasons indicate a specific "problematic cognitive state" (Lin & Belkin, 2000, p.28) that represents users' understanding of some aspects of the task problem that they need to solve by looking for information/knowledge that is not at hand.

Based on the review of theoretical literature and previous research on users' information seeking process and their search strategies, Lin and Belkin (2000) concludes that the properties of the objective context (e.g., types of problematic cognitive states, goal/task, a timeline of goal, priority, urgency or clarity, the complexity of information problems) can affect users' information-seeking activities and the"treatment" (referring as searchers' expectation about the useful information) (Lin & Belkin, 2000, p.13) in the search sessions of XSS. The information-seeking activities include different types of information-seeking strategies for articulating information problems, evaluating search results, monitoring their cognitive problematic states, etc. and the treatment's properties involve the types of data (i.e., soft data–abstract not directly observable, hard data–concrete, empirically derived) (Lin

& Belkin, 2000, p.14), the amount of information, its usage (e.g., applied/instrumental/immediately useful/know-how, substantive/descriptive/know-what), or the level of information (e.g., full documents, passages, sentences, etc.). Their later work (Lin & Belkin, 2005) confirmed that these properties could be observed when people search across sessions in a controlled lab study.

Besides, many other studies found the impact of task complexity on users' needs for information and their information search strategies even for different stages within a single-session search, often in a controlled environment. We reviewed some of the studies earlier in *Section 2.2.2*.

Together, these findings provide the foundation of the assumption that: there are correlations between (a) the resuming reasons (i.e., users' perceptions about their task needs), stopping reasons (i.e., the relevant strategies to deal with the information to complete their tasks), and (b) the properties of the sought information and the cognitive activities involved in the XSS process.

Therefore, my analysis specifically focus on the following two sub-questions:

RQ3a: What are the effects of the information types used on the stopping/ resuming reasons at the session level? In this question, I hypothesized that the types of information people interacted with during the sessions relate to the reasons why they started and stopped the session. (Note: we will also include the initial session.)

To answer this question, I used two types of data that have been mentioned earlier: (1) the pair of resuming and stopping reasons for each session that participants answered in the post-session questionnaire. (2) the involved information examples and any other information participants mentioned in the post-session questionnaire (e.g., their description of the search session, comparing the current session and the previous session).

RQ3b: What are the effects of the cognitive activities involved on the stopping/ resuming reasons at the session level? In this question, I hypothesized that the cognitive activities that are actually involved in a session relate to the session resuming and stopping reasons.

Similar to RQ3a, the resuming and stopping reasons were collected in the *post-session questionnaire*. The cognitive activities during the session were derived from participants' descriptions of how they think the found information can be helpful for their tasks, as well as from their descriptions of the search session process. There are two more sources of data that helped us gain more knowledge about participants' cognitive activities with the information they found: (1) the participant's daily review about how they used the found information to complete their task during the day; (2) their descriptions of what they had done with the task before they started a later session (in the pre-session questionnaire). To gain knowledge about their cognitive activities, first, I extracted participants' own words from the relevant questions. I categorized users' activities into the different complexity levels of the cognitive level group (e.g., retrieving, understanding, applying, etc.) following (Anderson & Krathwohl, 2001) and analyzed the relationships between their cognitive activities and the session resuming and stopping reasons.

RQ4: What are the effects of task stages on participants' perceptions of pre-task, post-task, and post-session difficulties? In addition to investigating XSS work task complexity, I also measured participants' perception of task difficulty on both the work task level and search-session level. In Section 2.2.3, I reviewed previous studies on task difficulty and their findings of the effects of task difficulty on users' information search activities for single-session search in the controlled environment. Participants' perceptions about task difficulty can be affected by the objective task complexity. Previous studies often measure task difficulty by asking participants to rate their evaluation of the efforts that require to complete the task before they start the search (as pre-task difficulty), and re-evaluate their efforts for completing the task after they finish the task (as post-task difficulty) (J. Kim, 2006b; Arguello, 2014).

J. Liu, Liu, Yuan, and Belkin (2011) designed different task structures (i.e., parallel verse dependent) to evaluate the changes in participants' perceptions of task difficulty when they search for subtasks across sessions. They found that in the dependent-structured task (the later subtask is based on the results of the previous task), participants' perceived task difficulty did not change before and after the task. However, for parallel tasks (subtasks are independent of each other), participants reported that tasks were less difficult than their expectations. It might be because the subtasks are similar to each other; participants become more familiar with the task topic and task structure in later sessions so that they felt less difficult by the end of the study.

In addition, there are some conflicting results about the effects of task descriptions quality and requirements ("prior determinability") on users' perceptions of task difficulty: J. Liu et al. (2011) found that compared to a task that includes specific descriptions about facts and requirements of goal, participants perceived higher task difficulty for completing a task whose description does not include factual information or specific requirement for the goal. In contrast, Capra et al. found that by specifying both the items (e.g., factual information) and the dimensions of the results in comparative tasks, participants reported greater post-task difficulty than tasks that contains less (or no description) about specific items or dimensions. The differences could be caused by the simulated task topics or different participants. More importantly, one significant difference between these two studies is that the former is about XSS and the latter is about single-session search. These differences indicate that it could be possible that even for similar tasks, people's perceptions about task difficulty could be different for tasks that spread over time compared to tasks that can be completed within one search session. Therefore, it is interesting to learn more about how their subjective perceptions of task difficulty at different levels (i.e., the whole task level, individual search session level). Especially, I analyzed the relationships for the following research questions: *(1) RQ4a: What are the effects of task stages on participants' perception of the overall difficulty of their tasks and search sessions?; (2) RQ4b: What are the effects of task stages on participants' perception of the difficulty to search for information? (3) RQ4c: What are the effects of task stages on participants' perception of difficulty to find enough information? (4) RQ4d: What are the effects of task stages on participants' perception of difficulty to integrate the found information? (5) RQ4e: What are the relationships between participants' perception of difficulty and the session resuming/stopping reasons?*

Using the same group of task difficulty questions adapted from (Arguello, 2014) (i.e., search difficulty, understand the difficulty, decide difficulty, integrate difficulty, overall difficulty), I measured participants' pre-task difficulty of the whole task by asking participants to answer the difficulty questions in the *pre-task questionnaire* (during the introduction meeting session), which is before they start searching for their task. I measured their evaluation of the post-task difficulty of the whole task in the last *daily review diary* when they completed all their searches for the task. During the study period, I asked participants to evaluate their experienced (post-) session difficulty for each search session in the *post-session questionnaire* so that I could determine the change in task difficulty across sessions and then compare the session difficulty to the pre-/post-whole task difficulty levels. Chi-square tests and ANOVA tests were performed to answer these questions. Table 3.3 shows the data sources for each of the variables involved in RQ4:

Table 3.3: Task difficulty and data source (RQ4)

Task facet	Data source
Pre-task difficulty (whole task)	Pre-task questionnaire
Post-task difficulty (whole task)	Last daily review diary
Post-session difficulty (session)	Post-session questionnaire
Session resuming reason	Post-session questionnaire
Session stopping reason	Post-session questionnaire

RQ5: What challenges did participants face due to conducting searches across multiple sessions? Previous studies on XSS often report supporting participants' XSS by providing assistance for refinding the information resources they found before, saving their search trails (e.g., sequential search queries), and extended functions for participants to record their task progress when work on tasks spread over time (e.g., Morris et al. (2008a), MacKay and Watters (2009), Gomes and Hoeber (2021)) (See Section 2.1.5). Many of these supports are developed based on surveys requiring people to reflect on their difficulties during past XSS experiences and specifically focused on information searching activities. In reality, complex tasks or long-term tasks may involve many different activities besides searching for information, even for tasks involving a single search session. Researchers found that people spent two times on the task after finding the relevant information (Toms, Villa, & McCay-Peet, 2013).

Our preliminary study found that 46% of users expressed that they don't need to review previously found information, and many users monitor their task progress through their ongoing work. Zhang et al. (2020) identified four types of challenges users encounter when they work on creative tasks, including challenges during creative processes (e.g., figuring out design options, clarifying ideas, making decisions), difficulties regarding information searching (e.g., finding exact information, extracting information, cannot find what they wanted), challenges about information organization, motivation to continue the task, and time pressure, etc.

A few studies addressed the challenges related to search across sessions that we aimed to address in the current study. During the interview, I asked participants about the challenges caused by searching across multiple sessions, and open-coded their responses to answer this question.

CHAPTER 4

Method: Diary Study

4.1 Online diary study

To investigate the research questions described in Chapter 3, I conducted an online diary study combined with questionnaires and interviews.

Diary study has been commonly used as a data collection method in information and library science and studies of human-computer interactions. Diaries hold great promise from some studies of information seeking and use because study participants record their thoughts concurrently with the information interactions at the moment when it happens, or as close as to the moment (Wildemuth, 2002), which can keep much more details in-situ while reducing recall biases. Diary studies are suitable for capturing the natural context information that is usually hard to gain by laboratory experiments or pure search log data and minimizes the effects of observers' participation (Carter & Mankoff, 2005; Bolger, Davis, & Rafaeli, 2003; Gunthert & Wenze, 2012; Czerwinski et al., 2004). On the other side, diary studies have some disadvantages. It is often associated with the risk of low control, and low participation because of the higher burden caused by the intensive repeated assessment over time, which might discourage participation or result in a sampling bias related to motivation and personality factors (Scollon, Kim-Prieto, & Diener, 2003).

Usually, diary study participants need to follow the study instructions and keep the diary recordings over time without a study moderator's presence. In some cases, participants might miss valuable input because they thought it is not sufficiently interesting to record it (Bolger et al., 2003). Besides, filling diaries over time can be a time-consuming task that needs effort from participants. Since it is not part of their previous routine, participants may easily forget to record a diary when they should (Czerwinski et al., 2004). Sometimes the circumstance may not be appropriate for participants to record a diary so they may try to make it up later based on memories. Therefore, a successful diary study needs good commitment from the participants who should be convinced to complete the diaries (Pickard, 2013).

Therefore, I carefully designed the study procedure and took possible issues into consideration. One of the strategies was to recruit UNC students as participants for this online diary study instead of Amazon Mechanical Turk (AMT) workers. Researchers found that AMT workers prefer to complete studies that require less effort within a short time (e.g., one sitting session) and often postpone complex tasks later (Wang, Sarkar, & Shah, 2018). A past study (Zhang, 2020) in our lab shows that college students are well-engaged with diary studies, willing to follow instructions, and less likely to quit before their studies end. Additionally, after I recruited participants, I set up an initial introduction meeting to address the study requirements at the beginning of the study. During the diary study, I monitored participants' responses to the questionnaires and daily work reviews, sent them reminder emails, and used monetary incentives to motivate their participation.

4.2 Study design

The diary study contains four main parts, as shown in Figure 4.1: (1) Participants recruitment and screening; (2) Initial introduction meeting; (3) Diary study; (4) Interview. Below I first briefly introduce the study protocol, and then I provide more details for each part separately.

4.2.1 Study protocol

Before starting the diary study, I set up an *introduction meeting* with each participant. During the introduction meeting, a participant went through and signed the consent form. I instructed them to set up a new profile on their computers in the Google Chrome browser and log in with a pre-set Google study account. Whenever they need to search for information for a task, they shall use the account to search. Their search queries, search engines, and any bookmarks would be recorded in the study account's search history, separated from their personal Google account (see more details in Section 4.4).

After setting up the account and logging in, the participant saw the following bookmarked links on the study account toolbar: a *pre-task questionnaire*, a *pre-session questionnaire*, a *post-session questionnaire*, and a *daily review diary*. First, the participant clicked and filled out the *pre-task questionnaire* during the introduction meeting. Then they were instructed about when (i.e., session

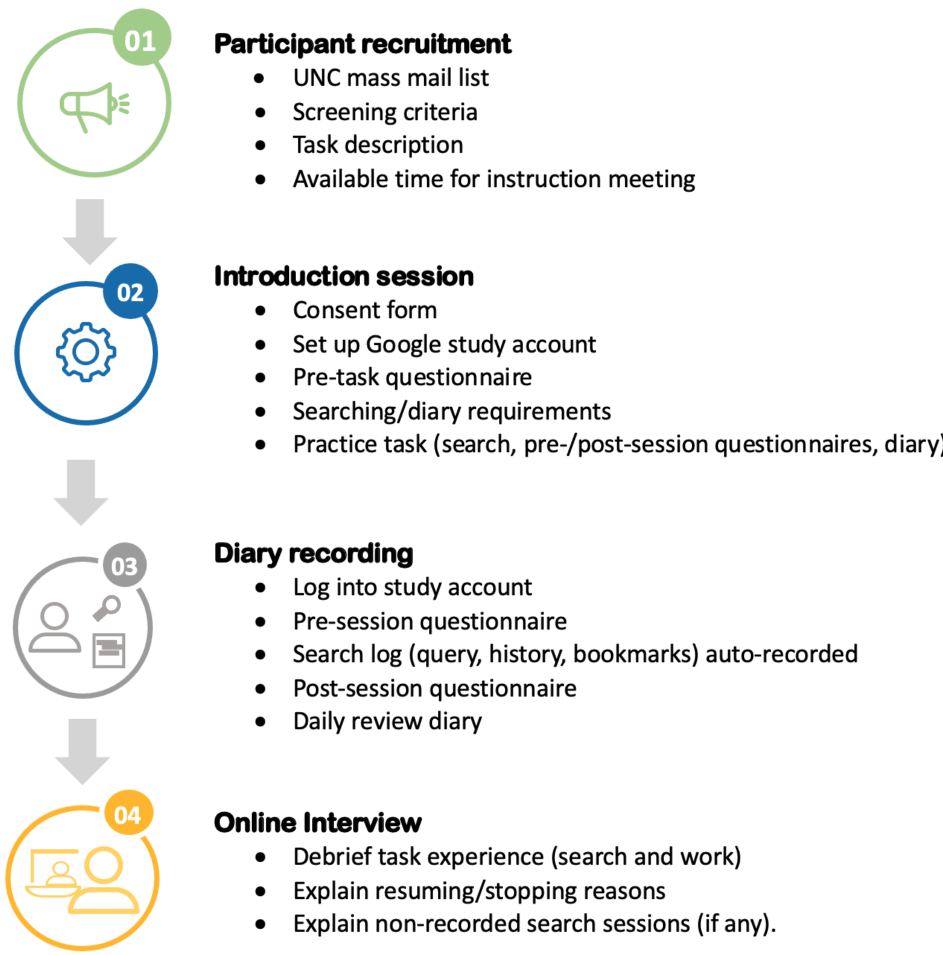


Figure 4.1: Cross-session search online diary study protocol

definition, the beginning, and endpoint of a session) and how to fill out the pre-/post-search session questionnaires and a daily review diary (see more details in Section 4.6.2). Then the participant conducted a practice task search session to familiarize themself with the steps to fill the questionnaires and diaries. They can also ask any questions about the questionnaires and the daily diary review to make sure they understand all the requirements.

After the introduction meeting, participants worked at their own pace to work on their tasks and complete the questionnaires and diaries during the process. The process was like this: when the participant decided to search for the task, they would switch to the Google study account and immediately fill in the *pre-session questionnaire* for the session, and then they could start to search. They were encouraged to bookmark useful web pages through the search process. When they decided to stop the current search session, they would follow the link on the toolbar to fill out the *post-session* *questionnaire*. They were required to fill out both the pre-/post-session questionnaire for each search session they conducted. On a day when a participant did some work for their tasks, including information seeking, and any other task activities, they were required to record a daily review diary by the end of the day about their task experiences. This process cycled till the participant indicated in a daily review diary that they had completed the entire task and decided not to search anymore for the task and would like to end the study or participate in the exit interview (Figure 4.2).

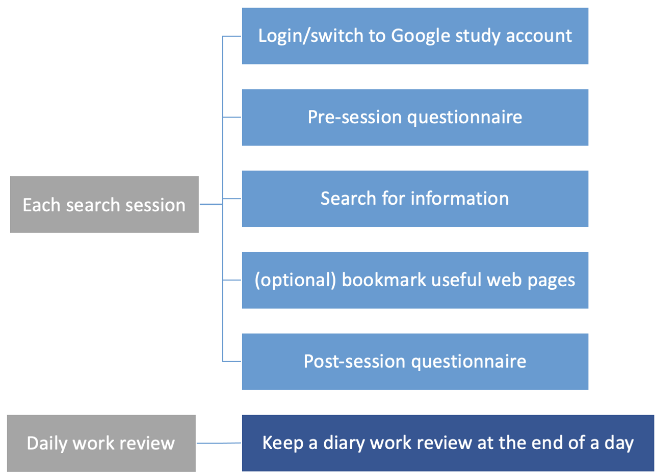


Figure 4.2: Participants' diary/questionnaire recording workflow

During the interview session, I presented a summary of participants' questionnaire responses, their diaries, and logged data. I asked them questions about their perceptions of the overall search experience, their interpretation of task phases of the task, challenges they met, etc. (See more details in Section 4.6.4). The interview lasted for 30 minutes. At the end of the interview, I sent an Amazon Gift Card Code with the appropriate amount of money (e.g., $50 for the diary recordings, $20 extra reward for higher-quality responses, $20 dollars for the interview) as compensation for their participation.

4.3 Participants' recruitment and screening

For the diary study, I planned to recruit 30 participants (by considering that 5 to 8 people may drop off, about 16% to 26%). The recruiting email (see Appendix A.1) was sent out through the UNC mass mailing list. The mass mailing list enabled us to reach a large population set with UNC students, faculty, and staff. In the recruiting mail, I explained the purpose of the study, the requirements for participants, the study procedure, and payment, etc. To select appropriate participants, I also set up some criteria. The basic criteria were listed in the recruiting email. About participants themselves, I required them to be: (1) more than 18 years old; (2) native English speaker, (3) affiliated with UNC as employees or students.

In terms of requirements for their tasks, I provided a link in the recruiting email to an online screening questionnaire (see Appendix A.2). The questionnaire asks the person to describe a task or project that s/he plans to do within the next three weeks. This task would be the *selected task* for the participant to complete during the study period, and it needs to fulfill the following requirements:

- The person has not started to work on the task yet.
- To complete the task, the person needs to search for information online.
- They plan to spend several days completing the task.
- They anticipate doing Internet searches at different times or days to complete the task.

The person also was required to provide information about the expected outcome(s) of the selected task, the time range s/he planned to start and complete the tasks, and brief explanations about why they thought the task would take time and why they needed to search for information online over time for completing the task.

There are multiple reasons why I required participants to select a task they plan to do shortly and set up these criteria for the selected task: (1) many tasks in daily work and life need us to search for multiple sessions. However, not all XSS tasks are expectable. Not all participants know exactly how many sessions they will search in advance. Sometimes, a "simple" task that the task doer thought would be completed by a single search session could turnout to be a complex one and require more search sessions later as the task doer becomes more engaged with the task. For these "unexpected" XSS tasks, it would be hard to know early whether there would be multiple search sessions. Considering that our purpose of this study is trying to capture users' perceptions about their XSS tasks before they start to search and work, limiting the task types to the expected XSS tasks will provide us with more details and cover as much as possible information from the beginning to end, which can help gain a holistic view of XSS. (2) Based on prospective participants' responses about the tasks they were going to do, I could select participants with different types of tasks and avoid certain types of tasks that may provide less diverse data. Previous work about XSS conducted with participants from the same institutions often found that people reported similar tasks (e.g., similar work projects or many school assignments by students(e.g., Morris et al. (2008a), MacKay and Watters (2008a)), which may cause bias or undermine the generalizability of the study results. I can also avoid participants who indicate that their XSS task is a daily routine task (e.g., reading news, checking emails, watching videos). Routine tasks refer to tasks that people perform repeatedly. When working on routine tasks, participants use the same information strategies, such as navigating to the same resource and displaying the same behavior (I. Xie, 2009). Previous research (Li & Belkin, 2008) found that 70% of routine work tasks are associated with short-term search tasks, while 83.3% of unique work tasks were related to long-term search tasks. Therefore, for the current study purpose, I decided to leave out the routine tasks. (3)By asking people pre-select an XSS task and keep diary recordings focusing on that task, the participants would have some expectations about when they need to fill the search session questionnaires and don't need to worry about whether they should fill out a questionnaire or a diary for any tasks that turn out to be XSS tasks. From this perspective, asking them to focus on the selected task will significantly reduce their workload and meet the study requirements.

Once I received responses, I discussed with my advisor to select the participants whose tasks were qualified for this study. The criteria for screening mainly focus on whether participants provide enough details about their tasks and whether the reasons why they needed multiple days to complete the task were related to the complex task process. Additionally, we also tried not to recruit too many similar tasks (e.g., school assignments, research papers). Then I sent an email to participants that had been recruited to schedule a one-hour introduction meeting (via Zoom). One of the advantages of a diary study is that different participants can start work simultaneously since they work without much interaction with the moderator. Thus, I could collect data from multiple participants at the same time. During the time between August 16th to October 15th, 2021, I recruited 29 participants in total. After 4 of them dropped off, 25 participants completed the diary study, and 15 participants who completed the diary study also participated in the retrospective interview.

4.4 Diary study set up

This study asked users to work at their own pace to search for information for their own tasks using their own computers and fill in multiple questionnaires/diary recordings that may spread over a few days. Therefore, I asked users to use Google Chrome as the web browser and sign in with a pre-set Google study account when they search for the task. For instance, for a pilot participant, I set up an account called: xssperson2@gmail.com, with a unique password generated automatically using password-generating software. I also set up the questionnaire and diary entry as bookmarks for each account, so when participants log into their study account, their page would look like Figure 4.3.

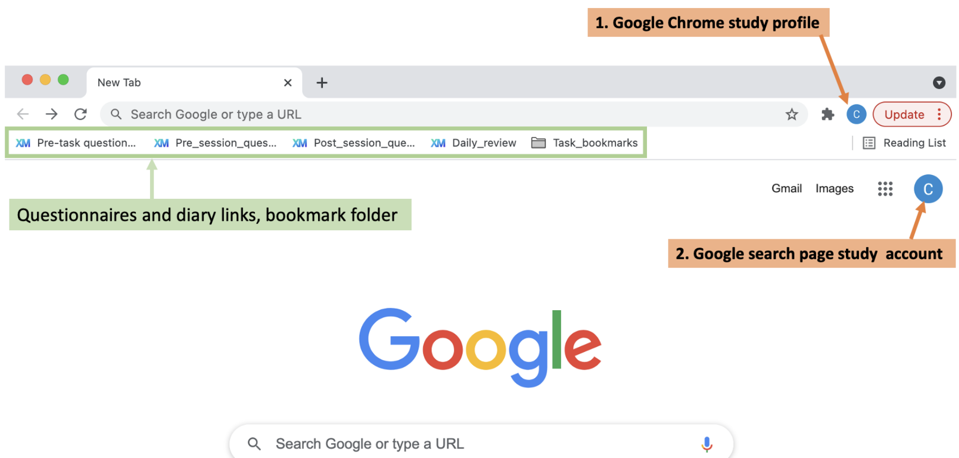


Figure 4.3: Google study account login page for XSS diary study

There are multiple reasons for using Google Chrome as the browser for this study:

- Google is the most popular browser in the United States since December 2013 (Figure 4.4)[1]. Therefore, participants do not need to study/switch to a different web browser they may not be used or familiar with. Even for some student participants do not use Google Chrome as their daily web browser, it would be easy to help them set up one.

- For participants who use Google Chrome as their daily web browser, switching between different accounts is much more convenient for them compared to using a completely different browser.

[1]https://www.statista.com/statistics/272697/market-share-desktop-internet-browser-usa/(retrieved by March.13, 2021)

- Instead of using their personal account, using a separate Google study account could prevent participants from mixing their personal information with the study data. This can also help protect their personal information since the study account will only include their search history related to the task. And participants can remove any history from this account if they won't like to share it.

- Compared to other designs that asked participants to review their entire search history and pick up the ones related to specific interesting tasks, separating search history would reduce their workload.

- Using a Google study account set up by the researcher and allowing the sync of searching data, so that I could monitor participants' search behaviors on a daily basis, and export their search data remotely without asking participants to do extra work (e.g., downloading their search history and send to me).

- For the current study, I am not explicitly focusing on analyzing particular search behaviors besides queries (e.g., clicks, dwell time, mouse movements), so there is no need for using a browser extension to record all participant's search behaviors for this study.

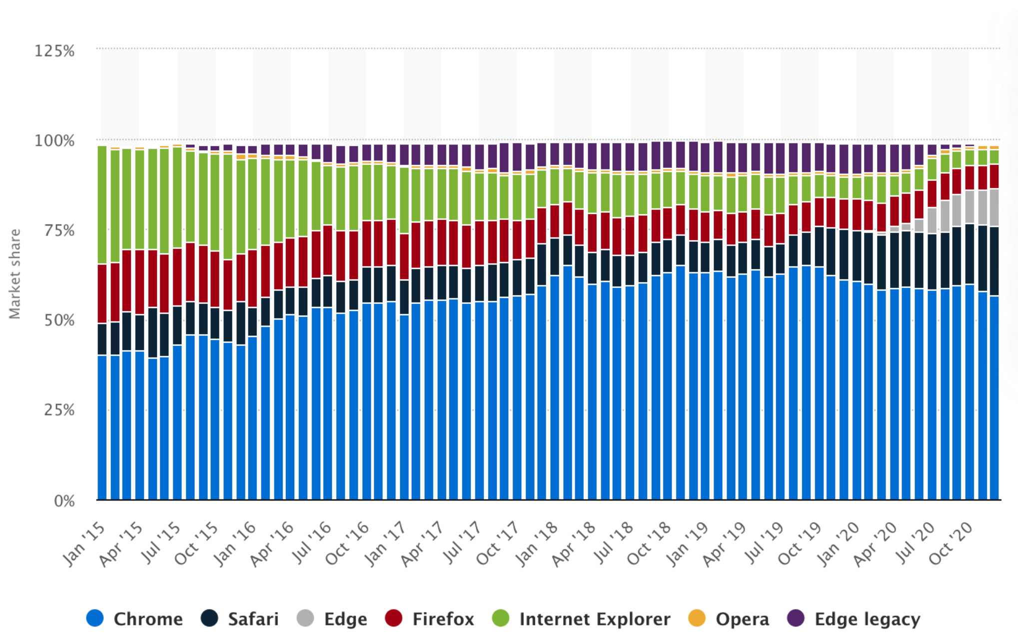


Figure 4.4: Market share held by leading desktop internet browsers in the United States from January 2015 to December 2020

I set up a new Google account associated with a unique password for each participant before the introduction meeting. All detailed steps are in described in Section 4.5 and Appendix C.

4.5 Initial introduction meeting

The initial introduction meeting (see Appendix C) is a one-hour online Zoom meeting. During the meeting, I talked with a scheduled participant about system set-up, and study protocol and answer any questions. I arranged the meeting based on the convenient time provided by the participants in their reply to the recruiting survey. A unique Zoom meeting link will be sent to the participants via email (see Appendix B).

During the initial introduction meeting, I asked the participant first to sign a consent form. Then I guided them to set up the Google study profile and sign in with the assigned Google study account.

After the system was set up, I first asked the participant to fill out the pre-task questionnaire using the first link on their Google study account page. Since this questionnaire only needs to be filled once, I removed the link after they completed it. Then I walked the participant through the requirements for the diary study, including how to switch back and forth between the Google study account and their personal account, search for information as usual and save bookmarks to the folder, how and when to fill the pre- and post-session questionnaires, and the daily diary review, as well as to review and clean their search history by the end of the day. To help the participant better understand the process, I asked the participant to do a practice task searching information for about buying a 4k television and practice with all the questionnaire/diary links. During this process, they could ask any questions about the task process or questions regarding questionnaires or diaries.

The introduction meeting aimed to help participants understand the study requirements, relevant taxonomy, and how to fill the questions in their questionnaires and daily diary review.

4.6 Questionnaires, daily review diary, and interview

I used mixed methods to collect data during the diary study, including the pre-task questionnaire, pre-/post-session questionnaires, daily review diary, and retrospective interviews. All the questionnaires and diary are semi-structured, including open-response questions, multiple-choice questions, and 7-point Likert-style questions. The interview is semi-structured. Next, I will discuss the design of the questions and requirements in each portion of these methods separately.

4.6.1 Pre-task questionnaire

I asked participants to fill out the *pre-task questionnaire* during the introduction meeting session. At the recruiting stage, participants already answered some basic questions about the task they would do if they were recruited. The screening questions (see Appendix A.2) told us about the participant's task topic, the final outcome, and the time range for completing the task, which can make sure that their tasks fit our study. The *pre-task questionnaire* aimed to gain more details about different facets of participants' tasks. And their perceptions of the task should not be affected by specific search activities or completed steps for the selected task.

Previous studies have investigated many different aspects related to XSS, for instance, task topics, searchers' motivations, and purposes, task complexity, time sensitivity, search sessions, etc.(e.g., Morris et al. (2008a), Agichtein et al. (2012), Spink (1996), Sellen et al. (2002), MacKay and Watters (2008a)). However, many of these data were collected either: (1) based on participants recalling responses for survey data (e.g., Morris et al. (2008a), Spink (1996), Sellen et al. (2002)), or (2) collecting data as participants search for/work on their XSS tasks (e.g., MacKay and Watters (2008a)), or (3) based on the coding of search queries logs without contextual information from the searchers' end (e.g., Agichtein et al. (2012)). Although these studies provide detailed descriptions of XSS tasks, relatively few know about users' perceptions of the complex tasks before they start them. For instance, whether and how much people already know about the potential causes before they start to search, and if they know some of the reasons, whether these reasons will change and how as their tasks spread out later. Understanding their knowledge of the task before the search can help us gain more insights about what aspects and to what extent those aspects lead participants to work overtime. It can also help us gain more knowledge about the sources of tasks that cause people to search multiple times. Therefore, the pre-task questionnaire was designed to focus on users' perceptions of task facets, familiarity, and difficulty.

The pre-task questionnaire (Table 4.1) includes fourteen questions grouped in three categories: The first section is about task characteristics, which includes seven questions (Q1 to Q7) focusing on task topic (Q1), task process (Q2), sources of tasks (Q3), collaboration (Q4), expect time to spend on the task (Q5), task goals (Q6), and needed information (Q7). These questions are developed based on Li and Belkin (2008)'s classification of task facets. And it also covers the questions asked by other previous XSS studies. The second group of questions and the third group of questions are about users' subjective perceptions of the task. Q8 to Q10 specifically ask participants about their familiarity with the task topic knowledge, required skills to complete the task, and solutions for the task. These three questions would tell us to what extent users were familiar with the task before they start. The third group of questions focuses on participants' subjective perceptions of the task difficulty about different aspects related to the entire task difficulty (Q11), search difficulty (Q12), monitor progress difficulty (Q13), and synthesize difficulty (Q14). Questions of both group 2 and group 3 are adapted from previous laboratory studies investigating the effect of task complexity and difficulty on users' information search behavior (e.g., Arguello (2014), Choi et al. (2019)).

Table 4.1: Pre-task questionnaire

Theme	Question Description	Work task facet
Task characteristics	**Q1.** Please describe what your task is about in detail.	Task topic
	Q2. Please describe how you will work to complete this task.	Task process
	Q3. Which of the following options can best describe why you undertake this task?	Source of task
	Q4. Will you require anyone's assistance to complete the task??	Interdependence
	Q5. How much time do you expect to spend to complete this task?	
	Q6. What are your goals for this task? Please list the main goals and any sub-goals.	Goal
	Q7. What information do you want to find for this task? If you wanted to find multiple pieces of information, please list them all.	Expected info.needs
Task familiarity	**Q8.** How much do you already know about the task topic?	Topic knowledge
	Q9. How much do you already know about the skills needed to complete this task?	Procedure knowledge
	Q10. How much do you already know about your solution to this task?	Outcome knowledge
Task difficulty (Whole task)	**Q11.** I think it will be difficult to complete this task.	Overall difficulty
	Q12. I think it will be difficult to search for information to complete this task.	Search difficulty
	Q13. I think it will be difficult to decide when I have enough information.	Progress monitoring difficulty
	Q14. I think it will be difficult to integrate the information I find to complete this task.	Synthesize difficulty

4.6.2 Search session questionnaires

An event-contingent diary study usually requires participants to make a diary entry whenever a specific type of event happens. For the current diary study, the participant was required to fill out a *pre-session questionnaire* whenever they would start a search session and complete a *post-session questionnaire* after they finish the session. In cases where they cannot fill out the questionnaires right away after the search session, they were encouraged to complete the questionnaires at their earliest convenience (e.g., by the end of the day). This can help participants to record what happens and how they perceive the event process immediately to reduce the effect of biased recall, such as misremembering or forgetting.

Pre-session questionnaire. The primary goal of the pre-session questionnaire is to gain knowledge about the complexity of the user's search session. There are various approaches to define task complexity, for instance, identifying the number of subtasks or steps involved in the task (Campbell, 1988; Singer et al., 2013; J. Liu & Belkin, 2010)), determining the *indeterminate nature* or *prior determinability* of the task process (Byström & Järvelin, 1995; Capra et al., 2018), or the needed query terms or operators to complete a task (Singer, Norbisrath, & Lewandowski, 2012), or designing tasks based on different types of cognitive activities that are required to complete a task (Kelly et al., 2015; Urgo et al., 2020).

However, many of these approaches are used in experimental conditions, meaning that researchers have to know the tasks well before asking participants to search for information for these tasks. Therefore, these approaches are not appropriate for the current study since (1) although we know the participant's overall task (by screening and pre-task questionnaire), we do not know how they would work on the task during the diary study. Participants could work on any parts or sections over time: they can use any search items and decide how much they want to complete by one search session; (2) we can gain information about their prior knowledge about the task topic, task steps, or potential solutions or outcomes for the entire task (which could be potentially used as *whole task complexity*), it is unsure about how much variability of the levels of *prior determinability* can be reflected across different sessions. Given these considerations, I decided to adapt the *cognitive complexity* measurements to explore the session complexity for XSS.

Table 4.2: Pre-session questionnaire

Section	Question Description	Session facet
Cognitive complexity (Open-ended)	**Q1.** What activities did you do with your task before you started this search session?	Pre-search action
	Q2. What are your goals for this search session? Why did you start this search session?	Goal
Cognitive complexity (All goals)	**Q3.** From the list below, select **all** that apply to your goals for this session.	Cognitive complexity
	A: I need to find some specific factual information.	*Retrieve*
	B: I need to learn about important concepts and their definitions related to my task.	*Understand*
	C: I need to find instructions and follow them to solve some problems for my task.	*Apply*
	D: I need to find information that can help me differentiate between related ideas.	*Analyze*
	E: I need to find and evaluate different alternatives to make informed decisions for my task.	*Evaluate*
	F: I need to find some examples and inspiration to generate my own ideas.	*Create*
	G: Other (please specify)	
Cognitive complexity (Main goal)	**Q4.** From the list below, select **the best** that can describe your **main goal** for this session.	Cognitive complexity
	A: I need to find some specific factual information.	*Retrieve*
	B: I need to learn about important concepts and their definitions related to my task.	*Understand*
	C: I need to find instructions and follow them to solve some problems for my task.	*Apply*
	D: I need to find information that can help me differentiate between related ideas.	*Analyze*
	E: I need to find and evaluate different alternatives to make informed decisions for my task.	*Evaluate*
	F: I need to find some examples and inspiration to generate my own ideas.	*Create*

Previous work usually pre-defined the task complexity based on the cognitive activities that predictably will be involved in a search task (i.e., retrieve, understand, apply, analyze, evaluate, create, from lower-level complexity to higher-level complexity) (Kelly et al., 2015; Capra, Arguello, Crescenzi, & Vardell, 2015; Choi et al., 2019). Urgo et al. (2020) also designed tasks using the Anderson and Krathwohl (2001)'s learning taxonomy, and their study found that knowledge types have effects on users' cognitive processes. Tasks that involve certain types of knowledge (e.g., conceptual knowledge) could be more complex and require more search activities. In this XSS study, I adapted Urgo et al.'s (2020) questions about task complexity and modified them to help measure participants' perceptions about the cognitive activities that may be involved in the session they were about to start. In the *pre-session questionnaire*, participants first answered a question about what they did for the task before starting this session (Q1), and then they would describe the goals for the current session and briefly explain why they need to search for the session (Q2). Then, Q3 provides a list of descriptions of the types of information and cognitive activities that can describe what participants wanted to achieve in that session. They would answer two questions based on the same list: (1) select *all* options that can match their goals for the session (Q3); (2) select the *best option* that can match their *main goal*for the session (Table 4.2) (Q4).

By analyzing these questions, I gained knowledge about the cognitive complexity of individual sessions, and then I conducted further analysis of its effects on users' search behavior during XSS process.

Post-session questionnaire. Table 4.3 provides an overview of the post-session questionnaire. The majority purpose of it is to collect specific session resuming reasons and stopping reasons associated with a past search session, the information examples participants sought, and how they cognitively dealt with the found information.

Q3 and Q10 are open-response questions that ask the participant to describe the reasons that motivate them to search for this session and why they stopped the session in their own words. Q4 and Q11 are multiple-option questions that require participants to select the most important reason from lists that can match their self-described reasons in Q3 and Q10. Asking participants to describe the reasons in their own words first can allow us to explore any reasons that lead to XSS from participants' perspectives that may not be included in previous findings. Asking them to then select the reasons from existing lists helped us to find (1) whether the reasons described by participants all have corresponding reasons that were discovered by previous studies (e.g., Spink (1996), MacKay and Watters (2008a), Lin and Belkin (2005)), (2) after about two decades along with the advanced development of computer systems, have these reasons changed or not, and what the differences between our participants described reasons and reasons identified by previous research.

Q7 asks participants to provide 3 examples of information they found through the search process. For each information example, they needed to briefly describe the information, provide the URL of the information source, and describe how they think the information is helpful in completing their tasks. Based on their responses, I wanted to learn the different kinds of information the participant sought during the session. By analyzing their explanations about how they process the information, I wanted to learn the cognitive activities involved when interacting with the information in that session.

Participants must also evaluate their experienced *session difficulty* in this post-session questionnaire. Using the same four difficulty questions (Q8), I gained information about the user's experienced session difficulty for each session immediately after they completed it.

Table 4.3: Post-session questionnaire

Section	Question	Session facet
Session resumption	**Q1.** Please indicate your starting time and stopping time for this search session, like 11:20 am to 12:23 pm	Session length
	Q2. What did you want to achieve in this session?	Session goal
	Q3. Why did you start this search session?	Resuming reason
	Q4. Which one of the following descriptions can best match the most important reason for you to start this session?	MISE modes
	Q5. Describe your search session process in detail.	Search process
	Q6. How is the search direction of this session different from your previous search?	Subtopic change
Results evaluation	**Q7.** Provide 3 specific pieces of information you found during this session and explain why you think they are useful for your task:	Information type
	e.g., Aboutness; URL; how they think helpful or not.	Cognitive activity
Session difficulty	**Q8.**It was difficult to search for information in this session.	Search difficulty
	It was difficult to understand the information I found in this session.	Comprehensive difficulty
	It was difficult to integrate the information I found in this session.	synthesize difficulty
	Overall, this search session was difficult.	Overall difficulty
Within session action	**Q9.** Beyond searching online, what other activities did you do for your task during this search session?	Non-search action
Session stopping	**Q10.** What are the reasons that made you stop this search session? Please list all the reasons.	Stopping reasons
	Q11. Select the most important task-related reason below that made you stop this search session.	
	Q12. Did you encounter any other difficulties during this search session? If yes, please describe	Other difficulties
	Q13. What kind of help did you wish the search engine provided during this session?	Wished function
Post-search activity	**Q14.** What is your next step for this task after this search session?	Follow-up activity Cognitive activity

Besides information search motivation and search activities during the session, I also asked participants to provide information about their search direction change (Q6), and descriptions of non-search activities during the session (Q9). During XSS, non-search activities are an important component of the broader work task that motivates the search. Although participants are required to fill out a daily dairy review (which will be discussed in the next section) in which they will be required to describe what they have done for the task during that day, the daily review diary may not catch all relevant activities, especially if participants conducted multiple sessions within one day, their descriptions about other task activities may be all mixed. Furthermore, their descriptions of the non-search activities they took during the session while they were searching could tell us about how participants dealt with the found information, did they apply the information immediately or do they set clear boundaries between searching and applying activities. The other two questions are about the challenges they encountered during this process (Q12)–these challenges may not completely prevent them from searching but make the search difficult where participants may want some help from the search system (Q13), as well as their plan for next step after the search (Q14). Together, these questions about participants' non-search activities can provide us with information about how they use the found information to help their task move forward.

4.6.3 Daily review diary

The daily review diary is a semi-structured diary form that a participant needs to fill out by the end of each day when they worked on their tasks. By "working", we mean any activities that are related to their task. Specifically, the daily review diary plays multiple roles: (1)reminding users' about the search session questionnaires (Q1), (2) collecting users' information-seeking behavior outside of search engines and the reasons (Q3, Q4), (3) collecting users' information use behavior (Q5), the challenges they met (Q6), the helps they wish the system can help with (Q7), (4) capture their evaluation of the whole task difficulty if it was the last day they work on the task (Q8, Q9). In addition, it also reminded participants to check the search history account to make sure there were no irrelevant search activities in that account (Q10) (Table 4.4).

Table 4.4: Daily review diary

Section	Question	Task facet
Reminding	**Q1.** Have you filled out the (pre/post) search session questionnaires for all your search sessions today?	Checkup
Work process	**Q2.** Please describe your today's work process for the task (include both search and non-search activities).	Task process
Off-line information search	**Q3.** Which of the following sources did you use to look for information related to this task today?	Info. sources
	Q4. What information were you looking for from the information source(s) you selected above, and why?	Off-line info type
Information use	**Q5.** How did the information you found online today help you with your task or not?	Cognitive activity
Challenges	**Q6.** What were the difficulties or problems you had when working on your task today?	Task difficulty
Wished help	**Q7.** What kind of help do you wish the computer or the search systems to provide for supporting your task activities today?	Task help
Task continuity	**Q8.** By now, are you finished your entire task or do you still plan to do more work on it?	Last diary review
Task difficulty (Whole task)	**Q9.** based on your information searching experience of the task, please indicate your level of agreement with the following statements:	Post-task difficulty
	Overall, it was difficult to complete this task.	Overall diff.
	It was difficult to search for information to complete this task.	Search diff.
	It was difficult to understand the information I found to complete this task.	Comprehensive diff.
	It was difficult to integrate the information I found to complete this task.	Synthesize diff.
	It was difficult to decide when I have enough information to complete this task.	Evaluation diff.
Clean search history	**Q10.** Please open and scan through your study account's search history today, and delete history links or bookmarks that are not relevant to the task or anything that you don't want to share with the research team.	

Answering questionnaires and recording diaries are not a part of their daily routine when they work on a specific task, especially at the beginning of a diary study. Repeatedly asking them to fill out the questionnaires and diaries is necessary to help participants get used to the study procedure. More importantly, it can help reduce the risk of losing data.

4.6.4 Post-task interview

I recruited 15 people from the diary study participants to conduct a retrospective interview. As an exploratory study, the interview focused on exploring various participants' perceptions about their XSS search stages and probing for more details about their session resuming and stopping reasons and the challenges and wished supports they mentioned in the questionnaires diaries. These analyses were qualitative instead of quantitative. And I adapted the *theoretical sampling* process for the *grounded theory methodology*.

Theoretical sampling is well-known for its feature that seeking additional data as the data collection progresses (Charmaz, 2000). As the diary study progressed with multiple participants may work simultaneously, I selected participants to interview as soon as they complete the task. The initial criteria were that the participants provide complete records of the questionnaires, diaries, and search history. After a couple of interviews, I focused on participants whose tasks and responses might be significantly different from those I interviewed. For instance, I chose those whose tasks lasted much longer than other tasks, tasks that contained much more sessions within the same days versus those who reported only one session per day, and tasks that contained more parts or steps. I selected the interviewees from the participants that could provide a richer and deeper understanding of XSS activities (e.g., a participant who provided a lot of details in their questionnaire might be more willing to tell their stories during the interview than those who just left a few sentences). Grounded theory studies suggested a different number of interviews for different types of research: For theory developing, Morse (1994) suggested 30 to 50 interviews, while Creswell and Poth (2007) suggested 20 to 30, but phenomenological studies need much fewer interviews, Creswell and Poth (2007) recommends 5 to 25 and Morse (1994) suggests at least six. Therefore, 15 interviews were enough for this study.

The interviews were scheduled within 1 to 2 workdays after a participant completed the diary study. The retrospective interview is a common research method used in diary studies. By talking directly with participants during the interviews, My main goals of the interview were to (1) gain a greater understanding of participants' XSS experience and the contexts that influence their XSS behavior, (2) help catch up with potential missing data during the diary recording period, and (3) clarify search log data.

To avoid the challenge of inaccurate recalling and response, I provided participants an overview profile of their search data (i.e., session time, used queries), responses extracted from their search session questionnaires, and daily review diaries to help them recall the details of their search process. Here I provided an example from a pilot participant's search log data (Figure 4.5 [2]):

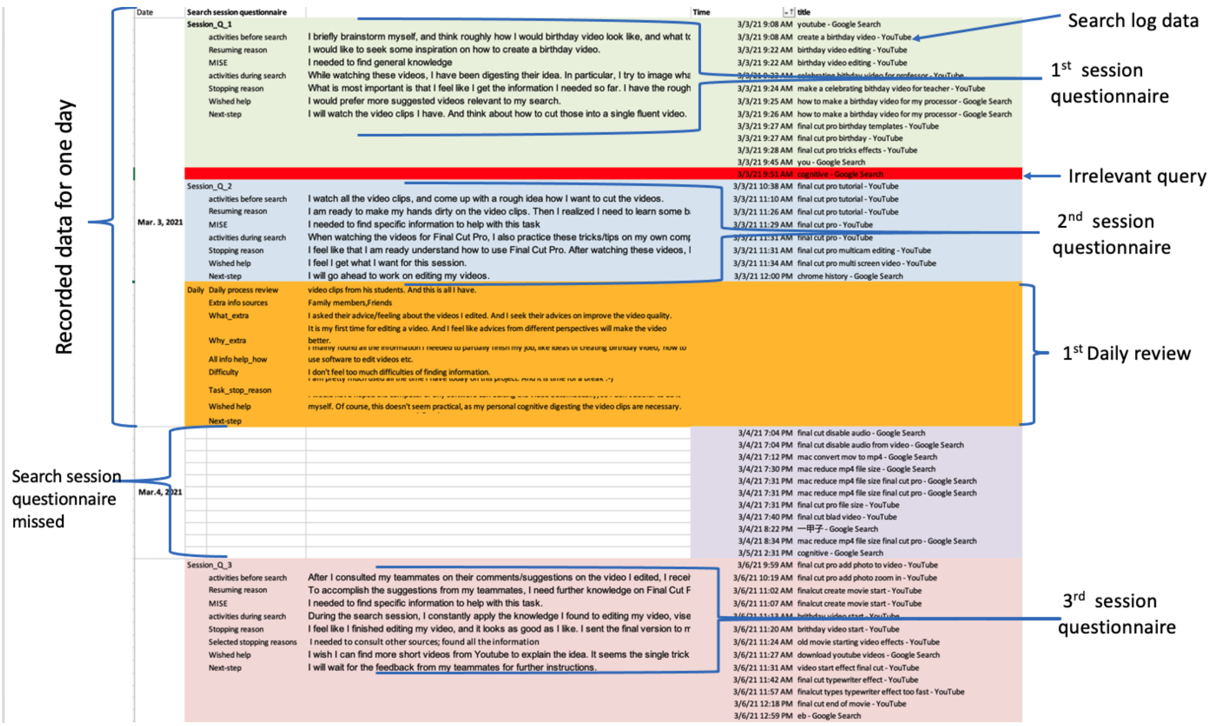


Figure 4.5: A profile of visualized summary of a participant's search-session questionnaires, daily review diaries, and search history.

The interview questions are semi-structured: While there are seven main questions I asked each participant, some of them were followed with more probing questions for more detail on a particular matter (see below, and Appendix N). Each interview lasted about 30 minutes and was conducted remotely using Zoom. The audio and screen were recorded for later analysis.

Q1: So far, what are the final outcomes of your task?

Q2: Here is an Excel spreadsheet of your search sessions and questionnaire and diary responses, thinking about the different stages as you work on this task, how do you differ the stages of

[2]This pilot participants' data don't include the pre-session questionnaire since the questionnaire is added later.

your search process and why?

Q3. Please scan your search queries for each session, briefly describe what kind of information you were looking for for that session and why?

Q4: During the search process, was there any unexpected information that you did not think of before you start this task? Can you describe how did you notice the unexpected information? And how did the information help your task? [Show the spreadsheet and point to the queries and useful info logged]

Q5: During the task process, what kind of challenges, or difficulties have you met? How did you overcome the difficulties?

Q6: In your search session questionnaire responses and daily diaries, you mentioned the following extra assistance that you wish the system can provide help with, can you explain more about them?

Q7: Comparing the recorded search session diaries with search log data, in cases when people searched but did not record a session questionnaire, point them to the search queries and links, ask them to recall why they started that search and how the stopped.

These questions worked as a guide for the interview but were not completely fixed. I probed participants with more questions depending on their task processes and the responses they provide.

4.7 Search history data

In this diary study, I captured users' search history and the bookmarks they made during the study period. Since they were required to use the assigned Google study account to search for information for the selected XSS task and bookmark useful web pages during the search sessions, all relevant search history and bookmarks were automatically stored within the account. Each day, I sent a participant a reminder email asking them to review their search history and delete irrelevant or any links they don't want to share.

On the other hand, to encourage participants to keep as much as their original search histories, I also made it clear that participants who provide higher-quality data (including questionnaire responses and search history) would receive $20 bonus as an award. The search history would record the timestamps when participants conducted a search, the search engines they used, the queries they used, and the links that they clicked. The primary purpose of collecting search history data was to gain the complete sequence of queries participants used for their tasks, and the cluster of queries for each search session over time (the right column in Figure 4.5). The search history log also provided us with information about sessions for which participants missed the search session questionnaire. For instance, in the pilot search history in Figure 4.5, the participant missed the search session questionnaire for the first search session on day 2 (highlighted in light purple). Therefore, during the interview, I showed the participant the missed session and asked them if they recall the motivations for that specific search session. In addition, by going through the search history, I also asked participants to explain more about some uncertain query entries (e.g., some queries that seem irrelevant will be highlighted in red, see Figure 4.5).

Participants were asked to bookmark web pages that contain helpful information for their tasks and save the links to the task bookmark folder. Also, in the search session questionnaire, they needed to list the useful information they found during the search session with the associated URLs. If participants retrieved all the useful information from the bookmarked pages, we could understand what information, where, and how it can help participants with their tasks. Suppose there were some information examples in the questionnaire from web pages they did not bookmark. In that case, we could get additional information that participants may forget to bookmark during the search session, in which case, the bookmarks can be used as complementary information that participants were interested in during a specific search session.

4.8 Data analysis

4.8.1 Qualitative analysis

I collected both quantitative data and qualitative data. To address the research questions, the analysis of data for this dissertation focuses on (1) the characteristic of the cross-session work tasks (XSWTs) and search sessions, (2) the reasons that caused participants to start and stop a search session, (3) the types of information examples participants provided, and (4) the types of cognitive activities regarding how participants processed the information examples, (5) the perceived difficulties for the task and sessions, and (6) the challenges they encountered that were caused by searching across sessions.

The quantitative data (e.g., task timeline, number of task doers, session numbers, pre-/post-session difficulty) were extracted from the questionnaires and the observation of participants' search processes based on the recorded search history from the Google study accounts. To analyze the qualitative data, I used both deductive and inductive coding methods.

The codes (or categories) and their definitions used in deductive coding are determined by existing theory, previous research, research questions, or a preliminary scanning of the text (Hsieh & Shannon, 2005; Crabtree & Miller, 1992). Researchers can use predetermined codes to analyze the target content. Any text that cannot be coded should be analyzed using new codes or categories (Hsieh & Shannon, 2005). Deductive coding can help researchers validate and extend existing theories or frameworks and focus on the research questions (Hsieh & Shannon, 2005)

In this study, I used the deductive coding method to analyze work task complexity, session resuming and stopping reasons, the types of information, and corresponding cognitive activities. Anderson and Krathwohl's (2001) taxonomy of cognitive complexity has long been used to design interactive information retrieval tasks (e.g., (Jansen et al., 2009; Kelly et al., 2015; Urgo, Arguello, & Capra, 2019)) which show that it is an effective way of categorizing tasks in lab-controlled studies. we used the same taxonomy to analyze the complexity levels of real-life work tasks in our preliminary AMT study and it worked well–we found all categories of the cognitive activities people conducted to complete their cross-session work tasks. (2) the information type categories were developed based on both Anderson and Krathwohl's (2001) taxonomy and the analysis of Choi et al.'s (2019) experimental study, which was also used by us for analyzing AMT XSS study (Li et al., 2021a). (3)Previous research (Spink et al., 2002; Li et al., 2020a) shows that some of the session resuming and stopping reasons included in Lin and Belkin's (2005) MISE model can be found from real-life tasks but under certain conditions (e.g., mediated library database search, provided a list of reasons for participants to recognize their session reasons). Using these categories to analyze users' own descriptions of the reasons can help us validate whether these reasons exist in real life and to what extent their meaning can match the previous definition. More importantly, it also can help us to find whether there are new categories and what their relationships with the existing codes are (e.g., are the new codes not included in the original template or they are some subcategories of the existing codes).

Based on Anderson and Krathwohl's (2001) taxonomy of cognitive complexity, I developed the codebook to analyze participants' work task complexity and the complexity level of their cognitive activities involved when dealing with the found information (Table 4.5). In terms of information types, I followed the template that was developed by Choi et al. (2019), which we also used to analyze information types in the preliminary AMT cross-session search study (Li et al., 2021a), except the metacognitive information type for which we did not find any examples from the preliminary scanning (Table 4.6). The codebook of resuming reasons kept the original definitions that were modified based on Lin and Belkin's (2005) MISE model (Table 4.7). As to the stopping reasons, the codebook 4.8 includes categories that were developed based on (1) Lin and Belkin's (2005) categories of two types of interruptions for multi-session search, (2) MacKay and Watters's (2008b) observations of session stopping reasons from their diary study, and (3) a preliminary scanning of participants' responses.

Table 4.5: Scheme of task complexity from

Complexity level	**Definition**	**Cognitive activity examples**
Retrieve	look for and locate specific information to fulfill the task requirements or information needs during a session	retrieve, recognize, identify
Understand	make sense and construct meaning from the retrieved information	interpret, exemplify, classify, summarize, explain
Apply	use the found information in the given situation to solve the problem, execute the steps by following what have learned from the found information	execute, carry out, implement
Analyze	mentally break down the found information into parts and determine how the parts relate to one another and to the overall problem or the task purpose	differentiate, organize, attribute
Evaluate	make judgments about the found information based on certain criteria and standards related to their own task	check, monitor, test, coordinate, critique
Create	put the found information together to form something new, coherent, such as a new pattern, or structure as a functional whole for the task	generate, plan, produce, design, construct

Table 4.6: Scheme of information types from taxonomy

Info. type	**Definition**	**Examples**
Factual info.	basic elements and specific details participants need to know about their task topic and to solve their problem	facts, numbers, terms, background information, introduction about the topic area, information about resources
Conceptual info.	information about interrelationships between two or more basic elements, or the relationships between specific elements/factual information within the task topic domain, or information about how something works (function)	theories, principles, categories, models, structures
Procedural info.	information about specific skills, methods, and techniques about how to do something, methods for inquiring, and criteria for using the skills or techniques	subject-specific skills, techniques, and methods, algorithms, criteria for determining when to use appropriate procedures
Opinions	perceptions made by other people, including objective opinions like scientific research or subject opinions based on personal experiences. Could be an opinion about the factual/conceptual/procedure information	conclusion, judgment, "other people's view"

Table 4.7: Scheme of session resuming (starting) reasons for cross-session search from

Resuming reason	**Definition**
Transmuting	task requirements were not clear and participants need more information about the topic, background, or how to complete the task
Spawning	participants need to find information for subconcepts or subtasks of the task
Rolling back	the information found previously did not work and participants need to find alternative solutions
Re-finding	participants need to search again to find information they had seen previously
Unanswered	participants want to continue a previous search that stopped without finding satisfactory information
Updated/cultivated	participants wanted to find updated (e.g., the most recent)information related to the task
Anticipated	participants did not have a specific goal for the session but looked for some general information that can help their task move forward (the information might be useful for the task in the future).

Table 4.8: Scheme of session stopping reasons

Stopping reason	**Definition**
Goal achieved	participants had specific information needs for the session and the search goal was fulfilled so that they wanted to move on to a different step/subgoal of the task;
Satisfied	participants stopped searching after they processed (e.g., assimilated, evaluated, synthesized) the information they found through the current session; or they felt that they gained enough information for the session or come to a natural stopping point. The information they found was more general than specific and may cover different aspects of the task
Cannot find the needed information	participants noticed a mismatch between the retrieved information and what they wanted to find for the task and stop searching (e.g., cannot find the needed information; inefficient query, rare information resources)
To validate the found information	participants stopped and wanted to make sure what they found was trustworthy and appropriate for their task
To consult other sources	participants stopped searching to consult with other information resources for further/extra information to be able to continue for the next step, including waiting for others' responses
Interruption	any non-task related causes that interrupt participants' searching, such as running out of time, mental fatigue, distractions, system errors, or being disrupted by others, etc.

To ensure the reliability of the coding, I invited another coder to analyze the data together with me. For all codes that involved counting and quantitative analysis, each coder independently coded all data, and the two coders then compared all codes and discussed all the differences to achieve agreements (Table 4.9).

Since the analysis of XSS challenges is more exploratory, I used thematic analysis/open coding that only involved one coder. Open coding was used for analyzing participants' motivation for cross-session search and the challenges caused by searching across sessions since there is limited research literature that focuses on these two aspects of cross-session search. Unlike deductive coding, open coding uses a bottom-up approach to identify the key ideas in the data. Following the six-step open coding workflow (Berg, 2001), I read and familiarized myself with transcripts of participants' oral responses from the fifteen interviews and noted the initial categories from the original data. I compared the codes to seek commonalities and differences among them. Then I sorted and grouped the similar categories into higher-level categories that reflect their commonalities. I will explain more about the commonalities in the corresponding sections in Chapter 5 together with the results.

Table 4.9: Intercoder agreement using Cohen's Kappa

Coding category	**Intercodder agreement** (CoHen's Kappa)
Session resuming reasons	0.723
Session stopping reasons	0.861
Session goal complexity	0.697
Information type	1
Cognitive activity	0.719
Work task complexity	0.747

4.8.2 Quantitative analysis

I used quantitative methods to analyze participants' responses to the multiple-choice questions in the pre-task questionnaires, pre-/post-session questionnaires.

In the pre-task questionnaire, participants were required to choose from the multi-option list about their (1) task source (Q3), (2) collaboration (Q4), (3)task time length (Q5), using a 7-point Likert-scale to describe their familiarity with task topic knowledge (Q8) and skills to complete the task (Q9), and knowledge of task solutions (Q10). They were also asked to evaluate the task difficulty (1–7 points), which measures the expected task difficulty levels. Similarly, their experienced task difficulty for each search session was also measured during the post-session questionnaire, and their experienced post-task difficulty was measured in the last daily review diary using the 1–7 Likert scale. The results of these questions were categorical (both ordinal and nominal) and discrete quantitative data.

I summarized the results of these quantitative data using descriptive statistics, like the frequency of different types of task sources, the distribution of session numbers across different types of tasks, the average length of sessions, and queries. By doing so, we gained an overview of the distributions of different XSS task features.

I used statistical methods such as Chi-Square and Repeated Measures ANOVA tests to investigate correlations among a range of variables, including task complexity, session goal complexity, reasons for session resumption and cessation, types of information, cognitive activities, and the difficulties of tasks or sessions. Alongside these, I integrated descriptive statistics to offer supplementary insights into trends not fully unearthed by the existing statistical analysis. This approach proved particularly useful when statistical significance was not established in the tests, yet intriguing trends were observable within these relationships.

4.9 Study risks and strategies

There were some risks associated with this study design: One possible risk was that a participant may forget to switch to the study account to search for information for the reported task, and we would miss some data. Missing data or failing to record a diary is typical for diary studies due to the absence of the moderator. Previous studies reported cases when participants forget to carry the paper diary book with them or choose not to record a diary at the moment of the event due to some inconvenience. In our case, this could be caused by forgetting to switch to the *Google study account*, or a participant forgot to fill the pre- or post-search session questionnaire when they conduct a search session. To minimize the chances of missing recording, I addressed the key points (e.g., switching to the study account, when to fill out a search session questionnaire, and when to fill out a daily diary review). Each day, I sent a reminder email to participants about filling out the questionnaires and daily review diary during the study period. If somebody forgot to record a search session diary, this email helped remind them to catch up on it later when their memory is still clear, although it may not be as fresh as when the search just happened. Besides, questions in the post-session questionnaire required them to list a couple of examples of helpful information along with the URLs of the information sources. Questions in the daily review diary asking them about useful information they found during the day can also help us catch up with some helpful information about their search behaviors even if we lose the search data. Additionally, after participants complete their study, I extracted their search history, which include specific timestamps, queries, and search links, to compare with the sessions' times reported by participants , so that we can distinguish search sessions from the history but are not associated with a search session questionnaire. If the participant attends the post-task interview, I asked them questions about those sessions.

Another risk of a diary study was that some participants may withdraw in the middle of the study process. This happened in Zhang et al.(2020)'s diary study as well, which is another common issue for diary studies. It could be because participants find that they don't want to continue the task before it completes. Or they maybe stop participating in the project if it is assigned by somebody else (e.g., a supervisor). Or they could quit just because they are tired of the study itself. In any of these cases, participants were allowed to withdraw from the study at any time. And they were paid based on the amount of work they completed for the study (i.e., $2 for each complete questionnaire/diary). Another way for motivating participants to keep with the study is that besides the initial 50 dollars, we awarded participants with an extra 20 dollars for those who provided higher-quality diary recordings (e.g., fully recorded search session questionnaires, and daily diaries, complete search log data that matches with the questionnaires, responses with rich details, etc.). And participants who were willing to join the post-task interview received another 20 dollars as compensation for their participation. As a result, we recruited 29 participants and 25 of them (86%) completed the study.

4.10 Pilot test

To make sure participants understand the study procedure and how to fill the questionnaire and record daily review diaries, I consulted multiple Ph.D. students for their comments about questionnaire content and conducted several rounds of pilot tests through the design process.

The pilot participant followed the instruction to set up and use the assigned Google study account to search for their tasks. Then he filled out the pre-task questionnaire and completed post-session questionnaires for four sessions across two days, and two daily review diaries based on his own work pace. The pilot participant also provided feedback on unclear questions, and we modified those questions based on his suggestions.

The pilot test shows that the pre-task questionnaire took about 8 to 10 minutes to complete. The post-session questionnaire took 9 to 14 minutes and 17 to 20 minutes for the daily review diaries. The pilot participant did report that he filled out some questionnaires and the daily diary report later; one potential reason may be that I did not send them reminder emails during the diary recording period (for a courtesy reason). It suggests that the daily reminding email is necessary when running the actual study.

I used some of the pilot participant's data as examples in an earlier session (see the response summary for the interview, Figure 4.5 in Section 4.6.4).

4.11 Data collection

In this section, I provide an overview of the diary study data collection and summarized the characteristics of our participants, and the characteristics of their cross-session work tasks based on the *pre-task questionnaire* responses.

4.11.1 Participants characteristics

In the *pre-task questionnaire*, we asked participants to provide responses for a set of demographic questions, including age, gender, the highest degree or level of school they have completed, their current status at UNC, and their occupation or major.

Among the 25 participants, there are 15 Females, 8 males, 1 nonbinary, and 1 did not report. Their age ranges from 18 to 62 (M=29.2, SD=12.66). In terms of their educational background, there are 9 with bachelor's degrees, 6 with graduate degrees, 9 with High school degrees or equivalent (for undergraduate students), and 1 with an associate degree. Regarding their current status at UNC, there are 9 undergraduate students, 7 graduate students, and 9 full-time non-student employees. Their occupations include students from the library and information science, Art department, school of public health, psychology, political science, mass communications, librarians, engineers, human resource staff, research assistants, and therapy professors.

4.11.2 Overview of diary study data

The diary study started on August 16 and ended on October 15, 2021. We recruited participants using university-wide and departmental mailing lists that included faculty, staff, and undergraduate and graduate students. The recruiting email included a screening questionnaire and outlined the requirements for participation: 1) the participant needed to have a task they planned to work on, 2) they have not started the task yet, 3) they know that they will need to search for information to complete the task, 4) they plan to spend several days to complete the task, and 5) they anticipate needing to do Internet searches across different times or days as part of completing the task. These criteria helped us focus our data collection on *expected* cross-session searches from participants who planned to work on a task that they knew would require multiple information search sessions online.

People who were interested in the study filled out the screening questionnaire and submitted it through Qualtrics. During the time from August 17, 2021, when the first round of recruitment emails was sent out, to October 6, 2021, we received 68 submissions of the screening questionnaires and selected 32 and invited them to participate in the diary study. Besides the selecting criteria listed above, we also considered the following aspects when deciding which participants to be invited: 1) include as many task types as possible, for instance, not all tasks should be school assignments; 2) exclude tasks that collect information from a one-single resource for maintenance purpose or resources rather than the open-web, for instance, checking stock websites to monitoring stock price changing, filling out materials to apply for a passport; and 3) exclude submissions that did not provide enough details about their task.

The purpose of these additional criteria is to help us identify a wide variety of work tasks that require cross-session searches in people's daily work and life. Excluding tasks that mainly focus on checking updated information from a single information resource or do not need to find information from the open web provides us with more opportunities to observe users' interaction with different information resources when they are searching across sessions over time. In addition, we expected that participants who provided rich details in the screening questionnaires are more willing to provide detailed information about their search process during the diary study than those who did not.

Table 4.10: An overview of data collected

P_ID	# of recorded sess.	# of pre_task Q	# of pre_sess. Q	# of post_sess. Q	# of daily diary
P101	4	1	4	4	3
P102	7	1	7	7	5
P103	6	1	6	6	5
P104	3	1	3	3	3
P105	3	1	3	3	3
P106	5	1	5	5	4
P107	4	1	5	4	5
P108	3	1	4	3	4
P110	4	1	4	4	3
P112	3	1	3	3	2
P113	6	1	6	6	4
P114	6	1	6	6	5
P116	3	1	3	3	3
P117	5	1	5	5	4
P118	3	1	3	3	3
P120	5	1	5	5	6
P121	3	1	3	3	3
P123	4	1	4	4	3
P124	3	1	3	3	2
P125	3	1	3	3	2
P126	3	1	3	3	3
P127	3	1	3	3	3
P128	6	1	6	6	6
P129	3	1	3	3	3
P131	3	1	3	3	4
Total	101	25	103	101	91

Among the 31 selected participants, 2 of them withdraw from the study immediately before it started (P109, P130), 4 participants did not complete the study and dropped out (P111, P115, P119, P122). In total, 25 participants completed the diary study. We also interviewed 15 of the 25 participants who completed the diary studies. An overview of the data collected is summarized in Table 4.10.

CHAPTER 5

Results

In this chapter, we show our observations and findings guided by the research questions. First, we present the observations of our participants and the characteristics of XSS tasks, such as participants' characteristics, task topics, task sources, timeline, etc.

5.1 Cross-session work tasks

5.1.1 Work task overview

In this study, we refer to tasks that cause people to search across multiple sessions as *cross-session work tasks* (XSWTs). In the *pre-task questionnaire*, we asked participants to provide information about their XSWTs by answering the following questions: 1) describe in detail the work task that they want to complete during the diary study, 2) describe the process of how they would work to complete the task, 3) the task resources (i.e., self-motivated, motivated through discussion with others, or assigned by others), 4) the information of task doers (i.e., the only person to complete the task, task doer but need to consult others, or work with other group members together), 5) the time needed to complete the task, 6) the goal for the task, 7) the information they want to find, 8) their levels of familiarity with the task topic, skills to complete the task, and knowledge about task solution, and 9) their perceived task difficulty. Next, I provide a summary of participants' responses to the above questions except for their answers for questions 6 (the goal of the task) and 7 (the information they wanted) since here we mainly focus on objective characteristics of the XSWTs to provide an overview of the context of participants cross-session search. We will discuss their task goals and information types later in other sections of this chapter.

Cross-session work task topics. From Table 5.1, we could find that our participants' tasks cover a wide range of topics, ranging from school assignments, research projects, and personal hobby development to self-training skills and projects for entertainment. Due to the limitation of participant recruitment, we do have multiple participants whose tasks are school assignments, but the tasks are not all about writing papers. For instance, one school assignment is to prepare a showing to present in class (P108), another is to write an essay on a topic based on the participant's own choice (P112), another essay assignment has requirements about the information resources to be used (P113), one is to prepare a group presentation (P125), and another one is about making a research outline on a specific topic (P126). From this perspective, our task samples are diverse that cover many different areas and work task types.

Table 5.1: An overview of participants' cross-session work task topics

P_ID	**Cross-session work task description**	**Shortened task name**
P101	Collect information to prepare for writing a film screenplay	Screenplay writing
P102	Look for a rental property for friends	Property renting
P103	Plan an graduation travel event	Graduation plan
P104	Look for summer theatrical opportunities	Summer internship
P105	Research potential guests for a brand new podcast channel	Podcast guest search
P106	Plan a solo trip to Disneyland in California	Trip plan to Disneyland
P107	Plan and Prepare for a baby shower	Baby shower
P108	Work on a showing about self-curiosity and discovery	Self-curiosity show
P110	Write a training report about sounds and audio	Sound/audio training report
P112	Write an essay about a public issue	Public issue essay
P113	Write a class paper on historical monuments	Historical monuments paper
P114	Learn how to conduct digital art as a hobby	Digital art learning
P116	Register a small consulting business company	Business register
P117	Plan a fall break trip (unknown destination) with friends	Trip planning with friends
P118	Recreate a Dual Screen (DS) terrarium sculpture	DS sculpture making
P120	Collect information about road safety issues for community members and stakeholders	Road safety proposal
P121	Shop multiple items for a triathlon training	Triathlon gear shopping
P123	Curate a [toy] collection composed by old discontinued sets	[toy] collection curate
P124	Complete a four-page policy memo based on selected topic	Policy memo essay
P125	Prepare a group presentation for a campaigns project on telemedicine and dentistry	Telemedicine group presentation
P126	Create a research outline about political policies about [two countries]	Political policy research outline
P127	Conduct research about a character to be played in a theater production	Theater character research
P128	Self-train how to do statistical sampling	Stats self training
P129	Complete a children book on parenting guide conversations	Children book writing
P131	Develop PPT presentation about non-profits organizations' strategies for advertising	NGO strategy slides

Table 5.2 provides more details about the 25 tasks including the task sources about where the task is from, participants' expected timeline for completing the task, and the number of tasks doers that would be involved.

Table 5.2: An overview of participants' cross-session tasks characteristics

P_ID	Task name	Task source	Expected time	Real time	Task doer(s)
P101	Screenplay writing	self-motivated	3-5 days	15 days	Work by self
P102	Property renting	by group discussion	>= 1 week	16 days	Need to consult others
P103	Graduation plan	self-motivated	3-5 days	12 days	Work by self
P104	Summer internship	self-motivated	>= 1 week	15 days	Work by self
P105	Podcast guest search	self-motivated	1-3 days	7 days	Work by self
P106	Trip plan to Disneyland	self-motivated	>= 1 week	21 days	Work by self
P107	Baby shower	by group discussion	>= 1 week	46 days	Need to consult others
P108	Self-curiosity show	assigned by others	3-5 days	12 days	Work by self
P110	Sound/audio training report	self-motivated	>= 1 week	22 days	Need to consult others
P112	Public issue essay	assigned by others	3-5 days	14 days	Work by self
P113	Historical monuments paper	assigned by others	3-5 days	14 days	Work by self
P114	Digital art learning	self-motivated	>= 1 week	34 days	Work by self
P116	Business register	self-motivated	3-5 days	15 days	Work by self
P117	Trip planning with friends	by group discussion	>= 1 week	16 days	Need to consult others
P118	DS sculpture making	self-motivated	>= 1 week	14 days	Work by self
P120	Road safety proposal	self-motivated	>= 1 week	35 days	Work by self
P121	Triathlon gear shopping	assigned by others	3-5 days	6 days	Work by self
P123	[toy] collection curate	self-motivated	1-3 days	3 days	Work by self
P124	Policy memo essay	assigned by others	>= 1 week	13 days	Work by self
P125	Telemedicine group presentation	assigned by others	>= 1 week	13 days	Need to work with others
P126	Political policy research outline	assigned by others	>= 1 week	28 days	Work by self
P127	Theater character research	assigned by others	>= 1 week	22 days	Need to consult others
P128	Stats self training	assigned by others	3-5 days	16 days	Work by self
P129	Children book writing	self-motivated	>= 1 week	6 days	Work by self
P131	NGO strategy slides	by group discussion	3-5 days	12 days	Work by self

Cross-session work task sources. The source of tasks is one of the frequently used task facets. It describes where tasks are from (Reid, 2000; Li & Belkin, 2008). Previous research classified task sources into two types: internally generated and externally generated/imposed and noted that the source of tasks could affect task representation (i.e., how the task is presented to the task doer), task doer's interpretation of the task, task complete process, and task outcome assessment (Reid, 2000). Li and Belkin (2008) described task source as a generic facet (i.e., external characteristics of tasks) summarized three task source types: internal, generated collaboration, and external assigned. We adapted these three task source types in the pre-task questionnaire to gain a general understanding of where our participants' cross-session work tasks were from.

Table 5.3 shows that there are twelve out of the 25 XSWTs (n=12, 48%) are self-motivated tasks, meaning that these tasks are generated by participants themselves based on their own information needs or personal context. Nine XSWTs (n=9, 36%) were assigned by others, usually somebody at a higher level of authority than the participants. Four XSWTs (n=4, 16%) were motivated through discussion with a group of people. And these tasks often involve collaboration among the group members to complete.

Table 5.3: Cross-session work task (XSWT) motivation sources

Task motivation sources	**# of XSWTs**
Self-motivated tasks	12
Assigned by others	9
Motivated by group discussion	4
Total	25

Cross-session work task timeline. Table 5.4 summarizes tasks based on participants' expected time to complete the XSWTs. As we can see, only two participants (n=2, 8%) expected to complete their tasks within a short-term period– by 1 to 3 days. Nine participants (n=9, 36%) expected to complete their tasks in a middle range of time from 3 to 5 days, and more than half of the participants (n=14, 56%) expected that their tasks would last more than one week (long-term tasks).

Table 5.4: Expected time for completing cross-session work tasks (XSWTs)

Expected task timeline	**# of XSWTs**
Short-term tasks (1-3 days)	2
Middle-term tasks (3-5 days)	9
Long-term tasks (>=1 week)	14
Total	25

Since our study required the participants who were interested in participating in the study need to complete the diary study within two weeks, we did not intend to recruit people whose tasks may take much longer or with an undetermined timeline. Therefore, we did not provide the "undetermined" choice as an option. But it should be noted that in real-life tasks, many work tasks that require users search for information over time could last longer than the time period we observed here. In addition, the summary we provided here only reflects the expectations that participants assumed they could complete the task. As we would see in later sections, many participants requested to extend the study time and took more time than they expected to complete the tasks. Some of them did not complete the entire task even with extended time. Instead, they stopped the task by the end by "partially completing their tasks" with multiple search sessions. We treated participants who "partially completed their tasks" as a completion of the diary study based on the pre-required criteria that 1) they provided questionnaire responses and search history for at least three search sessions, 2) they provided at least two daily review diaries, indicating that they worked on the task for at least at two different days.

By closely examining the task expected working time and the real-time (from the first search session time to the last search session time) (Tabel 5.5) participants spent on their tasks during the diary study, we can find that for the two tasks that needed to be completed within 3 days, only one participant who wanted to purchase a [toy] set (P123) completed the task as they planned by 3 days; the other participant (P105) who searched information about potential interviewee candidates for his podcast channel extended the task to 7 days and did not completely finish his task—they created a list through multiple search sessions and intended to continue to look for more information after the diary study and even after he started to interview people during the podcast program. On average, each participant in the short-term group spent 5 days on their tasks during the diary study period of time. For the 9 participants who planned to complete tasks within a middle-term (between 3 to 5 days), none of them completed the task within the expected time frame: the shortest one (P121) took 6 days, and the longest one took 16 days (P128). On average, each participant in this group spent 13 days working on the middle-term XSWTs during the dairy study (M=12.8, Median= 14, SD=2.97). For the 14 participants who planned that the task would take more than 1 week or month (long-term), the shortest time the participant spent was 6 days (P129) and the longest one (P107) spent 46 days working on the task. On average, each of the 14 participants in this group spent 22 days working on the XSWTs during the diary study period (M=21.5, Median=16, SD=10.83).

Table 5.5: Real-time participants spent working on their XSWTs during the diary study

Expected time	**1-3 days** (Short-term)	**3-5 days** (Middle-term)	**>= 1 week** (Long-term)
Real time on task	3	15	16
	7	12	15
		12	21
		14	46
		14	22
		15	34
		6	16
		16	14
		12	35
			13
			28
			22
			6
Mean	5	12.88	21.5
Median		14	16
SD		2.97	10.83

Overall, Table 5.4 and Table 5.5 show that within our sample group, many XSWTs are expected to take longer time (more than 3 days) than users' expectation to complete.

Cross-session work task-doer. Table 5.6 shows a summary of task-doers for the XSWTs in the diary study. Most of the participants (n=19, 76%) described that they would be the only person to complete the task. Five participants (n=5, 20%) reported that they need to consult others to complete the task, but they would work on themselves. And only one participant (n=1, 4%) said that they needed to work with others together to complete the task. However, similar to the task timelines, the description only reflects the task characteristics in this study, and the results here cannot be generalized to describe a broader scope of XSWTs in everyday work and life. Because both the recruitment process, criteria, and participants' choice of tasks could affect the results.

Table 5.6: No. of people working on individual cross-session work tasks (XSWT)

Expected task timeline	**# of XSWTs**
Work by oneself	19
Work by oneself but need to consult others	5
Need to work with others	1
Total	25

Cross-session work task complexity. In this section, we analyzed the cognitive complexity of the whole cross-session work task based on participants' descriptions of the task they wanted to accomplish during the diary study period.

We analyzed the complexity of the overall task based on participants' descriptions of the tasks they wanted to complete for the study (Table 5.7).

Table 5.7: Cross-session work tasks at different cognitive complexity levels

Task complexity level	**# of tasks**
Retrieve	0
Understand	3
Apply	3
Analyze	0
Evaluate	8
Create	11
Total	25

It needs to be pointed out that the complexity level of the task participants described in the pre-task description may differ from the work (i.e., the complexity of individual search sessions) they completed during the diary study because they may not complete the entire task as they expected. Consequently, the actual complexity level of the work they completed for the study might be changed. In some XSWTs (n=7), the level of task complexity is *more* complicated than the most complex cognitive activities that participants described in their post-session questionnaires. For example, one participant (P124) planned to complete a policy memo after selecting a specific topic that was classified at the *create* level. However, all three reported sessions for the task only contained cognitive activities like understanding, or applying/using the information but not actually working on the memo to create new ideas. In some other cases (n=5), the level of participants' task complexity is *less* complicated than the most complex cognitive example they provided in the post-session questionnaires. For instance, a participant (P110) searched for "fundamental sound and recording concepts to present these concepts to reporters at [a radio station] to help them with their equipment and recorded sound.." which task was labeled as "apply"; however, in his post_session questionnaires, he described many examples that involve evaluations by comparing different definitions for the same concept.

The examples participants provided in their post-session questionnaire may not completely represent the overall cognitive activities they conducted for the task, especially if a participant did not complete the expected task entirely during the diary study session. Our findings indicate we shall not assume that the complexity level of individual sessions during an XSS process is the same as the complexity level of the entire work task.

5.1.2 Cross-session search motivations

Before we dive into the specific starting/resuming and stopping causes of individual search sessions, we first analyzed users' motivation of XSS at the beginning–we want to understand why users expected to search for multiple sessions even before they started searching for their tasks. We analyzed participants' responses to the interview question "What made you decide to search for multiple sessions for this task?" Our qualitative analysis identified twelve themes that can be categorized into two groups: task-oriented motivations and cognition-oriented motivations.

It should be noted that, as we started to cite participants' responses from the questionnaires and interviews in the rest of this dissertation, we used P1XX to represent a participant, P1XX-interview after quoted content was used to represent the content came from participants' responses from the interview questions, P1XX-SX was used to mark content from participants' descriptions of their session process for session X, and P1XX-SX-Pre/Post were used for the quoted content from a participant's responses for the pre- or post-questionnaire of Session X.

Task-oriented motivations. In their interview responses, participants described expecting to conduct cross-session searches because of some of the characteristics of their work task and their perceptions of the task characteristics.

Multiple task components. Participants often mentioned that they expected their work task to involve multiple search sessions because the task contained different parts, or the participant perceived that it would be necessary to break down the task into multiple parts to be completed over time. For example, P110 worked on preparing a report for training purposes, he said:

> "there's so much out there that...it would be hours and hours of searching...there was so clear cut parts to each that I felt like that I could break it down into four sessions, and I could spend time on it" (P110-interview)

P125 described a group project for which she needed to complete searching on assigned subtopics before each group meeting. As she said:

> "because it was a group project...we schedule out what we have to bring to each group meeting. And so, for the first one, we just had to have the background, and then the next one, we had to have our slides done so that's when I started searching a lot more specifically. I knew that I could stretch it out, and I also knew, because I was like completely unfamiliar with it. And I'm just kind of unfamiliar with our client in general, though, that it would be good to have multiple sessions to be able to understand and like break it up a little bit instead of just like cramming all the night before [our meeting]" (P125-interview)

Ever changing task topic field. Participants described anticipating cross-session searches for tasks because the information about their topic on the Web is frequently changing. For instance, P104 was looking for an internship in the theater industry and noted,

> "there are always new [theatrical job opportunities] to be developed" and "people are still navigating the pandemic." (P104-interview)

P123 who was looking for "retired" [toy] sets described that:

> "the tertiary markets changed so much on a day-to-day basis...so every day you're going to see brand new listings." (P123-interview)

Participants also noted that some topic areas may have too much information to be explored within one sitting, while other areas may have rarely occurring information so participants need to keep searching over time.

Task completion expectation. Some participants had clear expectations that they would not be able to complete their work task in one sitting due to the scope of the task or their task timeline. For instance,

> "I knew I would not be able to complete a paper in one sitting, so I just did it over the course of two weeks and had to search multiple times because there was no way I was going to get all my information done in one setting. (P113-interview)

However, sometimes participants were uncertain about the task direction or the timeline. For example, P114 who was developing a personal hobby to learn digital art noted that

> "I didn't realize how much time would be taken up...I knew this was going to come up with different sessions." (P114-interview)

Cognition-oriented motivations. The second category of the motivations that lead people to search across sessions is related to participants' cognitive activities, meaning how they wanted to work on their tasks from the perspective of their mental capability and process. This category includes five types of motivations:

To understand the found information. Participants knew they would search for multiple sessions because they needed time to digest the information they found. For instance,

> “I think definitely across multiple sessions, because that time in between I think it’s important to be able to like soak in what you found in them, reevaluate what you need for the task. it helps me to like break it up into different sections so that it’s like organized in a certain way in my mind instead of like one big thing.” (P101-interview)

> “I knew that I could stretch it out because I was completely unfamiliar with [topic]... I split it up [so] I was able to think about it in chunks.” (P125-interview)

To explore new information and ideas. Participants pointed out that searching over time and leaving time in-between sessions can help them surface more information and ideas. For instance, P105 described that when looking for potential interviewees for their podcast channel, they “come across more names” as they continue to search.

> You know, as you continue to research, you come across more names, or when you’re no longer researching for that day like throughout the night or when you come across articles or you meet people and have conversations other names pop up or you’re searching on social media. You know you just discover more information, more people that you need to look into so. that’s why yes it’s always a constant..gathering of information. Even after you know the research...study is over, you know i’m still adding names and still booking interviews and you know it’s like I said it’s never ending.” (P105-interview)

Another participant (P108) described that:

> “I find it personally with multiple sessions is easier to complete the project, because with one big session sometimes there’s different things I want to explore, but I don’t have the time or... the idea I start with isn’t the one I always end with it was a later idea I was like oh that’s probably better than the initial so yeah always like I was like multiple sessions...yeah,it’s [my idea] deepening.” (P108-interview)

To process the found information for future use. Our participants also expected to search for multiple sessions since they would like to process (e.g., analyze, evaluate, compare) the found information and their ideas to help make adjustments or decisions about their next steps and task outcomes. For instance,

> "...searching for different individual elements... would be easier. Not to do it all at once, especially since a lot of them were working together, I felt like trying to do it all at once, would be too many decisions and trying to figure out every individual piece, at the same time, with no time in between, to try and narrow traces down or decide what I wanted to do." (P118-interview)

To organize and track the information and search process. Participants also anticipated doing cross-session searches because they thought it could facilitate their information organization and search process tracking. For instance,

> "if I am searching for too many things at once, opening too many tabs, then I forget things I wanted to search for and losing track of things." (P103-interview)

> "I have to break it down into parts because if I don't...my mind starts going to all these different places...it would be really hard for me to organize... a lot of information." (P110-interview)

To avoid mental fatigue and improve task productivity. Participants also preferred searching across multiple sessions since it can be difficult to concentrate for a long time on one thing. Searching across multiple sessions allows them to "turn [their] brain on and off" (P102) or "step away for a moment when things did not work" (P114). Other examples like:

> "if you spend so much time behind a computer doing the same thing your senses start to get compromised." (P110-interview)

5.1.3 Non-search task activities during the XSS process

During the process to complete their tasks, participants conducted various activities besides searching for online information. We defined these as *non-search activities*.

Our open-coded participants' responses in the pre_session questionnaire for the question "What activities did you do that related to your task before you start this search session?" and identified seven types of non-search task activities that participants conducted for their tasks, including:

Setting up/adjusting task scope and goals. Participants described that before they started a session or after the last session's search, they reflected on their understanding of the task goals and what they wanted to achieve either as a stage goal or for the entire task. For instance,

> "I've had two conversations with my sister about what she wants to do with her baby shower." (P107-S1-pre);

> "[before this session I] Reviewed my google doc sheet of notes to determine how I can complete my task." (P120-S5-pre)

Creating task work plans. Sometimes participants brainstormed or thought through how they wanted to work on the task and designed the process and activities such as how and what to search, where to find information, For instance,

> "I made a list of people and search terms that I wanted to focus on" (P105-S1-pre);

> "None. I didn't really prepare on a research plan other than what I had planned before like thinking I'll start with Wikipedia and search for videos. " (P128-S1-pre)

Creating task work plans and setting up/adjusting task scope and goals are related and sometimes overlapped with each other. But the difference is that work plans include more details about how to do something more than just describing what they wanted to achieve. For instance, "a list of people" (P105-S1-pre) can be treated as search goals, but "search terms" are more about the methods/strategies they would use rather than the purposes they wanted to achieve.

Working on the task using prior knowledge. Sometimes participants described working on their tasks based on their previous knowledge of the task topic and skills they already knew without any new searches. For instance,

> "[I] wrote a partial visual analysis of [topic] briefly from memory"(P113-S3-pre);

Applying the found information to address (sub)task problems. This describes such kind of activities often motivated them to conduct more searches as they identified some gaps that required further information. for instance,

> "I have read thoroughly the articles that I found in previous search sessions and drafted 3 slides about their content (VERY much a draft!). After that I realized that maybe I should search for some literature about public relations/public education strategies that I know some professions use, searching with those terms. This means changing my focus for this 3rd search session. (P131-S3-pre)

> "I booked my flight, which didn't involve any Google searches. I already had credit to use on [airline] so used their site directly" (P106-S3-pre)

Exchanging ideas with task collaborators. Tasks that involve collaboration with others often require the task participants to exchange ideas and their individual work progress with others through the XSS work process. For instance

> "I talked with my friend about the complexes I've found, and we narrowed it down to a list of complexes that sounded of most interest to her. We decided to eliminate the last complex I was going to add since she felt satisfied with what I had found so far." (P102-S7-pre)

> "I got in contact with all 6 sellers mentioned last time. Out of the 6 contacted, 4 saw my message and 3 responded. Out of those 3, only 1 said she might do shipping. I followed up with all 3 who responded, continuing an attempt to hash out some deals. " (P123-S2-pre)

It is interesting to be noticed that in the above example (P123-S2), the other part of the task participants were not actively involved in the task, as sellers they were a part of the task.

Consulting others for opinions and inspirations. This describes when task participants actively looked for suggestions and opinions from people that are not part of the task. For instance,

> "Last week I texted some of my friends to ask if they'd heard of the [name of fish], and, surprisingly, none of them had. I suppose that's a good thing if I'm going at it from the angle of bringing new content to their lives! I thought my friends who are into animals would've known about the richness of animal life there, but all the more to share with them in the future" (P101-S2-pre)

> "I had a very in-depth talk with two fellow theater artists about work that they have done in the past and it led to them discussing potential theatrical summer experiences. They shared insight and informed me of various ways to get involved with these institutions." (P104-S2-pre)

Monitoring task progress and ideas. In some cases, participants reviewed and evaluated their task progress, including both their search process and work process, to determine how they wanted to continue work on the task or whether they achieved the task requirements. For instance,

> "Reviewed my google doc sheet of notes to determine how I can complete my task." (P120-S4-pre)

> "I reviewed my notes and rewrote a list of things I would like to accomplish for this session." (P128-S6-pre)

5.1.4 XSS work modes

To investigate how the information search and non-search activities were intertwined and related to each other, we first abstracted all the non-search activities as *doing* (D), and then analyzed their relationship with the search activities (S) to determine participants' search and work modes for their XSWTs. We identified three different patterns: SDSD, DSDS, and SSD (Figure 5.1). We note that there are two main differences among these modes: (1) whether the process starts with a search activity or a doing activity, and (2) whether the search sessions are mixed with doing sessions through the task process.

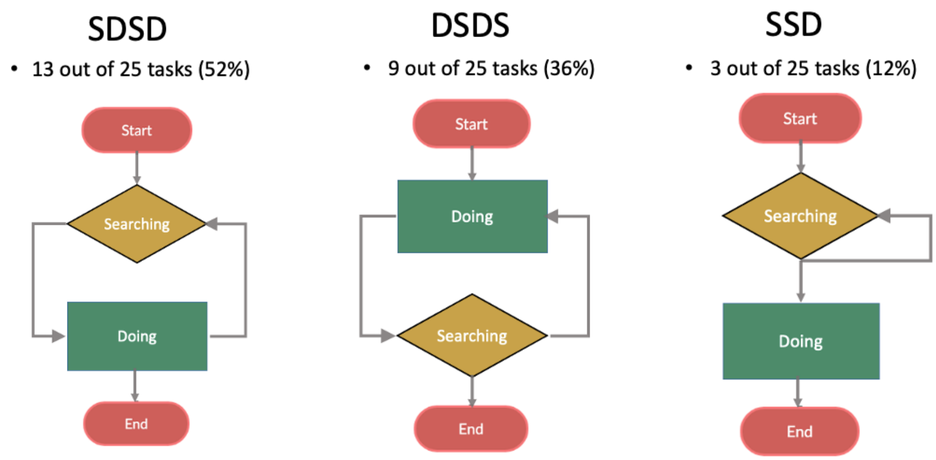


Figure 5.1: Search and work task modes for cross-session search work tasks

Search-Doing-Search-Doing mode (SDSD). SDSD was the most common pattern in our data (n=13). It describes a process when a person starts a work task by searching for information (S) (one session or multiple sessions) and then applying the found information to task (D). They may repeat this process until they have found all the information needed to complete their task. For instance, participant p108 wanted to create a dance show. She described that she started with a search session (S) in which she "searched [for] a lot of music, poems, literature and lyrics that I may be using in my piece." Then, she engaged in a doing phase (D) where she "drafted [an] idea on [based on her] research." Later she searched again (S) to find "some new songs" and "book and lyric quotes." Finally, she (D) "recorded text and lyrics for project use" and "rehearsed my piece from start to finish."

Doing-Search-Doing-Search (DSDS). DSDS was the next most common mode (n=9). It describes a process when a person starts to work on a task without searching. Instead, they begin the task based on their prior knowledge (D). As they move forward, they find that they need to look for more information, so they search online (S). As they find enough information needed for the current stage, they go back to work on the task again (D). After this the process may iterate similar to the SDSD mode – they keep switching between searching and working on the task till they achieve the task goals. For instance, participant p102 was helping his friends look for an apartment. Before he started to search online, he (D) "discussed *with my friends* and gave... my initial impressions...we specifically discussed *the apartment area* where I currently live." Then he started his search (S) to "look up apartment complexes...for more details." As he found information about different apartments, he texted his friends (D) to discuss "the complexes I've found and we narrowed it down to a list that sounded of most interest."

Search-Search-Doing (SSD). SSD was a less commonly observed pattern (n=3). It refers to the process when a person collects all the needed information across multiple sessions and then uses the found information to complete their task without further search. For instance, participant p101 who wanted to write a screenplay conducted three search sessions over 11 days. She searched for information and took notes about the background, local legends, and myths, but did not start to write. For the last search session, she described, "I wanted to do a... last sweep... to make sure I had all the links I needed before I start writing."

5.2 Session resuming and stopping reasons and their relationships

Search session overview In this study, we adopt the definition of session as a combination of both the search engine's view (i.e., search session is a *continuous period of time* when a participant search online) and participant's self-determined session boundary by reporting their session starting point and end point in the *post-session questionnaire*. First, the session length was calculated by using the participants' reported ending time point subtract the starting time point. We did not use the system log session times for two reasons: 1) the search log only recorded the time when participants clicked on a page, it did not include how long a participant stayed on the last page. Therefore the end time is hard to know in this case. 2) sometimes, participants took breaks during a long session when they worked on the task; it is hard to know how long the breaks (or the total time of all the breaks) took before the session ends. For instance, one participant (P103) filled out one of their session for 31 minutes while the system logged data show that the session (from the time of the first task link they clicked to the time when they clicked the post-task questionnaire after the last task-related link) is 128 minutes. On the contrary, another participant (P126) reported that one of their sessions took 85 minutes, while the system logged data showing that the session took 63 minutes, which is shorter than the actual session length. This might be because the first participant completed the session and took a break before returning to fulfill the post-session questionnaire. And the later participant may open the post-session questionnaire page once they decide their session is done to avoid forgetting the post-session questionnaire and then go back to the last page to complete the session.

Table 5.8 shows that, on average, participants conducted about 4 sessions per task (Min=3, Max =7). The medium session number is 5. The shortest session lasted about 9 minutes, and the longest session lasted 360 minutes (or about 6 hours). There are seven sessions that lasted about or longer than 2 hours (120 minutes).

Table 5.8: The minimum and maximum session lengths by participants (minutes)

P_ID	**# of sessions**	**Time of shortest session**	**Time of longest session**
P101	4	25	89
P102	7	20	50
P103	6	10	120
P104	3	24	31
P105	3	42	55
P106	5	9	16
P107	4	14	119
P108	3	162	308
P110	4	20	50
P112	3	20	69
P113	6	45	360
P114	6	16	59
P116	3	43	77
P117	5	17	28
P118	3	24	64
P120	5	24	44
P121	3	32	69
P123	4	23	124
P124	3	16	225
P125	3	22	40
P126	3	60	85
P127	3	15	35
P128	6	34	89
P129	3	22	99
P131	3	50	120

5.2.1 Session resuming reasons

We analyzed all the reasons that caused people to start a search session for their cross-session work tasks, including the reasons for the initial search session (which is not exactly "resuming" but starting reasons). There are multiple reasons: first, some proposed session resuming reasons in the MISE model do describe situations that may happen at the beginning search sessions, such as transmuting when people try to narrow down or clarify their task scopes, spawning when people start a session to find information for a related subtopic, etc. Secondly, in the preliminary AMT study, two initial sessions were started because of spawning and rolling-back, indicating that MISE resuming reasons can be applied for analyzing the initial session during cross-session search. Therefore, this section covers both the initial session's starting reasons and the resuming reasons for all the rest successive sessions for a task. In addition, for the convenience to conduct quantitative analysis, we consulted participants' self-selected main reason ("the closest reason that can match the main reason why you started the session") to determine the session's resuming reason if participants mentioned more than one reason in their open-ended description. In cases when participants' self-described reason cannot be fully covered by the MISE model, we still consulted their selection of the main MISE reason for quantitative analysis purposes but also took notes about the aspects that were not reflected by the MISE model. We will explain more details later.

Table 5.9 shows that among the 101 sessions, we identified five MISE resuming reasons: spawning (n=62), transmuting (n=19), unanswered (n=9), updated (n=9), and anticipated (n=2), from most popular to the least popular reason. We did not find any sessions that were mainly resumed because of the MISE resuming reasons of rolling back, transiting, or lost treatment, although we did find some similar situations that happened through participants' descriptions of their search process. In other words, our participants did not report any sessions that were started because the searcher wanted to find a new solution since the previous information did not work (rolling back), or the session was started to continue to search for a new task that was inspired by previous search (transiting) or wanted to find the information they saw before (answer lost).

Table 5.9: A Summary of Session resuming reasons and the related session stages

Resuming reasons	Initial Session	Middle Session(s)	Last Session	Total No.
Spawning	10	36	16	62
Transmuting	13	5	1	19
Unanswered	0	6	3	9
Updated	2	3	4	9
Anticipated	0	1	1	2
Total	**25**	**51**	**25**	**101**

As to the distribution of different resuming reasons across sessions at different task stages, Table 5.9 also shows that among the initial search sessions, spawning and transmuting are the two most popular resuming reasons. Spawning is also the most popular resuming reason for the middle sessions and the last session among the 25 cross-session work tasks. Unanswered is a frequent session for middle sessions compared to the initial and the last sessions. Looking for updated information is also a popular session resuming reason for the last search session.

Next, we will explain each of the five reasons in more detail.

Spawning Spawning describes the search resumed because sub-topics or problems emerged from the original task and ask searchers to find more information.

> "The spawning mode of successive searches is conceptualized as the search experience in which the information problem itself spawns sub-problems that extend over the surface of the original problem and that have a higher priority to cope with than the original problem. (Lin & Belkin, 2005, p.395)

Spawning is the most popular session resuming reason among all identified resuming reasons (N=62, 61%), especially for the middle sessions (N=36, 70%) and the last session (N=16, 64%) of participants' tasks. We identified different types of sub-problems that participants wanted to look for information for, including specific names, concepts, procedural information, different media type of information, and other details regarding a specific aspect of their overall task. For example,

> “ [I wanted to]look up information about small business insurance; look for tax information for an LLC.; I also wanted to go the [name of state] Secretary of State website to find out how to file within the state of North Carolina” (P116-S2)

> “[This session is] to start the first step and category in my project to discover a simple definition of sound and audio and the differences and similarities of the two” (P110-S1)

In addition to the types of information of sub-problems, we also identified several ways in terms of how these sub-problems were discovered by the participants:

(1) the subtopics (or sub-problems) were planned based on participants’ previous knowledge about the task topic and how they wanted to complete the task before they started. For instance,

> “I had a few names in mind beforehand...I started this session to learn more about [people’s names]...[and] I wanted to find contact information for [them]” (P105-S1)

> “I started this search session because I had been thinking about what I wanted to search the whole day, and I had a list of things that I really wanted to look up. ” (P121-S2)

(2) Subtopics emerged during earlier search sessions. In other words, participants might not be aware of the sub-topics till they found something as they started to search for information for the task. And they resumed more sessions to look for information for those new sub-topics. For instance,

> “part 4 of my project...so based on my 3 previous searches, I wanted to incorporate those” (P110-S4)

> “...these were my goals I had [for this session] based on my last study session and what I found through my search then” (P116-S2)

(3) Subtopic arose from the communication with others (e.g., task collaborator, or people that participants talked with during the time when they worked on the tasks. For instance,

> “I started this search because my friends and I had decided to hang out in one of our AirBnB rooms on Saturday night, and we will all bring snacks and different cocktails. ” (P103-S2)

"I had a very indepth talk with two fellow theater artists...they shared insight and informed me of various ways to get involved with these institutions. Hearing about other individuals' unique summer theatrical experiences peaked my interest to see if there very well may be a space for me." (P104-S2)

Another way that new subtopics were discovered is:

(4) Participants knew there were more subtopics related to their tasks, although they did not know exactly what they were before they resumed a new session, but they explored more subtopics during the resumed session. The purpose of the resumed session is to explore more subtopics and find information for the subtopics. For instance,

> "I started my searching on YouTube because I wanted to get some visuals of both the wildlife and [name of country] as a whole. There are a lot of nature documentaries and travel documentary/vlog-type videos, and I selected to bookmark the ones I think will be most useful and inspiring. I ran into a lot of subtopics here that I hadn't thought I'd get into"(P101-S2)

> "I started this search session to continue to find noteworthy guests across various sectors." (P105-S2)

Transmuting The transmuting reason describes that the task requirements were not clear to the searcher, and they want to find more information to help understand the task topic area and help understand how to complete the task.

> The transmuting mode of successive searches is conceptualized as the search experience in which the information problem is under formulation and gets enriched, polished, elaborated with new concepts, and finally changes its texture from the original state to the transmuted state (Lin & Belkin, 2005, p.394)

Nineteen (n=19) sessions' main resuming reason were coded as transmuting. Thirteen (n=13) of them are the initial search session of the task, whereas the other five (n=5) are middle sessions. One last session was started because of transmuting.

From the participants' descriptions, we found two main types of purposes for the transmuting sessions. The first purpose is to get oneself familiar with the topic, including general ideas about the subject/topic, potential information resources, and an overall idea about how to complete the task. Not all the information should be new to the participants. Sometimes people reacquaint themselves with what they knew already and what they did not in the topic area. Here are some examples:

> “I wanted to get a better sense of the information out there, what I already know, and get reacquainted with certain locations and aspects I love” (P101-S1)

> ” I would like to research the necessities for a short children’s book. What authors typically consider when writing and how to make sure a child will be able to take in and understand the topic of the book...To see how much it would cost for me to successfully execute writing the book” (P129-S1)

The second main purpose of transmuting is to find information to help formulate the scope and focuses of the task (e.g., create an outline, or find a starting point).

> ” [I wanted to] find a possible starting point [for the show], [to get] a firmer understanding of what my piece with a focus on as a whole...and a rough outline as to what the piece will be about, how long, the overall arc...” (P108-S1)

> ”I wanted to explore the costs and rarities of multiple [toy] sets, and narrow in on a few that I would be interested in further exploring to buy for my collection.” (P123-S1)

Interestingly, for the six non-initial transmuting sessions, their initial or previous sessions for the same task—the starting/resuming reasons for the sessions before those non-initial transmuting sessions–are also started because of transmuting. This indicates that for those tasks, it took multiple sessions from the beginning for the participants to gain general knowledge and information to help them clarify the task scope before they moved further. For instance, all three sessions of P120 were started because of transmuting.

> “I wanted to learn the definition of a "public issue" and research possible topics for my communications essay.” (P112-S1)

> “ I did not feel compelled to write about any of the topics I had found in my previous search session. I wanted more of a variety of topics to choose from.” (P112-S2)

> “To finalize my topic before drafting my essay [and] To learn more about the topics I am interested in” (P112-S3)

During this cross-session search process, the participant (P112)’s task goal was to identify a topic related to a “public issue” to write an essay. During the first session, s/he started to look for information to help understand the meaning of “public issue” and then started to explore possible ideas and topics to choose from. S/he continually explored different topics during the second session and the third session. In the third session/the last session, they finally decided on the topic and started to look for more specific information. As they described the S3’s search process:

> “ I browsed through my bookmarked topics and decided that I was the most passionate about the threat that child beauty pageants pose to the physical and mental well-being of young girls. After deciding on my topic, I searched for scholarly articles on the [university] Library website and on other news outlets that described the dangers of child beauty pageants. I bookmarked all of these tabs so that I can refer back to this information while I draft my essay. ” (P112-S3)

As the participant decided on their topic, they achieved their cross-session work task goal (find a public issue topic to write). Although this participant completed their task for this diary study, as we could predict, as they continue to complete the essay task, they may search for more specific information related to the selected topic for other reasons (e.g., spawn, undated, anticipated, unanswered).

Another participant (P128) whose first three sessions (out of seven sessions) were also started because of transmuting:

> “My goals are to get a general idea of the types of statistical sampling that are out there. I’m not entirely clear where this initial search will take me, so my only goal is to orient myself to the topic.” (P128-S1)

> ”I wanted to get through the Wikipedia article on sampling so that I could gain a general overview of the topic and move on to other pages. (P128-S2)

> “I wanted to learn more about statistical sampling in general. I didn’t have any specific topics that I wanted to learn about within this topic because I am too new to the topic

> to have any idea what I would be searching for other than the general topic itself." (P128-S3)

The above examples show that the participant (P128) started to explore a new knowledge area (statistical sampling) that she was not familiar with at all (S1) and followed a general information resource (Wikipedia page) (S2) and continued exploring (S3) without a specific goal in her mind. But as she learned more and more about the task, she started to have more clear directions for her search, and her latter sessions became more focused on specific subtopics to explore as *spawn*.

Unanswered/incomplete Resuming reason *unanswered* describes the situation when the searcher resumes a new search session to look for the information that they did not find in the earlier search session(s). As shown in the quote below:

> "The unanswered mode of successive searches occurs simply when the searcher cannot find satisfactory information objects as treatments by the termination of a search episode. Renewal of search sessions is necessary to complete the task." (Lin & Belkin, 2005, p. 396)

Lin and Belkin's original description indicates that during both sessions (the previous session and the successive session), the searchers' information needs to focus on specific information objects that could solve their problem, and the resumed session is inevitable to finish the task.

On the other hand, we found when participants chose "unanswered" as the reason that most closely describes why they started a session, they often described situations when they wanted to continually look for information they started during a previous session(s). In other words, their "unanswered" session was seldom because of "unanswered" not finding what they needed but more often because their previous session was *incomplete* or *unfinished*. This might be because of the complex nature of users' needs for complex tasks, which often ask people to look for various types of information. In our AMT study, we found that the information-gathering task is the most common cross-session work task that requires people to collect information from different resources over time. Therefore, we followed the same trend by grouping sessions that were resumed because participants wanted to continue an *incomplete* and *unfinished* session under the *unanswered* code. There are nine (n=9) sessions that were stopped because of *unanswered/incomplete*.

Specifically, we also identified several different types of *incomplete*: (1) A participant had a specific goal to fulfill with many different options and already found some information in the previous search session, but they wanted to continue to exhaust the available choices. The goal could be to find more similar information regarding the goal or to fully explore the information resources they already found during the earlier session(s). For instance,

> "I wanted to exhaust the rest of my current resources to see if anything intriguing was available for me to apply" (P104-S3)

> "[I wanted to] finish my list...[and] I was looking for a few more benefits to add to my list until it felt complete" (P120-S3)

In some other cases, (2) a participant did not have a specific goal but to explore for more information following a more general search direction as they did for the last session. It needs to be pointed out that this is different from *spawning*, for which the new session aims to look for information about a different subtopic; when the session is started for *unfinished*, the participant did not change their information needs/topic in the new session. For example,

> "I started this session because I did not finish looking for ideas earlier." (P103-S7)

> "I wanted to learn a bit more about complete streets policy to make sure I had the full spectrum of potential policies before taking the ideas to [stakeholder]" (P128-S3)

Table 5.9 also shows that unanswered as the resuming reason is associated with middle sessions (n=6) and the last session (n=3) but not for the initial sessions. It makes sense because a successive session started because the searcher did/not have to find all the needed information during the previous search session(s), and there is no previous session before the initial session. Unanswered sessions could be the reason for one session, or it could last for multiple sessions (e.g., multiple middle sessions, or a combination of the middle session and the last session). Among the 8 unanswered sessions, eight (n=8) participants had one session resumed because of unanswered reasons, indicating most of them found what they wanted by the second session (the resumed session). There is one (n=1) participant who had more than 1 unanswered session: P120 had two middle sessions (S3, S4, out of five sessions) that were resumed because of unanswered. In other words, P120 spent three sessions (S2, S3, S4) looking for the same information.

Updated/cultivated Updated, or cultivated, means a searcher resumed a search session to look for information that is up to date. As described in the MISE model:

> "The cultivated mode of successive searches refers to the search experience in which the searcher is trying to cultivate or to stay abreast of an area of information of interest, which Marchionini (1995) called accretional information seeking. the information problem... is fuzzy because a search in such a mode of successive searches tends to have general and less defined information problems. (Lin & Belkin, 2005, p.395)

Following Marchionini(1995)'s definition of accretional information seeking for which both the task and information needs (solutions) are ill-defined, the original updated/cultivated resuming reason for a successive search also suggests that people resume a session to look for the latest information when their information task has not been fully developed, and their interests in information needs are more general.

However, our findings show that for all nine (n=9) updated sessions, participants were clear about their task goals, and their information needs were specific: none of the nine updated sessions followed a transmuting previous session. Instead, five (n=5) updated sessions' previous sessions were started because of spawning–participants were looking for subtopic information in the previous session and then resumed the search to look for updated information. Two (n=2) updated sessions are also initial sessions. The rest two (n=2) updated sessions were by the same participant, and the first updated session followed an unanswered session (unanswered-updated-updated). Participants who started two initial sessions (P104-S1, P106-S1) to look for updated information also had a clear goal about what they wanted to find:

> "[I wanted] to identify if there are new and upcoming auditions for roles that I could be a strong candidate." (P104-S1)

> "I'd seen articles but hadn't had time to read when I saw the headlines. I needed to follow up on that information. I know this info is still changing, so I wanted the latest." (P106-S1)

Among the five updated sessions, there are two different types of information participants were looking for (1) Participants looked for updated information for previous information they searched during earlier sessions to make sure they stay abreast with the latest status of the found information. Usually, they could find such kind of updated information from the same information resources, involving refinding, comparing. For instance,

> “I was also curious to see if any of the flight information that I had looked up before had changed” (P117-S2)

> “Before I make a final monetary transaction, I just wanted to do a quick last search to make sure the deals I found were as great as I thought they were.” (P123-S4)

The second type of updated session is (2) participants knew the task topic areas keep changing and wanted to find the latest information that is available. In this case, the participant may did not search the information before, and they wanted to explore the most recent information related to their search topics. For instance,

> “I’d heard that the system had changed recently and wanted to know the latest.”(P106-S3)

> “I figured before I started officializing plans of shipping and payment to sellers I’ve already contacted, I should surf a couple of tertiary markets for a bit longer to make sure all my current options are on the table and maybe find a few additional deals.” (P123-S3)

Anticipated The problem anticipated predicts successive searches that might happen in the future for a future task. Instead of looking for information for an existing problem/task in their mind, a participant encounters some unique information first and assumes that the information would be helpful in the future, and they would search again (successive) when the predicted task really happens. In other words, the successive search session won’t happen if the anticipated problem/task does happen in the future.

> “The anticipated mode of successive searches depicts the search experience in which the information problem has not occurred but is anticipated because some elements of the current situation match patterns of past experiences which are estimated to revive the envisioned or planned situation in the future. ” (Lin & Belkin, 2005, p.397)

However, this original definition has two key factors that make it hard to capture such kinds of sessions: (1) the participant encounters the information during a search which is hard to predict, (2) the participant is attracted by the encountered information and process it in the scenario of a future task.

Taking all these factors into account, we adjusted the meaning of anticipated: Following the AMT study that people selected the "anticipated" resuming reason when they really were searching for information that related to the current task but would be useful for later use (e.g., plan a party and look for ideas about activities), we used the same description in the diary study that the sessions were resumed to look for information related to the current task but for future use when the participant took further actions as their task move forward. The goal of the session is to collect information that the participant thought would be useful for the task at a later time.

Overall, we identified two sessions (n=2) that were mainly resumed to find information for future use as the participants' task move forward (in the future). For instance,

> "I wanted to get inspiration for ideas for fall activities and clothing [for my graduation trip]." (P103-S5)

> "I simply wanted to find inspiration and other artwork to put toward later use" (P114-S3)

In addition, we also found *anticipated* could also be a secondary goal as the participant's main reason for resuming the session differs. For instance,

> "My first step is to try to define sound and audio in the simplest form possible and hopefully bookmark some websites for reference and possibly discover other sites that will help me further into the project."(P110-S1)

In the same session, the participant described that

> "I also found some links to microphones which is the next category I'll be researching." (P110-S1)

As we can see from these quotes, participants specifically mentioned that they wanted to find some information either for later search sessions or for future use as they apply the information to tasks at a later time but for the same projects.

Other MISE resuming reasons As we mentioned earlier, we did not find any sessions that were resumed mainly because of problem transiting, rolling back, or answer-lost. But we did find examples when participants described similar situations that happened during their search process. To be consistent, we still explain all the original definitions and povide quote examples from our participants.

Transiting. The transiting mode describes situations when the participant is inspired by the search and the results they complete earlier for a task and decide to conduct more searches following the inspiration for a new task.

> "The transiting mode of successive searches is when the information problem is transited to another; the two problems conceptually differ per se, but are connected in a cause-effect relationship (i.e., the result of A constitutes the cause of B). (Lin & Belkin, 2005, p.395)

The key points of transiting mode are that (1) the two problems are completely different; (2) the new task is motivated by the previous task.

In some cases, participants did describe a similar situation when they found some interesting, related information that is worthy of more search for a different task. For instance,

> "After that, I did a quick search to see if I could find some quick statistics on the benefits of taking space away from cars more generally, but determined that was another large task and stopped that search. " (P120-S4)
>
> "I changed my topic because I became interested in something related. " (P124-S2)

In the first example (P120) found some interesting information but decided to stop continuing their search for "another task" and back to the original task. In the second example, the participant, in fact, was in a "transmuting mode" as they tried to find a topic to write about. As they were searching for more information for the initial topic, they encountered some more interesting information and completely changed their task topic to the new one. Neither of these two examples matches the definition of transiting mode: P120 did not continue the search for the new task, and P124 did not complete the previous topic, and from the task perspective, s/he switched their task topic, but not to a completely new different task. This might be because the diary study only required participants to record their activities for the selected task, and participants did not want to be distracted, and they won't report activities they did for a different task.

Rolling-back. Rolling-back resuming mode describes sessions that are resumed because the searcher found out that the information found previously did not work, so they need to search more for different solutions for the task.

> "The rolling-back mode of successive searches occurs when the problematic situation that once was thought to have been resolved actually turns out unresolved. [the searcher] will need to search for alternative treatment." (Lin & Belkin, 2005, p.395)

We found one (n=1) example when the participant stopped an earlier session because what they found did not work. However, when restarting the search, the participant did not continue to look for substitutes. Instead, they switched to another subtopic. Therefore, this is not exactly a rolling-back case. As the following quote shows:

> "I found an app ...but [it] did not work on my laptop, so I decided to end my session out of frustration" (P114-S5)

and in the next search session, the participant described:

> "Once I cooled down from my incident with a poorly constructed and expensive application, I needed to 'restart' my initial search for artistic tutorials and artistic inspiration" (P114-S6)

Lost-treatment. Lost-treatment mode, or answer lost, describes sessions that are resumed because the searcher cannot access the information they found earlier. As a result, they need to redo the search to find information to solve their problem.

> "The lost-treatment mode of successive searches takes place when the information treatment of the information problem, once found, is not available in the treatment application stage... the problem rolls back, the searcher might be forced to re-examine the definition of his/her information problem and revise his/her expectation of treatment. Renewal of the search thus becomes necessary " (Lin & Belkin, 2005, p.396)

This original definition addresses that (1) the information was found earlier but no longer accessible later, (2) consequently, the participant who lost the treatment may not only want to find the lost information, but they may also reconsider their task requirements and make changes about how to solve the problem. For instance, they may think of finding information from different information resources or finding other alternative solutions.

We did not identify any sessions that were resumed mainly because of the lost-treatment reason. On the other hand, we found some participants described refinding information they saw earlier, a search strategy that could be used during lost-treatment mode when people need to regain the information they found previously. Noticeably, the purposes of refinding differ from treatment lost. For instance,

> "get a better sense of the information out there, what I already know and get reacquainted with certain locations and aspects I love [about the topic]..." (P101-S1)

> "I was looking for music, text, poetry, and literature that would start as a foundation for my project. I was also mainly looking for inspiration from things I knew of, but wanted to look up thoroughly and explore more as to how they may or may not relate to my piece and what I want to present." (P108-S1)

In the above two examples, the two participants were trying to refind the information they knew of (before they started the task). In other words, they refund information in order to review how their previous knowledge could help them with the current task. It could be some specific information that is useful for the current task (e.g., P101), or reacquainting themselves with the previous information can help them come up with new ideas for the current task (e.g., P108).

Refinding is also a strategy participants used to look up the information they saw during earlier sessions as the tasks moved to a later stage. They refound the information and information resources to help them gain a better understanding of their task and the information they found. For instance,

> "I revisited some websites in order to find the information I'd already seen but needed to review"(P113-S6)

> "I was looking for a few more benefits to add to my list until it felt complete, so I went back to some articles I had previously pulled up" (P120-S3)

As we can see, none of these refinding behaviors are caused by treatment-lost, or the searchers cannot access the information found previously. On the contrary, during the refinding process, people seem more confident about finding the information again with certain strategies (e.g., redoing the search, reviewing bookmarks). This might be because there is so much information online and it is easy for people to find what they need.

Resuming reasons that are not in MISE Although our diary study asked the participants to select one of the MISE reasons that are mostly close to the reason why they resumed each session and we also use the MISE to code the main resuming reason from participants' responses to the open-end question about why they resumed the session, we also identified some reasons that are not covered by the MISE resuming modes. We categorized these reasons into two groups (1) information needs-related reasons; (2) non-information needs-related reasons. Below we list these session resuming reasons with examples:

The information needs-related reasons are causes that are motivated by users' information needs for their tasks. Users started their sessions to look for more information. Specifically, there are five information needs-related reasons:

To exhaust available information resources. Participants mentioned they resumed sessions to go through (more) available information to make the most of the resources they have. For instance,

> "[I want to] Skim a few more research articles to make a *full* list of benefits" (P120-S3)

> "I wanted to find all of my quotes and attributions and collect any last-minute information."(P125-S3)

To explore more topic areas. Different from exhausting available information resources, participants who wanted to explore more subtopics did not mean to scout for everything that was relevant. Instead, their goals are more general, and would search by following whatever they find. Meanwhile, it differs from the *transmuting* mode because, by the time the participant had a clear goal about the task requirements, they were looking for information to help them achieve their goals instead of clarifying the tasks' needs. Exploring is also different from *spawning* because as they started the session, the searcher wasn't known whether they would search for a specific subtopic/concept or they would search more broadly. For instance, "I wanted to learn more about statistical sampling in general. I didn't have any specific topics that I wanted to learn about within this topic because I am too new to the topic to have any idea what I would be searching for other than the general topic itself." (P128-S3)

To find inspirations/examples. Participants often mentioned to resume search sessions to look for examples similar to their projects or find some inspiration to come up with ideas about how to work on their own tasks. For instance,

"I wanted to get inspiration for ideas for fall activities and clothing."(P103-S5)

" I was excited and ready to get to the inspirational portion of finding examples across the world." (P120-S3)

Sometimes, this might overlap with *spawning*, for instance, when the participant wanted to find some inspirations and examples that specifically focus on a new subtopic area that they had not yet explored. But as many spawning sessions have a specific goal to investigate information about certain aspects of a task, finding inspirations and examples are less concentrated: People may browse information resources to catch relevant ideas and thoughts. In addition, they may not have a specific goal about how much information they need, how deeply they want to understand the inspirational information, and whether or not they want to use the information later.

To review what they found earlier. Participants also mentioned resumed search sessions to review the information found in previous sessions. Sometimes people review to make sure/validate that the information they found earlier is correct. Sometimes they go through the information they found in previous sessions for a closer investigation to extract useful information as their task scope became set down in the latter sessions. For instance,

"I was specifically searching for information that I had saved in order to make sure I cited it properly" (P125-S3)

"My goal is to look through the materials I have already bookmarked and to take

> notes on what I plan to use from the sources and articles to write a plan/outline for my book." (P129-S2)

Resuming a session for review is different from refinding information that was lost or getting updated information, although reviewing also involves refinding behavior. In the reviewing mode, the participant knew that the information should still be there or they had the information at hand (no lost), and they did expect the information would be changed/updated. On the contrary, they purposefully find the original information as they found previously.

Consulting others triggers new search sessions. Talking with people about the task could also cause the searcher to start new searches. The input from others can provide new ideas and directions for search or change the original searchers' perspective about their tasks. The searcher may decide to look for what others suggest or find something that is different from what others say. Sessions resumed due to consulting others can happen to both tasks that involve other workers and solo task doer (tasks that were only conducted by one person). For instance,

> "I had a very in-depth talk with two fellow theater artists about work that they have done...hearing about other individuals unique summer theatrical experiences peaked my interest to see if there very well may be a space for me." (P103-S2)

> "I started this session because my friends said they wanted to book our flight today." (P117-S4)

The (2) non-information needs related reasons are reasons that not focusing on fulfilling the information needs of the task but other reasons related to task progress or one's mental/physical status:

To monitor task progress. Participants mentioned they wanted to conduct a search session in order to get a feeling about the workload or the progress of the task so that they could appropriately arrange their task and time. For example,

> "I was curious to see if only doing one apartment at a time would suit me better."

(P102-S2)

> "I also remembered that I have a deadline for this task because my race is at the end of September, so I need to start shopping if all of the components are going to arrive on time. " (P121-S1)

To complete search following a scheduled plan. For tasks that participants already have a work plan, participants often follow the plan to work. For example,

> "I'm now deciding to do one complex a day through the end of the list, so this was just my daily session!" (P102-S4)

Feeling in a mood/having the energy to search. Sometimes, participants wanted to conduct a search session just because they felt good and wanted to work on something to make the most of their time and energy. For instance,

> "I had a lot of energy compared to my usual amount of energy ...I felt in the research mood." (P101-S1)

> "I started this session because I felt like studying and had finished all of my work that was due this week." (P126-S1)

5.2.2 Session stopping reasons

Similar to the resuming reasons, we also analyzed the stopping reasons for all search sessions based on participants' responses to the open-ended question about why they stopped the session and their selection of multi-selection multi-choice questions from a provided list. We coded the main session stopping reasons from their free-style responses to the open-ended question and double-checked it against their choice of the multi-choice question. The rule of coding and selecting the main stopping reason is that if there were multiple stopping reasons, task-related/search-related reason goes first. If there were no task-related/search-related reasons, non-task/search-related reasons would be selected (e.g., interrupted by others).

Table 5.10 presents an overview of the session stopping reasons for the 101 sessions. Overall, *goal achieved* and *satisfied* are the most two popular session stopping reasons when people stopped searching for a session, and both commonly happened across the entire process of the cross-session search (i.e., initial sessions, middle sessions, and last sessions). Surprisingly, not all tasks ended with the last session's stopped for *goal achieved*: Besides 11 sessions (44% of the 25 last sessions), ten last sessions were stopped because participants felt satisfied with what they found, one participant stopped the last session because of need to consult others, and three were stopped because of interruption. Interruptions were the third most popular session stopping reason following goal achieved and satisfied. A small number of sessions were stopped because of could not find the needed information or consult others, and one session was stopped because of low task knowledge

Table 5.10: Summary of session stopping reasons

Stopping Reasons	**Initial Session**	**Middle Session(s)**	**Last session**	**Total No.**
Goal achieved	8	15	11	34
Satisfied	10	14	10	34
Interruption	6	13	3	22
Need to consult others	1	4	1	6
Cannot find the needed info.	0	4	0	4
Low task knowledge	0	1	0	1
Total	25	51	25	101

Goal achieved We coded users' stopping reason as "goal achieved" when (1) there were clear standards/quantifiable measurements of participants' goals (e.g., find two articles, purchase something), or (2) when they explicitly said that they achieved what they planned for the session, for instance, "I have finished researching and preparing for my project", "I found all the links I needed".

Overall, we identified thirty-four (n=34) sessions that were stopped mainly because participants achieved their goals by the end of the session. Specifically, eight (n=8, 24%) initial sessions were stopped for goal achieved, fifteen (n=15, 44%) middle sessions were stopped for goal achieved, and eleven (n=11, 32%) last sessions were stopped for goal achieved.

Considering that participants would not complete their final task goal until the end of the last session for their tasks, stopping search sessions that happened at early task stages for *goal achieved* indicates that participants had stage goals or sub-task goals throughout the process.

Many initial sessions and middle sessions that were stopped for goal achieved often described sessions that participants had clear aims, and by the end of the sessions, they were confident that they accomplished what they planned earlier. For instance, participant P102 had a plan about the amount of information he wanted to search through the process, and for multiple sessions, he expressed a similar stopping reason that

> "I am doing one session with one [topic example] a day, so after I finished I stopped." (P102-S4)

And P113 noted that

> "I had found four articles, which exceeded my goal of 2-3, searching just in the two journals I was targeting." (P131-S2)

Another participant P116 described

> "I stopped this search because I completed my intentions for this first search. "(P116-S1)

Differently, participants often described their goals were achieved for the last session by saying that they found all needed information. For instance,

> "I had found everything I absolutely needed." (P102-S7)

> "I stopped this search session because I found all the information I was looking for as well as guidance that I need to complete the [task]." (P116-S3)

It is worthy of being pointed out that out of the 11 last sessions that were stopped for goal achieved, only one participant (P113) explicitly said that s/he stopped because of *the work task itself is completed*. Otherwise, many participants' descriptions implied that there might be more work needed to be done after they stopped the search process. For instance,

> "I made a physical schedule of everything that I will have to do that weekend. I'm done with planning at this point." (P103-S6)

> "I decided to stop because I had reached a point where I had 2 sources for each of my body paragraphs and had an outline done. While I am not done with my paper yet, I have made significant progress on my research and have reached a milestone." (P126-S3)

Therefore, here the *goal achieved* is more for describing the search goal achieved than the work task goal achieved. From this perspective, it also means our codes were appropriate for describing the reasons leading to the stopping of search session rather than task progress.

In addition, there were situations when a session was without a clear goal at the beginning but ended with "goal achieved". It happened when a participant did not have a specific goal for searching but just a vague idea to explore; as they found more and more information or were inspired by information encountered during the search process, some goals arose. Then the search became more concentrated, and by the end, the participant accomplished what they wanted. For instance, In her pre-session questionnaire for the third session (P107-S3), the participant described that

> "I'll be looking for my baby shower gift, figuring out what baby shower games to play/look for inspiration ." (P107-S3)

Then in the post-session questionnaire, she described her search process at the beginning as that:

> "honestly, I had no idea what I was looking for..."(P107-S3)

and after encountering some interesting information, she said again:

> " What I found is that Amazon was having a buy 2 get one free sale for baby books, so I thought that was perfect!" (P107-S3)

and by the end of the session, the participant described that

> "I basically added all options ...and then dwindled down which one would be best....I think overall it'll be okay! A great successful task indeed!" (P107-S3)

However, sometimes, a session started with a clear goal could possibly end for other reasons instead of *goal achieved*: (1) participants did not find what they needed (coded as cannot find information), we will discuss this later; (2) participants only found part of what they needed but not all the information, they stopped the session for "satisfied", for instance,

> "I found almost all of the available information but had trouble finding photos...however, I still believe the information I found will be useful enough" (P102-S5)

(3) participants found the needed information for the initial goal of the session (goal achieved), but then they started to look for other information (e.g., a new goal, or next step of the task, or a different aspect of the task), the reason for stopping the session would depend on whether they find the needed information for the new search, for instance, participant P112 described,

> "I started my session by researching the definition of [topic related taxnonmy] I took notes on these definitions and then proceeded to search for articles about current public issues...from there, I compiled a list of at least five possible issues that I may want to write." (P112-S1)

This last scenario implies that a search session is not necessarily equal to a search task; participants might conduct multiple search tasks within the same session, as some search tasks may succeed, some others may not fully be satisfied.

Satisfied *Satisfied* described situations when people felt happy about the information search process and the results they gained by the end of the session. We used *satisfied* to code session stopping reasons when participants' session goal or task goal was more general than specific (e.g., qualitative goal vs. quantitative goal), they did not have ideas about what exact information they wanted or where to explore, but as they explore the information space, they collected information, digest what they found, evaluate how useful the information was for their task, and incorporate the information together (with their task or other information) and achieve a point where they felt good or enough to stop. These cases often include expressions related to attitudes or content feelings (e.g., enough, good, satisfied, etc.)

Satisfied is the most popular (n=34) session stopping reason. A majority of sessions that were stopped for *satisfied* were initial search sessions (n=10, 29%) and middle sessions (n=14, 41%), while ten (n=10, 29%) last sessions were stopped because of satisfied. The distinction between *goal achieved* and *satisfied* is that for goal achieved, participants explicitly said they found what they needed, they were done with the current search and wanted to move to the next sub-goal, whereas, for sessions that stopped because of *satisfied*, people might not evaluate the session based on whether they complete what they wanted. Instead, it was more about how they felt about what they achieved by the end of the session.

For instance,

> "It felt like something of a natural endpoint to the search session because I was at a good place to continue the next search more concretely" (P101-S1)

> "I used this session to gather the bare minimum of information regarding my paper topic in order to familiarize myself with what I will be writing about over the next 2 weeks. I stopped this session because I found enough to feel comfortable with completing the task." (P113-S1)

It is clear that participants felt *satisfied* after cognitively processing the found information during the search process. For instance,

> " I synthesized this information to form a critical opinion that will support the thesis of my paper. I felt this was a good stopping point before continuing my research over the next few days" (P113-S4)

> "For the article, I read through the entire piece to truly understand the content and grasp the art form. For the video, I also watched the whole thing to view how the dance would be in motion by a professional." (P127-S2)

These examples show evidence that participants processed the found information simultaneously as they searched.

We also found that *satisfied* was often associated with the expression of "natural stopping point" that refers to transit points where people could easily switch to another search topic or task stage next time. For example,

> "It felt like something of a natural endpoint to the search session because I was at a good place to continue the next search more concretely" (P101-S1)

> "I have plans and felt this was a good stopping point before continuing my research over the next few days." (P113-S5)

Interruption Interruption describes reasons that are not related to users' information needs or task progress but cause people to stop their ongoing search session. In order to distinguish interruption from other task-related reasons we mentioned in this research, here we define interruption as non-task-related reasons that lead people to stop searching, which is similar to the imposed interruption in Lin and Belkin's (2005) model. Then we defined self-motivated interruption (or internal interruption) as the stopping reasons that the task doer(s) themselves initiated, or related to themselves, such as mental and/or physical fatigue, stop to work on something else, meanwhile, the interruptions that initialed by others, or by external environmental factors that we labeled as an external interruption. But none of the interruptions are related to any task factors.

Our results show that interruption is the third most popular session stopping reason. There were 22 sessions out of the 101 that were stopped because of interruption. A majority of them (n=13, 60%) interrupted sessions were middle sessions. More importantly, almost all the interruptions were self-initiated, while only a few interruptions were introduced by external factors/people. Here are some examples:

Examples of self-initiated interruptions:

> "It was getting late and I needed to sleep! Otherwise, I would continue to search for more books or other additional gifts, like a closet organization system." (P107-S3)

Example of the external interruptions:

> "My mom and I are leaving the house to go on a hike. I thought we were leaving later in the evening and would have more time to search." (P103-S5)

> "Move locations from library to student union, as well as a low battery and need to charge laptop" (P124-S78)

Interruption also includes tasks people decide to stop before they complete them during study time. This is because of unknown reasons (i.e., participants did not explicate why they decide to stop the task by the last search session.). For instance, participant P107 described in her last session that "I needed to search how to cancel evites and the wording of how to delicately put this. " and mentioned in *other stopping reasons* "end of the study!" (P107-last).

Need to consult other information sources Need to consult other information sources describes that participants stopped a search session because they would like to look for information from other people who may or may not be involved in the tasks. We found six sessions stopped because the participants wanted to consult other people to gain information for their tasks.

For those that had been involved in the task, their information input often is an integral part of making the task move forward. For instance,

> "I emailed [name], which was the main thing that I wanted to cross off my to-do list. I feel like I have learned a lot of information tonight and want to process it. I can't officially settle on a dinner reservation until I hear back from [name]." (P103-S3)

For those who were not involved in the tasks, participants often thought those people could provide extra, specific (e.g., expertise, similar experiences) valuable information that could help them know more about the task. For instance,

> "I also wanted to get my friends' opinions on the [name] (just like a "hey random-ish question, any thoughts on..." type thing)" (P101-S1)

Other information sources could also be sources that they could not access during the search session, or need extra assistance to access, such as a physical library resource. For instance,

> "I also realized the source I wanted to use for further source finding is one I'll have to get physically through ILL." (P102-S1)

Cannot find the needed information Cannot find the needed information refers to situations when participants have some ideas about what information they wanted for their tasks (the needs could be vague or clear) and they spent a certain amount of time and effort to look for the information that matches their needs, but they did not find and decide to stop.

There were four (n=4) sessions participants stopped because they could not find what they needed and stopped the session. For instance,

> "I was mildly annoyed that I couldn't find a simple cocktail recipe and also frustrated that the cake websites didn't have an allergies session." (P103-S2)

> “I’m too frustrated with inconsistent information and websites of shuttle companies that don’t work. I think I’m going to just take a Lyft or Uber for simplicity and not have to deal with booking a shuttle in advance.” (P106-S3)

> “ I found the official Disneyland page about reservations, which says they open up 60 days in advance. But when I went to the actual restaurant’s reservation page, that availability wasn’t working. It just shows no reservations available. I then did more searching to try and figure out what was going on, and ended up on some Reddit threads. People there were having the same issue. I didn’t find a fix...I realized that the problem I found isn’t just me, others have it too, and all I can do is check back later on the reservation availability.” (P106-S4)

> “I found an app in the app store for which I paid a lot of money for but did not work on my laptop so I decided to end my session out of frustration.” (P114-S5)

From the examples above, we can find that there are several different reasons that made people cannot find what they needed: (1) the needed information was not provided or available (i.e., P103-S2, cake websites; P106-S4); (2) the provided information did not work (i.e., P106-S3; P114-S5), or (3) the participant did not find information that matches their requirements (i.e., P103-S2, simple cocktail recipe).

When participants could not find what they needed, they might take different strategies to address the problem: sometimes, they would try again to look for the information in a later time. For instance, Participant P103 did not find what she was looking for in her second session (P103-S2), and later in her last session, she mentioned that

> “so I looked for simple fall drinks that would also be cozy. I found another list by Oprah and looked through it because I liked the last Oprah list that I saw. I opened potential drinks in a new tab and got a few fun ideas to try... [I] wanted to check the price of cakes at a bakery in case I decide to bring cake.” (P103-S2)

And participant P114 described that

> “Once I cooled down from my incident with a poorly constructed and expensive application, I needed to ‘restart’ my initial search for artistic tutorials and artistic inspiration. I

> wanted to see what sites, what paywalls were there, if any, essentially a slightly more formal squirrel looking for his "nuts" of information for which I can come back to later. luckily most of these apps already have a large following." (P114-S6).

However, it has to be noted that for both of these later sessions that resumed after a previous session ended for *cannot find the needed information*, the latter sessions were not resumed for reasons of looking for the unfound information like unanswered, or rolling-back (previous information did not work). Instead, all the later sessions were resumed because of *spawn*, meaning that the participants started the new session to look for something else. This might be because the information they did not find was not that important, which could affect their task process.

Another strategy used for cannot find needed information is to find some alternative choices. For instance, participant P106 mentioned that after he could not find shuttle information, he decided to use Uber or Lift instead. In another two examples, although the participants did not find what they needed, they did not stop their search. Instead, they kept searching and found different solutions that could still satisfy their needs. For example,

> "I hit a dead end in my search. I wasn't finding much else, and I wasn't super interested in looking for busted [name] since most of them weren't ideal colors and I don't want to paint them. Nonetheless, I found enough options to be able to choose what I need later." (P118-S2)

> "I was content in my finding of 6 fantastically priced listings, even though they weren't the exact products I was looking for. I messaged all the sellers so all I can do now is wait on that front. My next step is to search eBay for the original two products I had in mind..but even if I can't find any reasonable deals I now have 6 other pending possible collection purchases awaiting seller response." (P123-S1)

Low task knowledge Lin and Belkin (2005) also mentioned that a low understanding of the goal could also cause the session to be interrupted (i.e., internal factors). We did find an example when participants mentioned they stopped the session because of the information overwhelming and were not sure to what extent the search should be continued.

"I felt overwhelmed by my uncertainty about going too far in one direction and eventually uninspired to continue." (P120-S3)

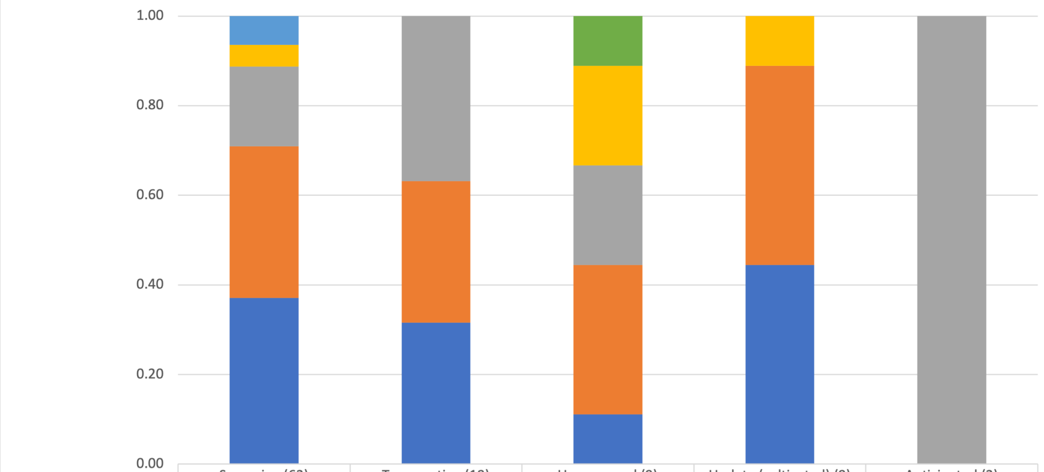
5.2.3 Relationships between resuming and stopping reasons

To evaluate the relationship between the session resuming reason and the session stopping reason, we performed a Chi-Square test, and the results are shown in Table 5.11. The relationship between these variables was not significant ($\tilde{\chi}^2$ (20, 101) = 29.139, p=.085).

Table 5.11: Chi-Square test for the relation between session stopping and resuming reasons

	Value	df	Asymptotic Sig. (2-sided)
Pearson Chi-Square	29.139	20	.085
Likelihood Ratio	26.798	20	.141
N of Valid Cases	101		

We also present a descriptive graph of the percentage of different stopping reasons to the associated session resuming reasons. The Figure 5.2 shows that:



	Spawning (62)	Transmuting (19)	Unanswered (9)	Update (cultivated) (9)	Anticipated (2)
Low task knowledge	0.00	0.00	0.11	0.00	0.00
Cannot find	0.06	0.00	0.00	0.00	0.00
Consult	0.05	0.00	0.22	0.11	0.00
Interruption	0.18	0.37	0.22	0.00	1.00
Satisfied	0.34	0.32	0.33	0.44	0.00
Goal achieved	0.37	0.32	0.11	0.44	0.00

Goal achieved Satisfied Interruption Consult Cannot find Low task knowledge

Note: each number in the cell indicates the percentage of a stopping reason that happened across all sessions resumed for a specific resuming reason. For instance, 0.37 (the bottom cell in the first column) means that across all 62 spawning sessions, 37% of them stopped because of goal achieved.

Figure 5.2: The percentage of different session stopping reasons to the session resuming reasons.

1) As the two most popular stopping reasons, *goal achieved* and *satisfied* happened almost across all different resuming reasons, except the sessions resumed for *anticipate*. *Goal achieved* is most popular for *updated (cultivated)* sessions (44%), followed by *spawning* (37%), *transmuting* (32%), and *unanswered* (11%), which indicates that participants found what they needed for most of the *updated (cultivated)* sessions. The reason for fewer (11%) of the *unanswered* sessions found the needed information might be because if a participant did not find what they needed in an earlier session, it is still hard to find in a resumed session when they returned. In other words, the needed information may be hard to be found. The *satisfied* stopping reason is the most popular stopping reason for sessions resumed because of *updated(cultivated)* (44%), followed by *spawning* (34%), *unanswered* (33%), and *transmuting* (32%).

2) *Interruption* is popular for *transmuting* sessions (37%), followed by *unanswered*(22%), *spawning* (18%) sessions. *Interruption* is the only stopping reason for sessions resumed for *anticipated*, although there were only two anticipated sessions. *Interruption* was not found for *updated (cultivated)* sessions, which might be because for *updated (cultivated)* sessions, participants already know the information

resources in previous sessions; the session might be stopped after a short search, or ended before an interruption happened.

3) *Consult other information sources* was found for sessions resumed because of *unanswered* (22%), *updated(cultivated)* (11%) and *spawning* (5%), meaning that in these kinds of sessions, people may more often found information that they needed to check by consulting other information sources such as talking with other people.

4) *Cannot find the needed information* was only found for four sessions resumed because of *spawning* but only took a small chunk of the all stopping reasons for the *spawning* sessions (6%). *Cannot find the needed information* rarely happens only if a participant had very specific search goals for a session or had to find a specific piece of information that is hard to be found on the web. *Cannot find the needed information* is not a popular stopping reason for XSS. This might be because participants often have multiple goals for their searches, and there are many alternatives that they would be satisfied with.

5) Participants rarely stopped a session because of the *low task knowledge* for a task. Only a session resumed because *unanswered* was stopped because the participant felt they did not know what information was appropriate or when to stop. Given that many participants resumed sessions to look for background or domain information *transmuting* to help understand the task requirements or task topic. It is possible that during the middle of the search session, they still did not thoroughly understand the task requirements, which can lead to stop for low knowledge.

5.3 Session goal complexity, information types, and cognitive activities

In this section, we will focus on the level of complexity of session goals, the types of information people found during their sessions, and the associated cognitive activities users conducted when they interacted with the found information during the search processes, Then we analyzed the relationships between 1) the complexity level of the session goal and the information types, 2) the information types and the cognitive activities during a session, and 3) the complexity level of the session goal and the level of the actual cognitive activities involved during a session.

5.3.1 Session goal complexity

The *session goal complexity* refers to the cognitive complexity level of users' goals for an individual search session during the XSS process. We decide the session goal complexity by focusing on their descriptions of the goal they wanted to achieve and the possible cognitive activities they would conduct or how they would interact and use the found information.

We extracted our data from participants' responses to two questions: 1) their open-ended answers for the session goal in the *pre-session questionnaire*, and 2) their open-ended answers for their session goal in the *post-session questionnaire*. In cases when people described multiple goals for a session or their descriptions of session goals in the pre-session questionnaire differ from that in the post-session questionnaire, we determined the level of the session goal complexity by the higher level of the goal complexity. We used the A&K's complexity taxonomy to classify the session goal complexity into six categories: retrieve, understand, apply, analyze, evaluate, and create. We will further explain the definitions of these categories in the corresponding subsections.

Table 5.12 shows the overview of the session complexity levels of all 101 sessions for the 25 tasks. It shows that overall, the complexity of our participants' session goals clustered at the lower complexity level of *retrieve* (n=20) and *understand* (n=22) and the higher complexity level of *analyze* (n=24) and *evaluate* (n=26). A few of the sessions' goals were associated with *apply* (n=2) and the highest complexity level goal *create* (n=7).

Table 5.12: Summary of the levels of complexity of session goals during XSS

Complexity level	Initial Session	Middle Session(s)	Last Session	# of sessions
Retrieve	4	8	6	20
Understand	6	13	5	22
Apply	0	1	1	2
Analyze	6	15	3	24
Evaluate	7	13	6	26
Create	2	1	4	7
Total	**25**	**51**	**25**	**101**

Retrieve *Retrieve* is to look for specific information needed for a task. We determined sessions' goals at the *retrieve* level if a participant described the purpose of the associated session as 1) to look for information from certain website(s) they already knew before the search; 2) they knew exactly what information they wanted to find but did not know (or mention) where to find the information; and when 3) they know information sources to search for certain types of information. The focus of the goal is on the information or information sources. For instance,

> "I want to find two [name] shells, preferably a mint green one and a pink one. " (P118-S2)

> "My goal is to look at the website for the race that I am doing and see what is officially allowed to be worn. I am also hoping the race website will have recommendations for what to wear and bring." (P121-S1)

Understand Compared to *retrieval* session goals, the session goals at the *understand* levels may not have a clear idea about what the needed information would be or the information sources where it was from. On the contrary, the main purpose of the goal would focus on the purpose of using the found information to help understand some parts of the task instead of use the information itself (to solve the task requirements). For instance, a participant may describe that they would like to find some information to help understand a specific topic/subtopic related to the task or to figure out what the task required for. For instance,

> "My goals are to find some sources that I can look at in more detail later as well as to

get a general idea of the relationship between [name of two countries]." (P126-S1)

> "My goals are to get a general idea of that types of statistical sampling that are out there. I'm not entirely clear where this initial search will take me so my only goal is to orient myself to the topic." (P128-S1)

Apply The session goals at the *apply* level focuses more on using the information found from the session to solve certain problems related to the task. Compared to the *retrieval* and *understand* goals, session goals at the *apply* level involve details about the interaction with the information or how the information fits their tasks. Many of such goals could involve searching for information about "how to" work on something. For instance,

> "Find tutorials on how to draw landscapes, find and gather tutorials on three-dimensional modeling, find information regarding working in Photoshop and other Adobe suite apps and work on finding character design tutorials and other concepts. " (P114-S2)

> "I want to find out if there is a specific guide on how to file taxes as an LLC" (P116-S3)

Analyze Session goals at the *analyze* level often involve necessary cognitive activities to decompose the found information into smaller units. This could be because participants wanted to gain a deeper understanding of the topic components, or it could be because they wanted to conduct more complicated cognitive activities with the smaller units, such as considering the information based on specific criteria or re-grouping information components from different sources to generate new ideas, etc. For instance,

> "determine how Complete Streets policy might relate to the other policies I was looking at, finish my list of benefits, and grab a few compelling images." (P120-S4)

> "I hope to read about the modern political climates of [name of two countries] and hopefully find some more specific aspects of their governments which can be compared." (P126-S2)

Evaluate Session goals at the *evaluate* level aim to make decisions by comparing the found information against each other or to some other criteria to find out solutions for the task problem. It is also possible that a participant evaluates the task progress based on their information search process and the amount of information they collected. For instance,

> "I want to finish the entire task of creating a document for my friend. I also want to make a distinction between the apartments for cheapest, highest reviewed, lowest reviewed, and most expensive." (P102-S7)

> "Today, I wanted to find a pair of goggles to purchase as well as a race belt. To do this, I intended to research what features often come with these items that I might want. Once I learned about the features, I would find the best and most affordable option that met my needs." (P121-S3)

Create Session goals at the complexity of *create* is the most complex level of goals. It indicates that participants clearly know that they wanted to generate some new ideas, come up with new solutions or make new things that are based on but differ from the information they found through the session. For instance,

> "To make a plan for fun fall activities and to plan some fall-inspired outfits." (P103-S6)

> "I want to find some basic information and lay out my first outline draft for my presentation slides" (P125-S1)

5.3.2 Information types

To understand the information users found to support their task process, we asked participants to provide three information examples found during a search session and to describe how the information helped them for their tasks. During the introduction meeting, participants were encouraged to provide three examples for each session, but they could decide how many they wanted to provide and skip examples depending on their search experiences. Among the 101 sessions, we collected 279 examples.

As mentioned in Section 4.6, we used a pre-determined coding framework to code the types of the 279 information examples: factual, conceptual, procedure, and opinion. As we started to code, we found that some participants provided detailed information examples, whereas in some cases, participants provided examples of information sources instead, such as "a Website for Chef [name]" (P105-S1) "[State] Secretary of State - online business creation website" (P116-S2), or articles "Article on the power of storytelling in PR (Kent, 2015)" (P131-S2).

To analyze the information at the same level of granularity, we follow the following coding rules: (1) the basic rule is to code an information example using the four codes; (2) if participants provided an information source example, then we would code based on the information they gained from the information source by reading their description about how they thought the provided example was useful for their tasks; (3) if it still cannot be decided based on #1 and #2, then click the URL link associated with the example to distinguish what type of information was provided on the web page. Most of the examples can be coded by following these rules. In addition, we coded "none" if participants did not provide an example or explicitly say they did not find useful information during the session.

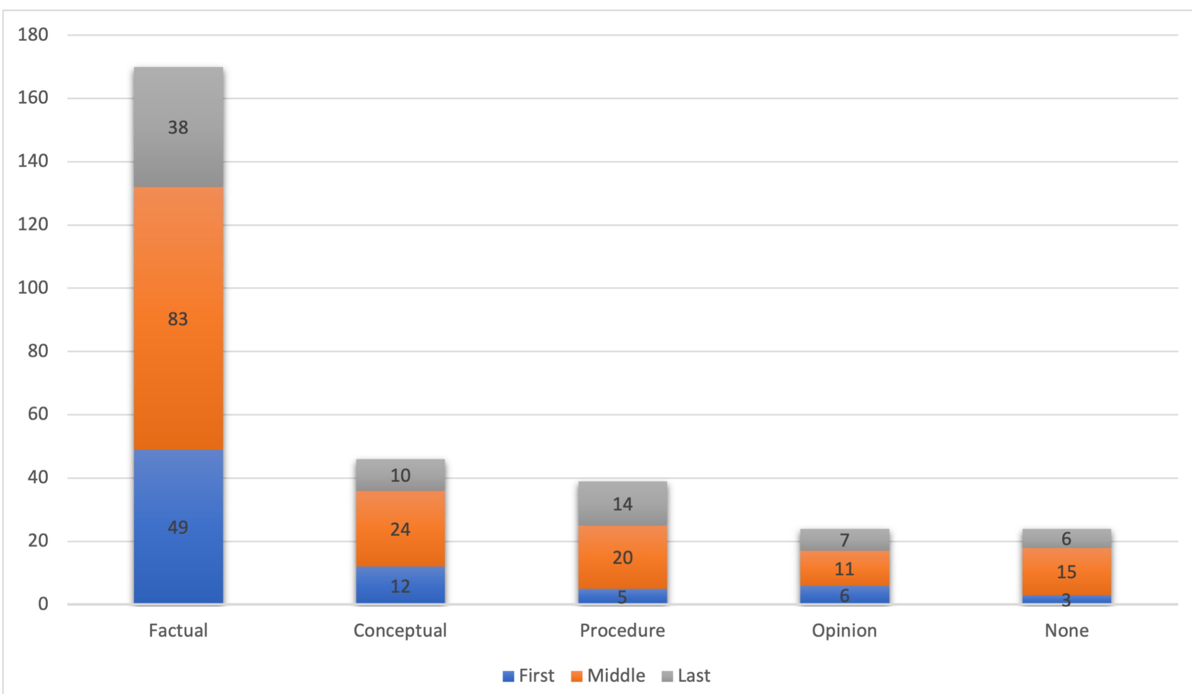


Figure 5.3: Summary of Information types participants found across all sessions

Except for 24 cases with no examples (i.e., none), our participants provided 279 useful information examples in total. Figure 5.3 shows a summary of the distribution of different information types across all search sessions, i.e., the first/initial sessions, the last sessions, and the middle sessions between the first and last for a participant.

Factual information Factual information includes basic information participants need to know about their task and to solve their problem, such as specific details like facts, numbers, terms, background information statements (e.g., the history of a topic, the context of the causes of a problem, or an introduction about the topic and related areas), as well as information about sources where to find relevant information.

Factual information is the most common information example compared to the other three types of information types among all examples and for each stage of sessions. The examples participants described include locations, contact information, stats, and description of physical objects (e.g., a book). For instance,

> " I saw the article on [name of website] last month and made a mental note. I revisited the article today to find out more about her [name of store]. I visited [name of website] to check the menu and get the address/phone number. I plan on paying them a visit personally and then pitching the idea of appearing on my show." (P105-S3)

> "General info surrounding the Unsung Founders monument...like the creator, dedication date and location" (P113-S5)

Factual information also includes statements about situations, rules, and descriptions of the process or events–the rules can be overlapped with conceptual information–factual rules are set up by human beings to help society to operate (not for specific individuals) while the conceptual rules are more fundamental about how things function in nature; Similarly, factual of the process more focused on some unique process that happened in the past, which is more like events; it does not include guiding information for how people to follow for a future event. For instance,

> "General information about how shared streets and open streets work in the US, which has increased during the pandemic" (P120-S1)

> “What to Wear for First Triathlon [required by the organizers]...This is another checklist, but it goes into more detail about the specific items. From it, I learned that I want to get a one-piece suit instead of a two-piece suit and that I do not need a wetsuit for my race.” (P121-S1)

The following example describes factual information about an event.

> “A description of the removal of a monument... I knew this to be outdated information because the [name] statue was removed in [date]. This information will be useful because I have taken note of the inaccuracy mentioned in my critique.” (P113-S3)

Factual information also includes information about information sources.

> “The [name of newspaper] article was the most helpful resource by far. It provided an in-depth story detailing the fundraising and location selection process for the monument. Although dated from [date], the information was very useful for describing the significance of the memorialization process in a tone that was clear and easy to understand. ” (P113-S5)

Conceptual information Conceptual information is about information of more abstract ideas, concepts, and the relationships among fundamental elements in an area (could be objective elements or abstract notions), like principles, structures, rules, and criteria in a specific area. As being defined by Anderson and Krathwohl (2001) that the conceptual rules are more about how the elements and structure work than the rules that are set up by humans for people to follow–the latter we classified as *factual information.*

For instance, P110 searched for information on a series of definitions, including audio definition, microphone definition, frequency definition, audio compression, etc. Participant P116 found information about “Professional Liability Insurance" (P116-S2) useful for their task.

Participants also found information about the relationships among key elements involved in their task topics. For instance, Participant P126 provided an example of “[name of country]n Protectionism in Agriculture” (P126-S3), and Participant P127 searched for “The dance style of the Peking Opera” (P127-S2). Participant P120 said “Slow streets and shared streets are interchangeable terms, added the benefit of reduced pollution with a policy ” (P120-S2)

Conceptual information could also be information about structure or framework and how the different elements work to achieve (or not) certain functions. For instance,

> "A technical report on the economic benefit of reducing VMTs" (P120-S5)

> "The Rise of Telemedicine... really illustrates how telemedicine has rapidly risen since the start of the pandemic and how both providers and patients have become more familiar with how it works. I also think that the quote about how telehealth has been accelerated by ten years during the pandemic will be useful and is a good introduction for our project" (P125-S2)

Participant P128 described that she learned about "Population definition and sampling frame" (P128-S1)

The conceptual information could also be hidden behind their representations rather than being expressed explicitly. For instance, the following example shows that the participants explored how different functional elements work of a specific tool to prepare for her project.

> "Graphic Vs. Parametric EQ... along the way of my searching, I was reminded that there are different kinds of EQ, the two primary being graphic and parametric. These types can't be left out in an EQ definition." (P110-S3)

Procedural information Procedural information is about "how to do" something or information about using specific skills or techniques. This is different from the conceptual information about "how something works functionally" (i.e., conceptual information) as conceptual information is about abstract concepts, whereas procedural information focuses more on specific skills or techniques. Anderson and Krathwohl (2001) explains that procedural knowledge often comes in the form of "a series of sequence of steps to be followed"[p.52]. From this perspective, procedural information differs from factual information about processing information (i.e., the process of an event) that already happened in the past and is not purposeful for people to follow. As knowledge is more coherent than information, here we use procedural information to address any sequential information about working on something or a piece of information, like brief instructions about a step or tips during a working process. Particularly, the information is created for people to follow in the same/similar situation or with the same purposes.

For instance, participants described useful procedural information found about sequential steps of actions, such as " long island ice tea recipe" (P103-S2), "ways to cancel evites" (P107-S4), " how to record in a nonstudio setting" (P110-S2). "Pinterest-collection of color palettes, brush ideas, and art" (P114-last)

Some procedural information is related to methods, policies, rules, and strategies of use of systems, and this information is generated purposefully to ask people involved to follow in specific situations when applied. such as: "The AirBnB house rules" (P103-S6), "Chicago style notation guidelines" (P113-S6) "LLC business registration information and documents" (P116-S1)

Others like:

> "EIN information from the [name of website]...I found this link which provided an easier way to apply for an EIN online as opposed to filling out the SS-4 form and sending out for approval." (P116-S3)

> "a tutorial on how to use clipping masks, which is a main feature of this app. I lack knowledge on this and plan on using it to gain greater skills in procreate. this should be helpful in bettering my skills " (P114-S6)

> "Adobe suite 3D rendering tutorial..[to]edit and create abstracted shapes within Adobe suits...I plan on using it for creating abstract shapes and further editing software " (P114-S2)

Procedural information also includes criteria or requirements to take action. For example,

> "generally require registration with the secretary of state... I have to research if registering in the state of NC is applicable" (P116-S1)

It should be noted that participants do not necessarily follow the procedural information they found. They may use it for different purposes at different task stages, such as gaining an overview of their task process or learning about an unfamiliar area. For instance,

> "This article provides an overview of the process of good storytelling, several "plots" that create good stories, and how to use storytelling in PR. "(P130-S2)

> “I started today’s task work by reviewing one of the sources I bookmarked the last search that included steps for writing a children’s book. Google Children’s Book and a phrase that describes my book. I thought that this was a great idea and decided to try it to see whether there were points from other books on my topic that include points I should remember when writing my own.” (P129-S2)

Another participant (P107) looked for NICU Care Packages because her sister gave birth to a premature baby. She said

> “ this website gave helpful hints of how to handle NICU babies, speak to parents and some wording that is helpful.” (P107-S4)

And participant P120 sought for an “Implementation guide for a Complete Streets policy" (P120-S4) to include in a proposal that would be submitted to a community board.

Opinion/ideas Information on opinions is about viewpoints or perspectives from others or a group of people. This type of information was not included in A&K’s taxonomy of knowledge types. But we included it here because our previous research on exploratory searches often found people sought others’ opinions or comments when they searched online. Opinions are about the human perspectives and feelings for almost everything ranging from the natural world, society, individuals, and events, as well as thoughts that come into their mind. Opinions include comments, conclusions, judgments, and feelings. Opinions could be objective, such as scientific research results or judgments made based on certain rules or criteria; opinions and ideas could also be subjective that is perceived based on an individual’s personal experiences. While opinions often coexist with other types of information, they are distinctly unique. For instance, a person might present facts, concepts, or describe a process before sharing their personal judgment. These underpinning elements serve to frame and contextualize the opinion, yet they are separate and distinct from the viewpoint itself.

Compared to other types of information, there were many fewer examples of opinions compared to the other three types of information. Our participants found useful opinion information ranging from product reviews to scholarly articles. For instance, “Negative reviews for [name of place]” (P102-S3), “Crepe is a shiny material” (P103-S1), “Scholarly article lending evidence to the economic benefit of open streets policy” (P120-S3), “an article about Best Triathlon Suits for Women” (P121-S1)

“Benefits of open streets policies” (P120-S1), “A research article on the argument for taking space from cars in urban areas” (P120-S4), “American perceptions of the effectiveness of telemedicine” (P125-S1), “Patient attitudes toward teledentistry” (P125-S2)

More examples are like:

> ‘Pros and Cons of Child Beauty Pageants..it provided in-depth reasoning and statistics for why Child Beauty Pageants are harmful to children’s physical and emotional health” (P112-S2).

> “Information on gender, cross-dressing, and the Peking Opera...This article did a good job at explaining how the actors and roles of the Peking Opera were viewed by the Chinese since the 1600s” (P127-S2).

Specifically, we also include creative ideas as a subcategory of opinion. It describes information/knowledge generated by others. The information is about the creators’ view or products of their mental process based on the information they knew of (e.g., factual, conceptual, procedure, ideas). An example of creative ideas information is the literature content, such as novels, poems, and lyrics. For instance, participants described that they found information like “Poem about Icarus’ Fall", ”monologue”, “Thee Odyssey” and "Outro: Tear lyrics (translated)" (P108-S1 to S3)

5.3.3 Cognitive activities

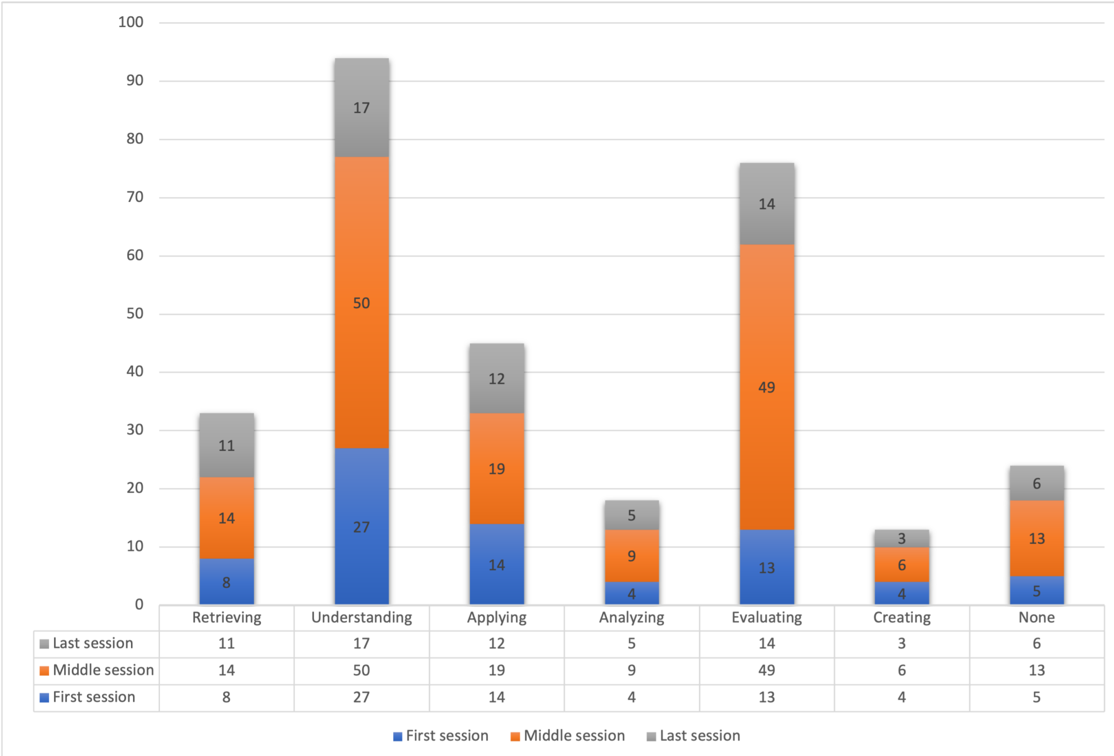
We are also interested in the cognitive activities participants actually conducted as they processed the information through the search processes and task processes. Cognitive activity here is the mental actions or processes about how people think about the information found that can be used to complete their tasks. The data were extracted from their responses to the post-session questionnaires regarding their thought process for the information examples on how the information is useful for their tasks.

Using the same coding framework of the A & K taxonomy, we coded participants’ cognitive activities into six categories: retrieving, understanding, applying, analyzing, evaluating, and creating, following the same levels of cognition complexity. Here we use the *-ing* to indicate the action for distinguishing the activity from the goal complexity level, which just used the verb. During the coding process, Some of the cognitive activities can be directly extracted from participants' descriptions, such as "I found..." (*retrieving*), "It helps me understand..." (*understanding*), "I used this as an example..." (*applying*). Other times, we synthesized participants' activities based on the processes they reported in the description of the sessions' search process. For instance, " I used ...as a keyword to search..." (*retrieving*), "this means...." (*understanding*), "unfortunately many are not within my desired time frame" (*evaluating*), "I used all of this information in order to write the introductory paragraph of my paper" (*creating*).

According to the original taxonomy, the six cognitive activities are associated with different levels of cognitive complexity but not exclusive from each other: Some higher levels of cognitive processes are built upon the processing results of lower-level cognitive processes. For instance, retrieval of the information in need is the foundation for almost all other cognitive activities, from understanding to creating. On the other hand, it should be noted that the higher-level cognitive activities do not necessarily mean participants conducted all lower-level complexity cognitive processes. For instance, evaluating is based on retrieval and understanding but may include applying process. Consequently, multiple types of cognitive activities might be associated with the same information item. To avoid confusion for analysis, we selected the cognitive activity with the highest complexity level for the same information example when a participant described multiple types of cognitive activities for the same information example.

It should be noticed that this is different from the session goal complexity, where the session goal complexity is users' perceptions about session complexity before they start a session. The selected cognitive activity from the provided examples is the complexity level of actual cognitive activities that happened during a session. In addition, because each task has only one initial session and one last session but at least one middle session; as a result, the overall middle stage accounted for over half of all the cognitive activities (147/279).

Figure 5.4 provides an overview of the coding results.



	Retrieving	Understanding	Applying	Analyzing	Evaluating	Creating	None
Last session	11	17	12	5	14	3	6
Middle session	14	50	19	9	49	6	13
First session	8	27	14	4	13	4	5

Figure 5.4: Summary of the highest-level cognitive activities of information examples participants described for all sessions and at different stages.

There were 24 cases with no examples of information or any descriptions of cognitive activities. For the 279 information examples, we identified the cognitive activities at different levels, including: 33 examples of retrieving, 94 examples of understanding, 45 examples of applying, 18 examples of analyzing, 76 examples of evaluating, 13 examples of creating.

The results (Figure 5.4) show that overall, *understanding* is the most common cognitive activity participants described in the examples when they interacted with the found information during search processes. *Evaluating* is the second most popular higher-level complex cognitive activity, followed then by applying, retrieving, and analyzing. *Creating* is the least common example we found across all sessions. It also shows that the six different levels of cognitive activities spread across all three stages of search sessions, from the first search session, and middle search sessions to the last search sessions. Chi-Square tests did not find a significant association between the sessions' most complex cognitive activity and the task stages (x^2 (10, 279) =10.751, p=.377) (Table 5.13). No particular type of cognitive activity is only associated with a specific task stage (Figure 5.5).

Table 5.13: Chi-Square tests for the relationship between the session's most complex cognitive activity and the task stage

	Value	df	Asymptotic Sig. (2-sided)
Pearson Chi-Square	10.751	10	.377
Likelihood Ratio	10.697	10	382
N of Valid Cases	279		

Cog * Stage Crosstabulation

Count

		Stage			
		First	Last	Middl	Total
Cog	Analyze	4	5	9	18
	Apply	14	12	19	45
	Create	4	3	6	13
	Evaluate	13	14	49	76
	Retrieval	8	11	14	33
	Understan	27	17	50	94
Total		70	62	147	279

Figure 5.5: The crosstab results of Chi-Square tests between the session's most complex cognitive activity and the task stage

Figure 5.6 shows the percentage of different types of cognitive activities across the task stages. It can be seen that *understanding* and *evaluating* are the two most common cognitive activities across different task stages. Another two higher complexity level cognitive activities *analyzing* and *creating* are about same across different task stages. Particularly, *creating* can be found from some initial search sessions (6%) at the beginning of the search stage, indicating that although not common, some participants started to create new ideas at early task stages involving XSS. Figure 5.6 also suggests that evaluating was a larger percentage of the middle sessions (33%) than it was for the first (19%) or last (23%) sessions. And it looks like applying was lower in the middle session (13%) than in the first sessions (20%) or for the last sessions (19%). Retrieving was also a bit higher in the last sessions (18%) than it was in the first sessions (11%) or the middle sessions (10%).

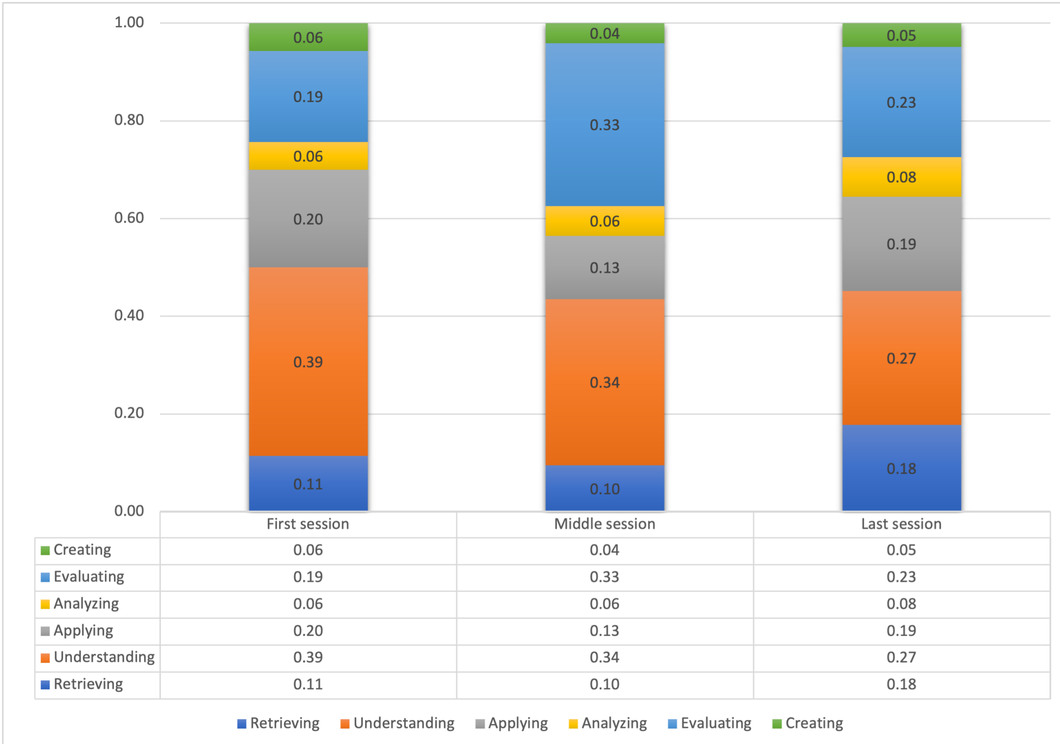


Figure 5.6: The percentage of different cognitive activities happened across the three task stages

The results also provide clear, strong evidence that besides searching activities, participants conduct many other mental activities to process the found information synchronously as they search online.

Retrieving Retrieving describes participants' actions of looking for specific information or information sources related to their tasks. We applied this code to examples when participants only described what they found through the session without further explanations about how they processed the information. It should be noted that the cognitive activities we coded here are purely based on the participants' description of their activities during the session described in the corresponding post-session questionnaire, and if there were multiple cognitive activities associated with one specific information example, we selected the most complex activity for that information example. For example, Figure 5.6 shows that in 18% of the information examples provided from the last sessions, participants did not go higher than the retrieval cognitive activity level. It is possible that participants might also conduct other/multiple cognitive activities for the same piece of information during and after the search session. This rule applied to all other cognitive activity examples during our coding process.

Compared to other cognitive activities, retrieving is the fundamental step for all other higher-level of cognitive actions since participants need to find the needed information first. Therefore, it is not surprising that participants retrieved all types of information. The information could be facts, details, a specific document/article, or a tool that would be used in the task. Participants may or may not have clear thoughts about what they wanted to find, but through the search process, they collected some information that might be useful for their tasks. That means retrieving activity is not just for a session with retrieval goals. But it is a general activity. In addition, in our case, we also include re-retrieval, which happens when participants are searching for information they knew previously. For instance,

> "I wanted to find the community amenities for my friend in case she was ever interested in using them. She didn't express any interest in a community amenity in general, but I still wanted to include them. One specific piece of information I had trouble finding was whether or not this complex had a fitness center. At first, I didn't think they had one based on the list of amenities provided. However, when I looked for photos of the complex, later on, I saw photos of a gym. I then checked apartments.com and confirmed that there is a fitness center on the property. " (P102-S3)

> "Honestly, I just clicked on this link so I could get the image of Squidward. I have no interest in the blog about oil painting in fresco, but the art was funny and interesting. all I wanted to was to isolate the image, and instead, it brought me to the blog. " (P114-S3)

There are also cases when participants conducted searches, but they did not find what they needed. Similarly, this could happen at any session across the initial, middle, and last sessions. For instance,

> "I thought that [name of retailer] might have cute pajamas because I have seen them have nice sweatpants before. I went to look at their loungewear section, but nothing fit the aesthetic that I was looking for. I decided to try another site." (P103-S6)

> " I was thinking that I might be able to discover possible summer internships or consultant positions that might be happening locally or within the state that might fit my field of

> interest. It was somewhat useful. Nothing particularly appealing to me or helpful for what I am hoping for." (P104-S1)

Participants might also refind information they interacted with (i.e., saw, heard) previously. For instance,

> "This is exactly the article that I'd heard about on a podcast a while ago. I never would have found it otherwise; I've never heard of [name of website] in any other context. I really appreciate that she put an updated date at the top of the article since this process changes regularly. " (P106-S5)

> "Here I returned to the information I had originally encountered in the beginning when I was using the Wikipedia site. I couldn't remember if I had looked at this information on the [name of website], so I reviewed it again. Then I realized I had reviewed it and went to the sidebar, which had the table of contents and went on to another topic." (P128-S6)

Understanding Understanding describes participants' behavior when they actively construct meaning from the retrieved information in various meanings, especially in their task context, such as interpreting, exemplifying, classifying, summarizing, inferring, comparing, and explaining, etc. For instance, a participant might provide a summary of the information they learned from a specific information source; they would interpret the found information from their task perspective or based on their own past experience, and identify connections between the found information and their task requirements.

Understanding activity is more complex than retrieving because people take further steps to interpret the information and infer the relationships between the found information and their own tasks. On the other hand, different from higher-level cognitive activities, participants did not describe cognitive steps like using, analyzing (i.e., breaking down), or making judgments about the found information with the examples at the understanding level. For instance,

> "The webpage describing the [name of person] monument did not provide much information on his actual life, so I searched elsewhere for a more in-depth biography. Contextualizing his life benefitted my understanding of who exactly is being commemorated." (P113-S3)

> "I think this information will be useful because it really illustrates how telemedicine has rapidly risen since the start of the pandemic and how both providers and patients have become more familiar with how it works. I also think that the quote about how telehealth has been accelerated by ten years during the pandemic will be useful and is a good introduction to our project." (P125-S2)

Participants' mental models could be changed based on the found information. For instance,

> "This is critical information to know because I would have assumed that [name of country] only sought independence once, but this is not the case. [name of country] achieved independence for a short period of time but was eventually forced into [name of country] and occupied by the [name of army]. It eventually regained independence but since [name of country] was occupied, it makes sense that [name of country] would be weary of [name of country], fearing another occupation." (P126-S1)

> "This section was helpful because it prompted the thought of how many words could keep the attention span of my age group. It also made me think about the fact that I still need to identify an age group because doing so now would probably help craft my searches better." (P129-S2)

In some cases when people did not find useful information, there were a couple of reasons, such as being distracted, did not have a specific goal for the session. For instance,

> "I didn't find this information useful at all this time around because I was so distracted. First my mind was not focused on the information, second I decided to email my professor which took me a while... I also felt very tired and that made it hard to stay on task and focus to comprehend that information I was reading." (P128-S2)

Applying Applying describes situations when participants use the found information in given situations (i.e., tasks) or carry out the steps by following what participants have learned from the found information to solve part or the entire task. During the coding process, we used the code *applying* when participants described some next steps after they explained their understanding of the meaning of the found information, which is also the key point to distinguish *understanding* and *applying*. We coded examples as *applying* when a participant described how to use the information as elements, examples, pieces of evidence, and "building blocks" for their tasks.

Participants used the information for all different purposes, including: 1) as an initial point to start their task or for inspiring new ideas, for instance, 2) as supportive examples, details for the final outcome, for instance, 3) to help make decisions, for instance, 4) as instructions/guidelines (i.e., procedural information) to work on the task. For instance,

> "This seems particularly important to review before I start writing. I want to approach this in the best possible way to the best of my ability, so this seems like a great place to start. as I write, I'll continue to research more using the links I find in this research as a starting point. " (P101-S2)

> "My search on "Facebook Marketplace" unfortunately didn't help me find the two originally desired products, however, it showed me 6 separate listings of similar [name of product] products from the same era, all selling for a price far lower than what I know to be market standard off the top of my head. All 6 listings are in the bookmarks folder. I messaged all 6 sellers on my personal account, and if any of them respond, I will do more specific research on the comparative prices of each newly desired product." (P123-S1)

> "I found this information useful because it shows a clear, statistical difference between the amounts of people who used telemedicine options before the pandemic versus after the pandemic. I will use this information to illustrate the rapid rise of telemedicine and how it is much more common now than it was at the beginning of 2020. " (P125-S1)

Analyzing Analyzing is defined as the activity when people mentally break down information into parts and determine how the parts relate to one another and to the overall problem situation or the task purpose by differentiating, organizing, attributing, etc.

We found the following examples:

> "This site lists the official transportation options and prices from the [name of airport]. It was useful but didn't answer everything. I liked that it showed me exactly how much a taxi would cost. But their links to shuttle services are confusing and outdated. And no real info on how Uber or Lyft work at the airport." (P106-S3)

> "After seeing the amazing brick-based [toy] listing on FB-Market, I went to [name of site] to breakdown the price and see how much of a deal I was getting; turns out pretty good. [name of site] broke the lot down by individual set, figure, and part cost, showing me that I got $80 worth of a lot for the low cost of $25. [name of site] is great for accurate cost breakdown comparisons." (P123-S3)

Evaluating Evaluating is the activity when people make judgments about the found information based on certain criteria and standards related to their own task, such as checking, criticizing, monitoring, testing, or coordinating. Participants evaluated various things, including 1) the information quality, 2) the amount of information on specific topics or aspects, 3) alternative task solutions, or 4) one's own thoughts about the task requirements based on the found information. And their criteria might be related to the information they found from other sources, their previous knowledge of the topics, or some criteria that were developed through the search process for the tasks. For instance,

> " For this session, I actually had to think a little more deeply about what reviews to pick from those listed on Google Reviews. I noticed a large majority of the negative reviews were from 5+ years ago, which seemed a little too far in the past to include in a relevant list of reviews. So, I narrowed my criteria down to anything 3 stars or less within the last 2 years. I tried to pick the most detailed reviews from this criteria in order to provide my friend with the most thorough information." (P102-S3)

> "While still considered secondary, the website [name of site] is far more mainstream and "official," giving me a good idea of what typical realistic listing prices look like. Thanks to this information, I was able to deem that the eBay prices for [name of product] were unreasonable, but I found a potentially pretty good deal for [name of product]. As described previously, the [name of site] listing I found came out to be about $30-$40 cheaper than the next cheapest option on [name of site]." (P123-S2)

Creating Creating describes the activity of putting the found information together by generating, planning, designing, and producing to form something new and coherent, such as a new pattern or structure as a functional whole. In our analysis, creating activities happened when participants generate new thoughts, solutions, and ideas by adapting the found information to fit the specific requirements of their own tasks. In other words, they did not apply the information exactly as it was but adjusted it in some ways through creative thinking processes.

The found information was creatively used as 1) components as part of participants' generated content or product, 2) inspirations for novel ideas or directions to follow in later task progress, 3) functional tools to help generate new ideas or products, etc.

In some examples, creating activity is often performed together with other cognitive activities ranging from retrieving, understanding, analyzing, and evaluating. In Anderson and Krathwohl (2001), creating can also be seen as a special way of applying the found information in novel ways that may not be the primary purposes of the original information as "implementing"[p.78].

> "This was the most useful page I found in this session, as it explained the goals and intentions behind the website's creation. It was created in order to provide an accessible resource relating to historical monuments of [name of state]...I used all of this information to write the introductory paragraph of my paper, ending it with a thesis statement based on my opinion of my website." (P113-S2)

> "[Name of web tool] is both simply amazing and frustratingly annoying. I created the layout and template, uploaded photos, and made everything nearly perfect and then it crashed and didn't save any of it! I'll have to use it later and redo it.... " (P107-S4)

> "This is helpful information because it gave me a list of questions to prompt my creation of my character. Reading this section made me think about the importance of distinguishing my main character in a special way. I plan to review some books with a similar topic to see what unique ways other authors have distinguished a main character." (P129-S2)

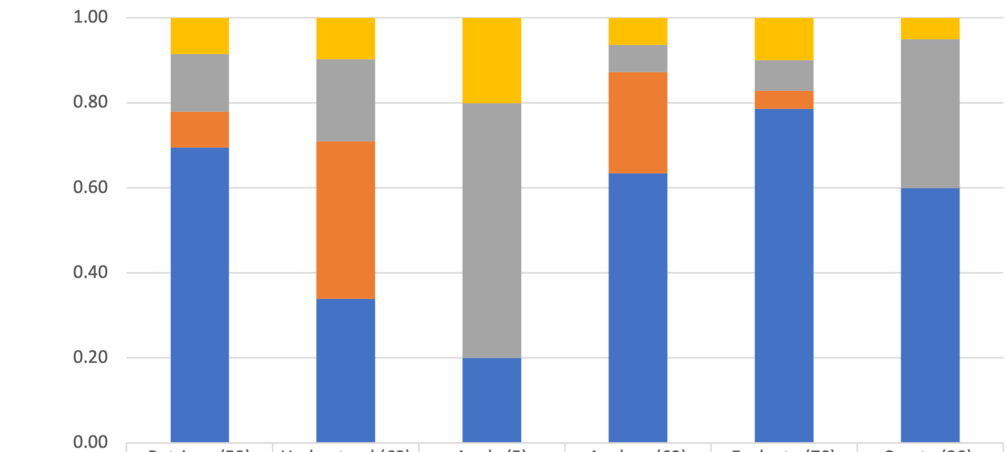
5.3.4 Session goal complexity and information types

For this question, we wanted to see whether there are relations between the complexity levels of a session's goal and the information types participants found useful during XSS.

Table 5.14 and the associated Figure 5.7 provide descriptions of different information types for the associated levels of session goal complexity.

Table 5.14: The number of different information example types associated with different session complexity levels

Session complexity \ info. types	Factual	Conceptual	Procedural	Opinion	**Total**
Retrieve	41	5	8	5	59
Understand	21	23	12	6	62
Apply	1	0	3	1	5
Analyze	40	15	4	4	63
Evaluate	55	3	5	7	70
Create	12	0	7	1	20
Total	170	46	39	24	279



	Retrieve (59)	Understand (62)	Apply (5)	Analyze (63)	Evaluate (70)	Create (20)
Opinion	0.08	0.10	0.20	0.06	0.10	0.05
Procedural	0.14	0.19	0.60	0.06	0.07	0.35
Conceptual	0.08	0.37	0.00	0.24	0.04	0.00
Factual	0.69	0.34	0.20	0.63	0.79	0.60

Note: the number in a cell indicates the percentage amount of a type of information participants retrieved for the sessions whose goals are at a specific complexity level. For instance, 0.69 represents that among all the 59 information examples found for sessions with the goals at the retrieve level, 69% of 59 (n=41) of them are factual information.

Figure 5.7: The percentage of information types for the session goals with different complexity levels.

A Chi-Square test (Table 5.15) shows that the session goal complexity has effects on the information types participants searched during the sessions ($\tilde{\chi}^2$ (15, 279) = 65.765, p<.001).

Table 5.15: Chi-Square test for the relationships between session goal complexity and the information types.

	Value	df	Asymptotic Sig. (2-sided)
Pearson Chi-Square	65.765	15	<.001
Likelihood Ratio	65.798	15	<.001
N of Valid Cases	279		

To further explore the relationships between the session goal complexity and the information types, we analyzed the proportions of distribution of different information types for session goals at different complexity levels. Figure 5.7 shows that: *Factual* information can be found from all six levels of session goal complexity. In addition, it is the most popular useful information type, especially for session goals at the *evaluate* (79%), *retrieve* (69%), *analyze* (63%), and *create* (60%) levels, but not quite popular for session goals at the *apply* (20%) level. The reasons why factual information is very popular for *retrieve* and *evaluate* sessions might be because participants' purposes for retrieve sessions mainly focused on factual information. Meanwhile, when their session purposes are about evaluation, they might also concentrate on comparing and assessing factual information.

Conceptual information can be found for four complexity levels of session goals, except the session goals at the *apply* and *create* levels. It is popular, especially for session goals at the *understand* (37%) level, followed by that of session goals at *analyze* (24%) level. It was found but not quite often for session goals at the *retrieval* (8%) level and *evaluate* (4%) level. The reason we did not find conceptual information for sessions whose goal is to create might be because, during the creative stages of a task, participants already have thorough knowledge about the topic areas, and they no longer need conceptual information to help understand the basic elements related to their tasks.

Procedural information can be found for all six levels of session goals. It is very popular for sessions with goals at the *apply* (60%) and *create* (35%) levels, followed by that of session goals at the *understand* (19%), *retrieve* (14%), *evaluate* (7%), and *analyze* (6%) levels. It indicates that sessions aimed at using the information found to solve problems or generate new products often need information about processes or skills to work on something.

Opinion information is only a small part of useful information but appeared among all six goal complexity levels: *apply* (20%), *evaluate* (10%), *understand* (10%), *retrieve* (8%), *analyze* (6%), and *create* (5%). It indicates that across all different complexity levels of session goals, participants paid attention to others' perspectives (either subjective or objective) as they searched for other types of information. However, unlike other types of information, opinion is rarely found as a main type of information for session goals at any complexity level.

5.3.5 Session goal complexity and cognitive activities

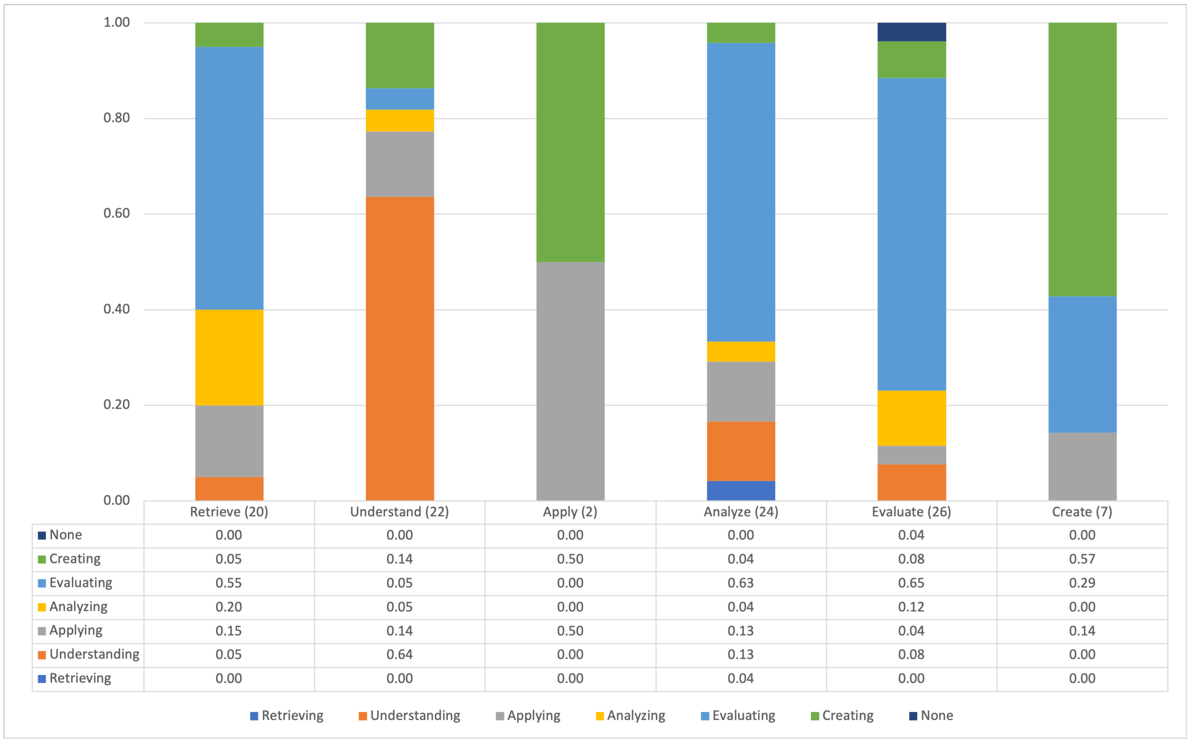
Chi-Square tests were performed to analyze the relationships between the session goal complexity and the representative cognitive activity (i.e., the most complex cognitive activity conducted through a search session) of the session. Table 5.16 shows that there was a significant association between the session goal complexity levels and the highest level of cognitive activity carried out during the sessions ($\tilde{\chi}^2$ (30, 101) = 73.552, p<.001). It means that the complexity levels of session goals affect the cognitive activities conducted during the sessions.

Table 5.16: Chi-Square test for the relation between session goal complexity and the most complex cognitive activity of the session

	Value	df	Asymptotic Sig. (2-sided)
Pearson Chi-Square	73.552	30	<.001*
Likelihood Ratio	63.817	30	<.001
N of Valid Cases	101		

Table 5.17: The number of highest cognition for sessions with different complexity levels.

Goal complexity \ Highest cog.	Retrieving	Understanding	Applying	Analyzing	Evaluating	Creating	none	**Total**
Retrieve	0	1	3	4	11	1	0	20
Understand	0	14	3	1	1	3	0	22
Apply	0	0	1	0	0	1	0	2
Analyze	1	3	3	1	15	1	0	24
Evaluate	0	2	1	3	17	2	1	26
Create	0	0	1	0	2	4	0	7
Total								101



	Retrieve (20)	Understand (22)	Apply (2)	Analyze (24)	Evaluate (26)	Create (7)
None	0.00	0.00	0.00	0.00	0.04	0.00
Creating	0.05	0.14	0.50	0.04	0.08	0.57
Evaluating	0.55	0.05	0.00	0.63	0.65	0.29
Analyzing	0.20	0.05	0.00	0.04	0.12	0.00
Applying	0.15	0.14	0.50	0.13	0.04	0.14
Understanding	0.05	0.64	0.00	0.13	0.08	0.00
Retrieving	0.00	0.00	0.00	0.04	0.00	0.00

Note: the number in a cell indicates the percentage of a cognitive activity conducted for sessions whose goals are at a specific complexity level. For instance, 0.05 represents that, among the 20 sessions with goals at the retrieve level, 5% (n=1) of the sessions' most complex cognitive activity is creating.

Figure 5.8: The percentage of the highest complexity level of cognitive activities for session goals with different complexity levels.

Figure 5.8 shows that:

1) *Retrieving* as the most complex cognitive activity was only found in the session whose goal was at the *analyze* (4%, n=1) level. This might be because during that session, the participant only found the information but did not further process the information for the task or because they did not process the provided example or provide examples of other activities (e.g., explaining how they used the information). On the other hand, it indicates that *retrieving* is rarely the only cognitive activity for a session, no matter the complexity level of the session goals. Even for sessions whose goals are at the *retrieve* level, participants conducted more complex cognitive activities than the complexity level of their goal indicated. For example, during the sessions with goals at the *retrieve* level, the most popular cognitive activity for these sessions is *evaluate* (55%), and the least complex cognitive activity is *understanding* (5%).

2) *Understanding* as the most complex cognitive activity can be found for four complexity levels

of session goals *understand* (64%), *analyze* (13%), *evaluate* (8%), and *retrieve* (5%). It is very common for sessions with goals at the *understand* level compared to those at other complexity levels. It is also the most common cognitive activity for the *understand* sessions compared to other cognitive activities conducted for these sessions (64% v.s 14%, 14%, 5%, 5%)

3) *Applying* as the most complex cognitive activity can be found for sessions at all six different complexity levels: *apply* level (50%), followed by sessions with goals at *retrieve* (15%), *create* (14%), *understand* (14%), *analyze* (13%), *evaluate* (4%) levels, from the most popular to the least. It indicates that for sessions at all complexity levels, participants could find some information that can be applied to their tasks. It is interesting that the sessions with goals at apply level only involve applying and creating as the most complex cognitive activities. One of the reasons probably is only a few sessions with the goal of applying; secondly, apply sessions focus on executing and implementing, which involves creating that adapts the found information to solve one's own task.

4) *Analyzing* as the most complex cognitive activity can be found for sessions with goals at the complexity levels of *retrieve* (20%), *evaluate* (12%), *understand* (5%), *analyze* (4%). It indicates that within these kinds of sessions, participants often found information that they need to break down into smaller units. On the other hand, it could also mean that during these sessions, participants often looked for more complex information than the basic units or complex structure that needs the further process to be used. *Analyzing* is not quite common for sessions with goals at the *understand* level (5%). It might be because for sessions with goals at the *understand* levels, people purposefully searched for information at a lower level of granularity–they already broke down what they needed into smaller units before they searched for the information. The reason why analyzing is not common as the most complex cognition either for sessions with goals at the *analyze* level might be because it often was the pre-step for evaluation. In other words, many analyzing activities during evaluation sessions could be transferred to evaluation.

5) *Evaluating* as the most complex cognitive activity can be found in sessions with goals at five complexity levels, except sessions with goals at the *apply* level. *Evaluating* is the most popular highest complex cognition for sessions at *evaluate* (65%), *analyze* (63%), *retrieve* (55%), *create* (29%) levels. It is least common in sessions with goals at the *understand* level (5%). It indicates that sessions with goals of *retrieve, analyze* and *evaluate, create* involved a lot of evaluating activities.

6) *Creating* as the most complex cognitive activity can be found in sessions with goals at all six

complexity levels: It is popular for sessions with goals at the levels of *create* (57%), *apply* (50%), *understand* (14%). It is not quite often for sessions with goals at the levels of *evaluate* (8%), *retrieve* (5%), and *analyze* (4%). It means that for sessions with goals of *create*, people indeed conducted creative activities quite often than sessions with goals at other levels. Meanwhile, it also indicates that *create* could be an important part when people apply the found information to their tasks.

5.3.6 Session goal complexity and resuming and stopping reasons

Session goal complexity and resuming reasons As shown in Table 5.18, Chi-Square tests did not find statistical significance for the relationship between the complexity of session goals and the session resuming reasons ($\tilde{\chi}^2$ (20, 101) = 21.156, p=.388).

Table 5.18: Chi-Square test for the relation between session goal complexity and the session resuming reason

	Value	df	Asymptotic Sig. (2-sided)
Pearson Chi-Square	21.156	20	.388
Likelihood Ratio	22.883	20	.295
N of Valid Cases	101		

The percentage of different resuming reasons associated with each complexity level of session goals is provided in Figure 5.9. We can find that:

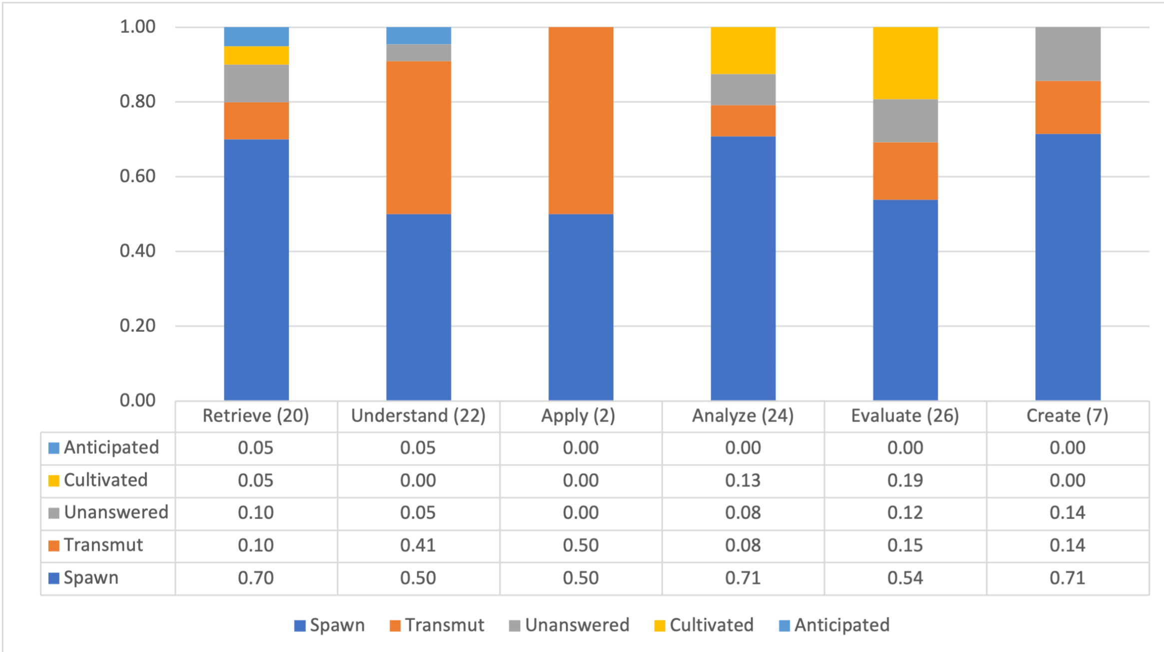
Spawning, the overall most popular session resuming reason, is also the most common session resuming reason for sessions whose goals are at each complexity level: *create* (71%), *analyze* (71%), *retrieve* (70%), *evaluate* (54%), *apply* (50%), and *understand* (50%).

Transmuting resuming reason can also be found across all six complexity levels of session goals. However, it is more popular for sessions with goals at the *apply* (50%) and *understand* (41%) levels. On the contrary, it is not quite common for session goals at other complexity levels like *evaluate* (15%), *create* (14%), *retrieve* (10%), and *analyze* (8%). It indicates that during sessions with goals at the *understand* and *apply*, participants often resumed the sessions to look for information to help them understand background knowledge or task requirements since it can help them to understand the task and know better about how to apply the information to solve their task problems.

Unanswered/incomplete resuming reason can be found for five levels of complexity of session goals including *create* (14%), *evaluate* (12%), *retrieve* (10%), *analyze* (8%), and *understand* (5%). It means that it is more common for participants who did not find what they needed and resume sessions to create, evaluate, and retrieve. However, it is not quite common for sessions with the goal at the *understand* level might be because participants often found some alternative information to achieve their *understanding* goal.

Cultivated/updated resuming reason can be found for sessions with goals at the level of *evaluate* (19%), *analyze* (13%), and *retrieve* (5%). It is not found for sessions with goals at the levels of *understand*, *apply*, or *create*. This might be because, at the beginning of understanding the task topic, participants still looked for information to help them become familiar with the task topic. Time may not be the top priority of their search. On the contrary, when it comes to generating new ideas or creating novel products, they may already find the latest relevant information and no longer need to check for updated information.

Anticipated as the least popular session resuming reason overall, there were only two sessions associated with it, one is at the *retrieve* level, and the other is at the *understand* level. Because sessions resumed/started for the *anticipated* reason, aim to look for information that might be useful in the future as their tasks move forward. The information was not required as a necessary part to complete the current task. The main intent of using the information is for inspiration.



Note: the number in a cell indicates the percentage of a type of session resuming reason for sessions whose goals are at a specific complexity level. For instance, 0.70 means that, among the sessions with goals at the retrieve level, 70% of the sessions were resumed because of anticipated.

Figure 5.9: The percentage of different session resuming reasons for session goals with different complexity levels.

Session goal complexity and stopping reasons Next, we analyzed the relationships between the different levels of session goal complexity and the session stopping reasons for the associated reasons.

As shown in Table 5.19, the results of Chi-Square tests show that there are no statistically significant associations between the complexity level of session goals and the session stopping reasons ($\tilde{\chi}^2$ (25, 101) = 26.970, p=.357).

Table 5.19: Chi-Square test for the relation between session goal complexity and the associated session stopping reason

	Value	df	Asymptotic Sig. (2-sided)
Pearson Chi-Square	26.970	25	.357
Likelihood Ratio	30.154	25	.219
N of Valid Cases	101		

The percentage of different session stopping reasons for session goals with different complexity levels is summarized in Figure 5.10. It shows that:

1) *Goal achieved*, as one of the most popular session stopping reasons, can be found across session goals at all six different complexity levels. Among them, it is the most popular one for session goals at the *create* level (57%), followed by sessions whose goals at *apply* (50%), *analyze* (42%), *evaluate* (38%), *retrieve* (30%), and *understand* (14%). It is interesting that on average, the *goal achieved* is more popular for sessions with the goal complexity at higher levels (i.e., *create, evaluate, analyze, apply* than lower levels (i.e., *retrieve, understand*.

2) *Satisfied*, the other most popular session stopping reason, can be found for sessions with goals at five complexity levels: *retrieve* (45%), *understand* (41%), *evaluate* (35%), *analyze* (25%), and *create* (14%). Overall, sessions with goals at lower-complexity levels (i.e., *retrieve, understand* are more often associated with *satisfied* than the sessions with goals at the higher-complexity levels (i.e., analyze, evaluate, create). Together with the *goal achieved* as a more popular reason for sessions with goals at the higher-complexity levels, it would be interesting for future studies to explore the reasons behind them.

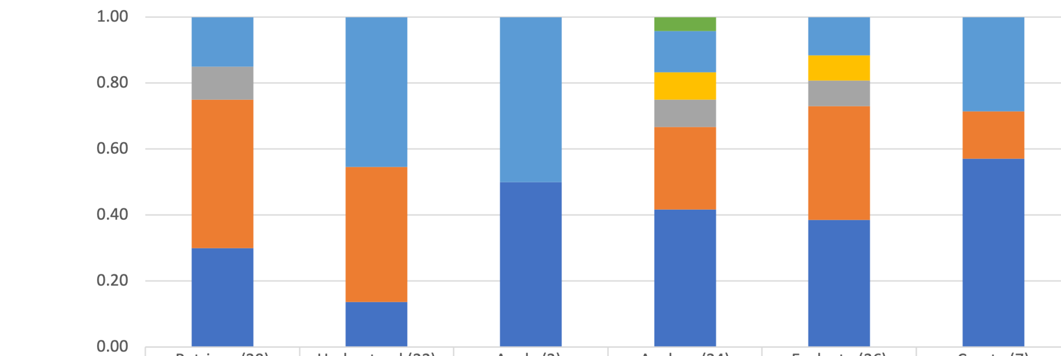
3) *Consult* as the session stopping reason were only found for sessions with goals at three complexity levels: *retrieve* (10%), *analyze* (8%), *evaluate* (8%). Given that the previous findings (Figure 5.7) show that sessions whose goals at the *retrieve, analyze, and evaluate* were associated with the most factual information examples, there might be some relations between the information type (i.e., factual information) and the reasons why people need to consult others, which requires further exploration. In addition, it might also be because that, for many analysis and evaluation sessions, participants were nearing a decision point and needed to consult with others.

4) *Cannot find the needed information* was only found as the stopping reason for sessions with goals at two complexity levels: *analyze* (8%), and *evaluate* (8%).

5) *Interruption* is a common stopping reason. It can be found across sessions with all six complexity levels: *apply* (50%), *understand* (45%), *create* (29%), *retrieve* (15%), *analyze* (13%), and *evaluate* (12%). It means that any sessions at any complexity level could be interrupted–either by self-interruption or by external interruptions. Interruption is often for sessions with goals at the *understand, apply, and create* levels than the other three. It is hard to know whether this is because the process of understanding, applying, and creating are easier to be interrupted or because

participants were more tolerant of interruptions during these sessions.

6) *Low task knowledge* is the least common session stopping reason and was only found for one session in our sample whose goal is at the *analyze* complexity level. This might be because as the participant dived deeper into the task topic by breaking down the information into lower levels or directions they did not aware of, they then found that it was hard to continue their progress without going back to reconsider their understanding of the task topic and requirements.



	Retrieve (20)	Understand (22)	Apply (2)	Analyze (24)	Evaluate (26)	Create (7)
Low knowledge	0.00	0.00	0.00	0.04	0.00	0.00
Interruption	0.15	0.45	0.50	0.13	0.12	0.29
Cannot find	0.00	0.00	0.00	0.08	0.08	0.00
Consult	0.10	0.00	0.00	0.08	0.08	0.00
Satisfied	0.45	0.41	0.00	0.25	0.35	0.14
Goal achiieved	0.30	0.14	0.50	0.42	0.38	0.57

Note: the number in a cell indicates the percentage of a type of session stopping reason for sessions whose goals are at a specific complexity level. For instance, 0.30 means that, among the sessions with goals at the retrieve level, 30% of the sessions were stopped because of the goal achieved.

Figure 5.10: The percentage of different session stopping reasons for session goals with different complexity levels.

We also noticed that the distribution of stopping reasons for analyzing and evaluating sessions are very similar, indicating that these two types of sessions could often be stopped for the same reasons.

5.4 Information types, cognitive activities, and resuming and stopping reasons

Based on previous analysis of the information types participants collected throughout the XSS process, the cognitive activities they conducted, and the associated resuming and stopping reasons for the sessions, we wanted to further explore the relationships among these factors.

5.4.1 Information types and resuming and stopping reasons

First, we were interested in whether the useful information types relate to the session resuming and stopping reasons.

Information types and session resuming reasons Table 5.20 summarizes the numbers of different information type examples across the sessions resumed for different reasons.

Table 5.20: The number of information type examples for sessions resumed for different reasons.

Resuming \ info. types	Facutal	Conceptual	Procedural	Opinion	**Total**
Spawn	109	27	21	12	169
Transmut	24	16	6	9	55
Unanswered	15	1	9	1	26
Cultivated	20	0	1	2	23
Anticipated	2	2	2	0	6
Total					279

Note: each cell contains the total number of information examples for a session that was resumed by a specific resuming reason. For instance, 109 factual information examples are found from sessions resumed because of spawning.

A Chi-Square test was performed and found there are significant associations between the two variables, indicating that the information types people found useful for their sessions are affected by the session resuming reasons ($\tilde{\chi}^2$ (12, 101) = 36.287, p<.001) (Table 5.21).

Table 5.21: Chi-Square test for the relation between session resuming reasons and the information types

	Value	df	Asymptotic Sig. (2-sided)
Pearson Chi-Square	36.287	12	$<.001$
Likelihood Ratio	37.867	12	$<.001$
N of Valid Cases	101		

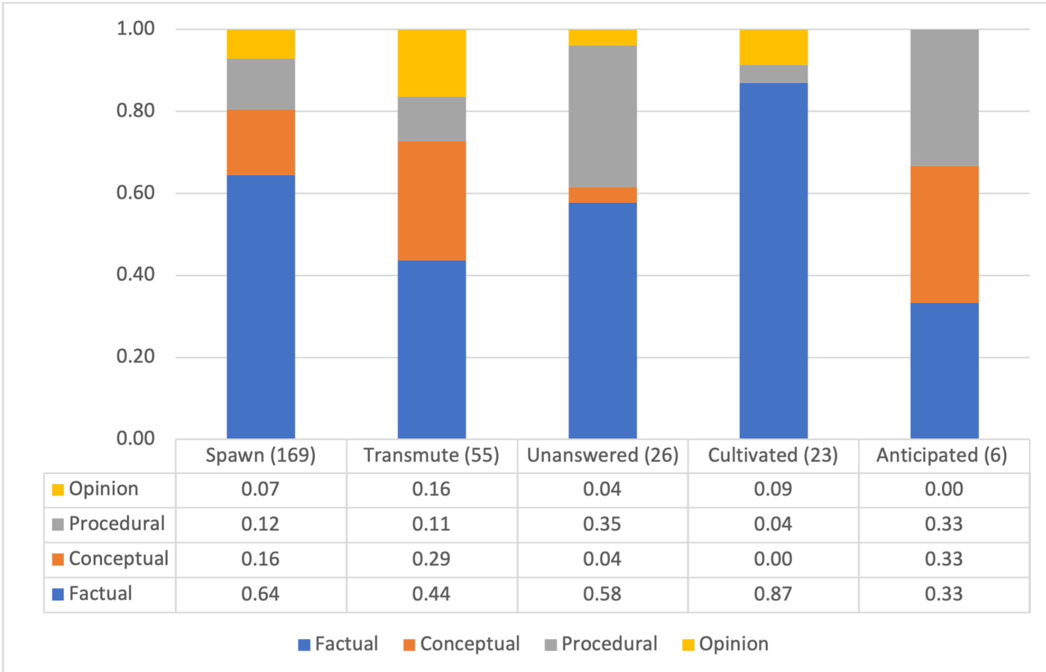
Figure 5.11 is the associated bar chart of the percentage of the distribution. We can find out that:

Factual information generally took the largest proportion across sessions resumed for all five different resuming reasons. It is the most popular information type for *cultivated, spawn, unanswered, transmute* sessions and as equal as other information types for *anticipated* sessions: *cultivated* (87%), *spawn* (64%), *unanswered* (58%), *transmute* (44%), *anticipated* (33%),

Conceptual information is very common for *anticipated* (33%) sessions when people looked for information that potentially will be used in the future as their task moves forward, and *transmute* (29%) sessions where participants looked for information to help them understand the task requirements. *Conceptual* information is less common for *spawn* (16%) and *unanswered* (4%) sessions.

Procedural information can also be found for sessions resumed for all five types of reasons. It is more popular for *unanswered* (35%) and *anticipated* (33%) sessions. It is the least popular information type for *cultivated* sessions (4%), indicating that participants did not often search for procedural information for sessions resumed to look for the latest information.

Opinion information was found across sessions resumed for four reasons: *transmute* (16%), *cultivated* (9%), *spawn* (7%), and *unanswered* (4%). Although *opinion* information was not very common during XSS sessions, it is more popular for *transmute* sessions (16%) than sessions resumed for other reasons. On the other side, opinion information was found but not quite popular for *spawn, unanswered, cultivated* sessions and, it was not found for *anticipated* sessions, meaning that when people need to find information about a subtopic/subtask when they wanted to continue a previous search, searched for updated information or future information for inspirations, they are less likely to look for opinions.



Note: Each cell below a bar represents the proportion of an information type found for the sessions resumed of a specific reason. For instance, 0.64 means that, 64% of the information found for sessions resumed because of spawn was factual information.

Figure 5.11: The percentage of information types for sessions resumed for different reasons.

Information types and session stopping reasons Table 5.22 presents the summary of the numbers of different information type examples across the sessions stopped for different reasons.

Table 5.22: The number of information type examples for sessions stopped by different reasons

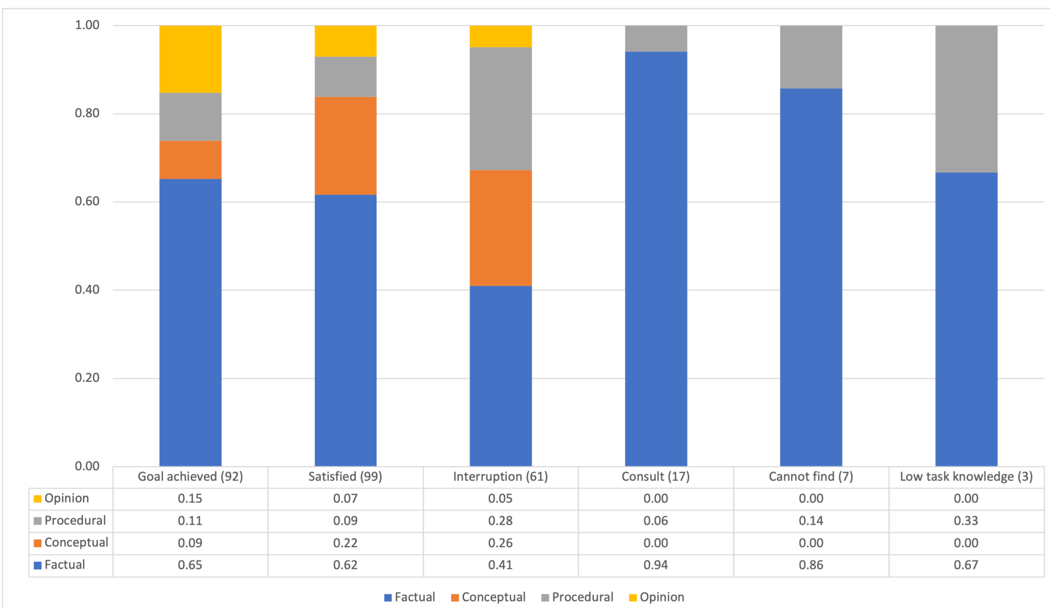
Stopping \ info. type	Facutal	Conceptual	Procedural	Opinion	**Total**
Goal achieved	60	8	10	14	92
Satisfied	61	22	9	7	99
Interruption	25	16	17	3	61
Consult	16	0	1	0	17
Cannot find	6	0	1	0	7
Low task knowledge	2	0	1	0	3
Total					279

Similarly, the Chi-Square test in Table 5.23 found that there are significant associations between the session stopping reasons ($\tilde{\chi}^2$ (15, 101) = 41.973, p<.001) (Table 5.23) and the information types participants found during the sessions,

Table 5.23: Chi-Square test for the relation between session stopping reasons and the information types

	Value	df	Asymptotic Sig. (2-sided)
Pearson Chi-Square	41.973	15	<.001
Likelihood Ratio	46.419	15	<.001
N of Valid Cases	101		

Figure 5.12 shows the percentage of the distribution of information types for sessions stopped for different reasons. We can find that:



Note: each cell below a bar represents the proportion of an information type found for the sessions stopped of a specific reason. For instance, 0.65 means that, for sessions stopped because of goal achieved, 65% of the information found during those sessions was factual information.

Figure 5.12: The percentage of information types for sessions stopped because of different reasons.

Factual information is the most popular information type associated with all six different stopping reasons and the most popular information type within the sessions stopped for the same reasons: *consult* (94%), *cannot find* (86%), *low task knowledge* (67%), *goal achieved* (65%), *satisfied* (62%), and *interruption* (41%).

Conceptual information can be found for sessions stopped because of *interruption* (26%), *satisfied* (22%), and *goal achieved* (9%). This indicates that when participants were looking for conceptual information, they were more often to be interrupted or satisfied than achieved their goals than other session stopping reasons. No *conceptual* information examples were found for sessions stopped because of *consult other information sources, cannot find the need information, or low task knowledge*, meaning that when they looked for conceptual information, they rarely stopped the session to consult others, or cannot find the conceptual information, or stopped because they lack understanding for their task requirements.

Procedural information can also be found for sessions stopped for all six reasons: *low task knowledge* (33%), *interruption* (28%), *cannot find the needed information* (14%), *goal achieved* (11%), *satisfied* (9%), *consult* (6%). It is more common for sessions stopped because of *interruption, low task knowledge* than those for the other four reasons, meaning people were more likely to be interrupted or stopped for low task knowledge when they searched for procedural information.

Opinion information was only found associated with sessions stopped because of *goal achieved* (15%), *satisfied* (7%), *interruption* (5%). *Opinion, conceptual* information was not found for sessions that stopped for *consult, cannot find, and low task knowledge*, which means that for conceptual and opinion information, people seldom reported they stopped to consult others or did not find what they needed, or did not have enough background knowledge to consume the information. However, it is possible that we did not find examples because overall, there are few session examples stopped for these three reasons (i.e., consult, cannot find, low task knowledge).

5.4.2 Cognitive activities and resuming and stopping reasons

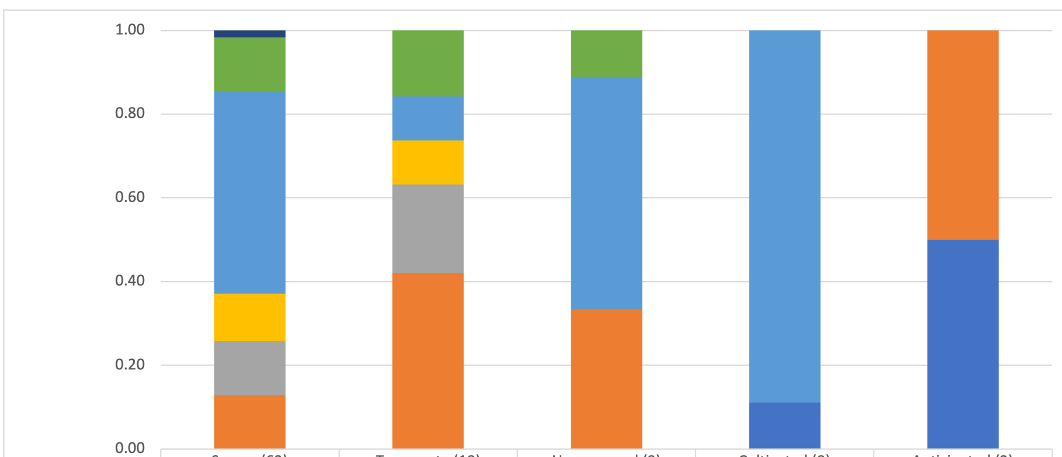
The decision of participants to start/resume or stop a search session may be influenced by the cognitive activities they are required to perform or those they engaged in during the search session. To investigate whether the session resuming and stopping reasons are related to the cognitive activities participants behaved during the session, we conducted Chi-Square tests. It should be pointed out again that for the cognitive activities of the session, we selected the most complex cognitive activity to represent the session's cognitive complexity level, as higher-level cognitive activities frequently build upon and incorporate lower-level cognitive activities.

Cognitive activities and session resuming reasons In Table 5.24, the Chi-Square test results show that there was a significant relationship between the session's resuming reason and the most complex cognitive activity conducted in the sessions ($\tilde{\chi}^2$ (24, 101) = 37.413, p=.04).

Table 5.24: Chi-Square test for relation between session resuming reason and the most complex cognitive activity of the session

	Value	df	Asymptotic Sig. (2-sided)
Pearson Chi-Square	37.413	24	.040*
Likelihood Ratio	39.422	24	.025
N of Valid Cases	101		

Meanwhile, the distribution of data (Figure 5.13) shows that:



	Spawn (62)	Transmute (19)	Unanswered (9)	Cultivated (9)	Anticipated (2)
None	0.02	0.00	0.00	0.00	0.00
Creating	0.13	0.16	0.11	0.00	0.00
Evaluating	0.48	0.11	0.56	0.89	0.00
Analyzing	0.11	0.11	0.00	0.00	0.00
Applying	0.13	0.21	0.00	0.00	0.00
Understanding	0.13	0.42	0.33	0.00	0.50
Retrieving	0.00	0.00	0.00	0.11	0.50

Note: Each cell below a bar represents the proportion of one complexity level of cognitive activity found for the sessions resumed for a specific reason. For instance, 0.13 means that 13% of the sessions resumed because spawning's highest complexity level of cognitive activity is understanding.

Figure 5.13: The percentage of the highest complexity level of cognitive activity for sessions resumed because of different reasons.

Retrieving as the only cognitive activity conducted can be found for sessions resumed because of *cultivated* (11%) and *anticipated* (50%). This might be because, for some of the *cultivated* and *anticipated* sessions, participants were mainly searched to see if there were any updated changes to the information they previously found or look for some potentially useful information without any further processing activities.

Understanding, as the most complex cognitive activity was found for sessions resumed because of *anticipated* (50%), *transmuting* (42%), *unanswered* (33%), and *spawn* (13%). This might be because, for these sessions, participants found information helpful for these tasks (e.g., can help them understand task topic, requirements, provide inspiring ideas, etc.) but not necessarily required for their tasks, so they did not process the information further. *Understanding* was not found for *cultivated* session might be because participants conducted higher-complexity levels of cognitive activity (i.e., evaluate) which overpass the *understanding level*.

Applying, as the most complex cognitive activity was found for sessions, resumed because of *transmute* (21%) and *spawn* (13%), indicating that in these two types of sessions, there are more examples participants directly applied the found information than for other sessions. *Applying* was not found for *unasnwered, cultivated, anticipated* sessions might be because participants did not find information that can be applied to their tasks in those sessions or because they conducted higher level cognitive activities than *applying* (i.e., evaluating, creating).

Analyzing presents a similar pattern as *applying*: it was only found for sessions resumed because of *spawn* (11%), *transmute* (11%), but was not found for sessions resumed for other reasons. It might be because, in *spawn, transmute* sessions, participants searched for more information that required them to break down to be used for the task. And in other sessions, the information they break down may result in *evaluating and creating.*

Evaluating as the most complex cognitive activity can be found for sessions across all types of resuming reasons, except for *anticipated* sessions: *cultivated* (89%), *unanswered* (56%), *spawn* (48%), *transmute* (11%). It is the most popular cognitive activity for sessions for *cultivated, unanswered, spawn* sessions. But it is not quite popular for *transmute* sessions, which might be because, during *transmute* sessions, participants focused on general information like background information or other information to get themselves familiar with the task topic but not require much evaluation of the information.

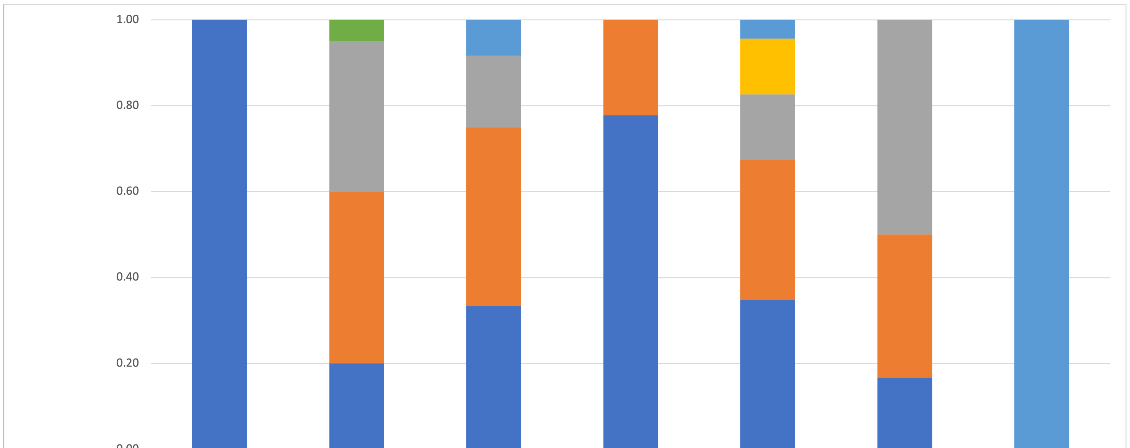
Creating can be found for sessions resumed because of *spawn* (13%), *transmute* (16%), *unanswered* (11%). Since the purposes of these sessions are different, it would be interesting to investigate whether the intents of the creative activities for sessions resumed for different reasons are the same.

Cognitive activities and session stopping reasons Similarly, Table 5.25 shows that there was a significant association between the sessions' stopping reasons and the highest level of cognitive activity carried out within that session ($\tilde{\chi}^2$ (30, 101) = 56.724, p=.002). This might indicate that the most complex cognitive activity carried out during the session has an effect on the decision to stop the session.

Table 5.25: Chi-Square test for the relationships between session stopping reason and the most complex cognitive activity of the sessions

	Value	df	Asymptotic Sig. (2-sided)
Pearson Chi-Square	56.724	30	.002*
Likelihood Ratio	41.604	30	.077
N of Valid Cases	101		

Figure 5.14 shows that:



	Retrieving (1)	Understanding (20)	Applying (12)	Analyzing (9)	Evaluating (46)	Creating (12)	None (1)
Low task knowledge	0.00	0.05	0.00	0.00	0.00	0.00	0.00
Cannot find	0.00	0.00	0.08	0.00	0.04	0.00	1.00
Consult	0.00	0.00	0.00	0.00	0.13	0.00	0.00
Interruption	0.00	0.35	0.17	0.00	0.15	0.50	0.00
satisfied	0.00	0.40	0.42	0.22	0.33	0.33	0.00
Goal achieved	1.00	0.20	0.33	0.78	0.35	0.17	0.00

Note: Each cell below a bar represents the proportion of sessions stopped for a specific reason over the sessions that involved a specific cognitive activity as the most complex cognitive activity. For instance, 0.20 means that, 20% of the sessions involved understanding *as the most complex cognitive activity was stopped because of goal achieved. that stopped because of goal achieved's highest complexity level of cognitive activity is understanding.*

Figure 5.14: The percentage of sessions that stopped for different reasons over the sessions that involved a specific type of cognitive activity as the most complicated cognitive activity.

Goal achieved can be found as the stopping reason for sessions that involved any complexity level of cognitive activities as their most complex cognitive activity: for sessions whose most complex/only cognitive activity is *retrieving*, *goal achieved* is the only stopping reason (100%). *Goal achieved* is also popular for sessions whose most complex cognitive activity is *analyzing* (78%), followed by sessions with the most complex cognitive activity as *evaluating* (35%), *applying* (33%), *understanding* (20%), and *creating* (17%).

Satisfied stopping reason can be found for sessions whose most complex cognitive activity at five complexity levels, except *retrieving*: *applying* (42%), *understanding* (40%), *evaluating* (33%), *creating* (33%), *analyzing* (22%).

Interruption stopping reason can be found for sessions whose most complex cognitive activity at five complexity levels: *creating* (50%), *understanding* (35%), *applying* (17%), *evaluating* (15%). In other words, sessions' whose most complex cognitive activity is *understanding, creating* are often stopped because of *interruption* than sessions whose most complex cognitive activities are *applying,* *evaluating*.

Stop to consult other information sources as the stopping reason can only be found for sessions whose most complex cognitive activity is *evaluating*: 13% of the *evaluating* sessions are stopped to consult other information sources.

Cannot find the needed information can be found as the stopping reasons for sessions whose most complex cognitive activity involves *applying* (8%), *evaluating* (4%). The session that was not associated with any cognitive activity (i.e., none) was stopped because the participant could not find what they needed.

Low task knowledge as the stopping reason was found for sessions whose most complex cognitive activity is *understanding*, indicating that during the process when the participant wanted to make sense of the found information, they noticed that they could not move forward because they lack some knowledge about their task.

5.5 Task stages, session difficulties, and session resuming and stopping reasons

We conducted Repeated Measures ANOVA (RMA) to compare users' perceptions of task difficulty at three stages: 1) each participant's perception of the task difficulty before they started the task. This is collected during the pre-task questionnaire they filled out during the introductory meeting; 2) the average perception of post-session difficulty, which is the mean of the post-session difficulties of all the search sessions conducted by a participant. For instance, one participant searched for four sessions for her task, and after completing each session, she rated the session's difficulty in the post-session questionnaire for the session. This repeated for four sessions. During the analysis, we summed the post-session difficulty for all four sessions and then calculated the mean as the average post-session difficulty; 3) the post-task difficulty obtained from the post-task questionnaire. Once participants indicated they finished the last search for the task in the corresponding post-session questionnaire, they were asked to fulfill an exit questionnaire in which they rated the task difficulty again based on their experience.

A repeated measures ANOVA is used to determine whether or not there is a statistically significant difference between the means of three or more groups in which the same subjects show up in each group. We conducted an ANOVA test with participants' perceptions of the overall task difficulty, the search difficulty (search_diff), find enough information difficulty (enough_diff), and the difficulty to integrate the found information (integrate_diff), respectively. The result is shown in Figure 5.15.

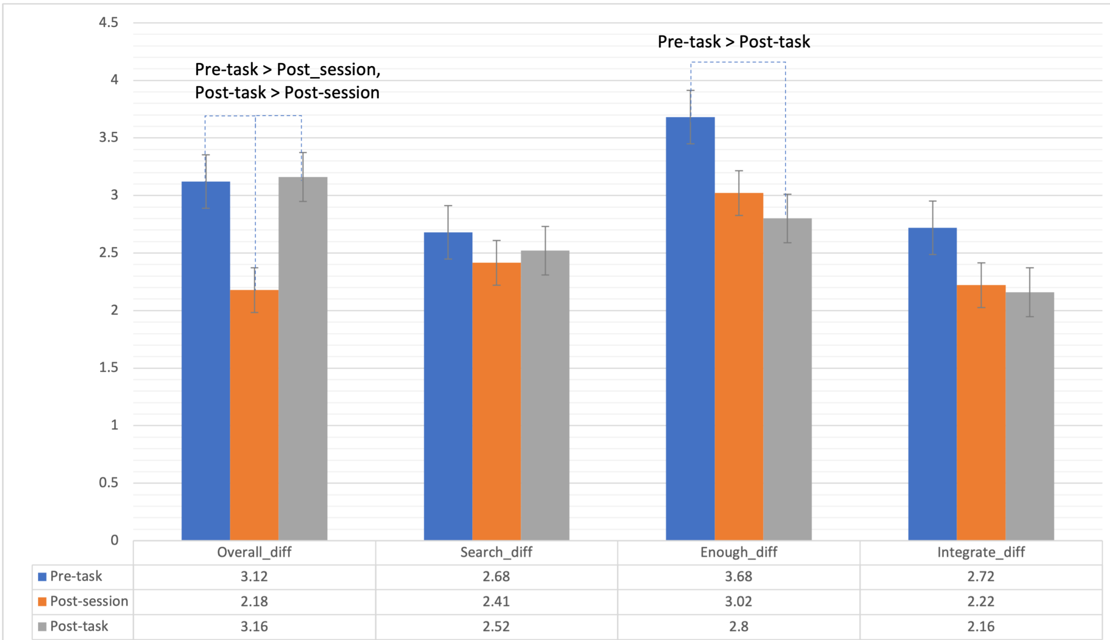


Figure 5.15: Means of difficulties tested before tasks, post-sessions, and post-tasks. Means found to be significantly different with repeated Measures ANOVA are indicated with dotted lines and labels.

The descriptive statistics associated with participants' perceptions of the different types of task difficulty across the three stages are reported in Table 5.26. In order to test the hypothesis that the task stage (pre-task, post-session, post-task) had an effect on participants' perception of the task difficulty (i.e., overall task difficulty, search difficulty, finding enough information difficulty, integrate difficulty), ANOVA tests were performed. In the following subsections, we will explain our results for each of them.

Table 5.26: Descriptive statistics for participants' perceptions of task difficulties measured before the task (pre-task), after search sessions (post-session), and after the task completed (post-task)

difficulty measure	Pre-task *M (SD)*	Post-session *M (SD)*	Post-task *M (SD)*
Overall difficulty	3.12 (1.27)	2.18 (.89)	3.16 (1.60)
Search difficulty	2.68 (1.25)	2.41 (1.02)	2.52 (1.29)
Find enough info. difficulty	3.68 (1.46)	3.02 (1.33)	2.8 (1.83)
Integrate difficulty	2.72 (1.49)	2.22 (1.00)	2.16 (1.55)

5.5.1 Difficulty of the overall task and sessions

From Table 5.26, it can be seen that participants' post-session perception of overall session difficulty (M=2.18, SD=.89) is lower than their perception of pre-task overall task difficulty (M=3.12, SD=1.27) and their post-task perception of overall task difficulty (M=3.16, SD=1.60).

The Repeated Measures ANOVA (RMA) test yielded a statistically significant difference [F (2, 48) =10.077, p <.001]. A post-hoc pairwise comparison (Table 5.27) show that the average perception of post-session overall difficulty (M=2.18, SD=.89) was significantly lower than the pre-task overall difficulty (M=3.12, SD=1.27), and lower than the post-task overall difficulty (M=3.16, SD=1.60), respectively.

Table 5.27: Pairwise comparisons of participants' perceptions of overall difficulty for pre-task, post-session, and post-task

(I)Overall_diff	(J)Overall_diff	Mean difference (I-J)	Std. Error	Sig.^{b}	95% confidence^b	95% confidence$_b$
Pre-task	Post-session	.941*	.213	**<.001**	.502	1.381
	Post-task	-.040	.291	.892	-.641	.561
Post-session	Pre-task	-.941*	.213	**<.001**	.-1.381	-.502
	Post-task	-.981*	.231	**<.001**	-1.458	-.505
Post-task	Pre-task	0.40	.291	.892	-.561	.641
	Post-session	.981*	.231	**<.001**	.505	1.458

5.5.2 Difficulty to search for information

We also compared participants' perceptions of the difficulty of searching for information at the three task stages. From Table 5.26, it can be seen that participants' average pre-task rate of difficulty to search for information (M=2.68, SD=1.25) is highest among the three stages, while their average post-session rate of search difficulty (M=2.41, SD=1.02) is the lowest, and their average post-task rate of search difficulty (M=2.52, SD=1.29) is in-between.

The RMA test did not find significant differences [F (2, 48) =.738, p =.484].

5.5.3 Difficulty to find enough information

Different from the difficulty of searching for information, the difficulty to find enough information intends to evaluate to what extent participants felt to achieve their search goals of getting enough information from the search process.

From Table 5.26, it can be seen that participants' average rate of difficulty to find enough information decreased from pre-task, post-session to post-task. Their pre-task rate of difficulty to find enough information was the highest among the three (M=3.68, SD=1.46), followed by the average rate of post-session difficulty to find enough information (M=3.02, SD=1.33), and then the post-task rate of the difficulty to find enough information (M=2.8, SD=1.83).

The RMA test yielded a statistically significant difference [F (2, 48) =4.159, p = .022]. A post-hoc pairwise comparison (Table 5.28) show that the average perception of post-task difficulty to find enough information (M=2.8, SD=1.83) was significantly lower than the pre-task perception of the difficulty to find enough information (M=3.68, SD=1.46), p=.029. The difference between participants' perception of pre-task difficulty to find enough information (M=3.68, SD=1.46) and the post-session perception of difficulty to find enough information (M=3.02, SD=1.33) was marginally significant (p=.054).

Table 5.28: Pairwise comparisons of participants' perceptions of the difficulty to find enough information for pre-task, post-session, and post-task

(I)Enough_diff	(J)Enough_diff	Mean difference (I-J)	Std. Error	Sig.[b]	95% Confidence[b]	95% Confidence[b]
Pre-task	Post-session	.659	.325	**.054**	-.011	1.329
	Post-task	.880*	.380	**.029***	.096	1.664
Post-session	Pre-task	-.659	.325	**.054**	.-1.329	.001
	Post-task	.221	.229	.344	-.251	.694
Post-task	Pre-task	-.880*	.380	**.029***	-1.664	.-.096
	Post-session	-.221	.229	.344	-.694	.251

5.5.4 Difficulty to integrate information

Another important aspect of search is to integrate the found information. Participants were asked about their perceptions of this factor at the three stages.

From Table 5.26, it can tell that the mean of participants' rate for the pre-task difficulty to integrate information (M=2.72, SD=1.49) is the highest among the three, followed by the mean of the post-session difficulty to integrate information (M=2.22, SD=1.00), and then the mean of the post-task difficulty to integrate information (M=2.16, SD=1.55). In other words, the overall trend of integration difficulty is decreased along with the progress of the tasks.

The RMA test did not find any significance [F (2, 48) =2.303, p =.111].

5.5.5 Post-session difficulties and resuming and stopping reasons

We analyzed the relationships between the post-session difficulty and the session resuming and stopping reasons to see whether different reasons are associated with different levels (1 to 7) of the four session difficulties (overall session difficulty, difficulty to search for information, the difficulty to find enough information, and the difficulty to integrate the found information, respectively.

Post-session difficulty and session resuming reasons. Regarding the relationships between participants' perceived search session difficulty to the reasons that motivated them to start the session, the null hypothesis is that the resuming reasons are not associated with the levels of difficulty for each difficulty type. The chi-square tests did not find any significant differences (Table 5.29).

Therefore, the null hypothesis is not rejected.

Table 5.29: Chi-Square tests for post-session difficulties and session resuming reasons

		Value	df	Asymptotic Sig. (2-sided)
Diff_overall	Pearson Chi-Square	14.771	24	.927
	Likelihood Ratio	15.898	24	.892
	N of Valid Cases	101		
Diff_search	Pearson Chi-Square	15.672	24	.900
	Likelihood Ratio	16.743	24	.859
	N of Valid Cases	101		
Diff_enough	Pearson Chi-Square	13.814	24	.951
	Likelihood Ratio	16.610	24	.865
	N of Valid Cases	101		
Diff_integrate	Pearson Chi-Square	20.744	24	.654
	Likelihood Ratio	23.397	24	.496
	N of Valid Cases	101		

Post-session difficulty and session stopping reasons. We also conducted Chi-Square Test of Independence to assess the relationships among the levels of the difficulty of different difficulty types to the session stopping reasons. The null hypothesis is that the session stopping reasons are not associated with the levels of difficulty of the different difficulty types.

From the Tabel 5.30, it can be found that: there are significant relationships between the session stopping reasons and the level of all the different types of difficulties: 1) there was a significant relationship between the overall session difficulty and the session stopping reasons ($\tilde{\chi}^2$ (30, 101) = 105.983, p<.001). 2) there was a significant relationship between the session difficulty to search for information and the session stopping reasons ($\tilde{\chi}^2$ (30, 101) = 47.817, p=.021). 3) there was a significant relationship between the session difficulty of finding enough information and the session stopping reasons ($\tilde{\chi}^2$ (30, 101) = 48.628, p=.017). 4) there was a significant relationship between the session difficulty of understanding the found information for the tasks and the session stopping reasons ($\tilde{\chi}^2$ (30, 101) = 62.357, p<.001). 5) there was a significant relationship between the session difficulty of integrating the found information and the session stopping reasons ($\tilde{\chi}^2$ (30, 101) = 71.080, p<.001).

Table 5.30: Chi-Square tests for post-session difficulties and session stopping reasons

		Value	df	Asymptotic Sig. (2-sided)
Diff_overall	Pearson Chi-Square	105.983	30	<**.001***
	Likelihood Ratio	46.932	30	.025
	N of Valid Cases	101		
Diff_search	Pearson Chi-Square	47.817	30	**.021***
	Likelihood Ratio	40.614	30	.094
	N of Valid Cases	101		
Diff_enough	Pearson Chi-Square	48.628	30	**.017***
	Likelihood Ratio	44.871	30	.040
	N of Valid Cases	101		
Diff_integrate	Pearson Chi-Square	71.080	30	<**.001***
	Likelihood Ratio	41.579	30	.078
	N of Valid Cases	101		

However, the post-hoc tests did not find statistical significance between the session stopping reasons and the levels of difficulty of different difficulty types. This might be because of the small size of our sample and the lack of statistical power. Figure 5.16 provides an example of the crosstab examination.

Diff_overall * stop Crosstabulation

Count

		stop						
		Cannot find	Consult	Goal achieved	Interruption	Low task knowledge	satisfied	Total
Diff_overall	1	0	2	16	12	0	15	45
	2	0	1	12	2	0	7	22
	3	1	2	3	1	0	8	15
	4	1	1	3	4	0	3	12
	5	0	0	0	2	0	0	2
	6	0	0	0	0	1	1	2
	7	2	0	0	1	0	0	3
Total		4	6	34	22	1	34	101

Figure 5.16: Crosstab examination of participants' perceptions of post-session overall difficulty to the session stopping reasons.

5.6 Challenges of searching across sessions

We also identified a group of challenges that participants often faced during the cross-session search processes. Below we describe the four main reasons:

5.6.1 Changing ideas between sessions

Participants found that when searching across sessions, their focus and perspective kept changing, which could cause problems. For instance, P101 mentioned that "you are not like the same exact person each session...[and] don't have the consistency of... one big session." P114 said that "even if it's different times of day I'm thinking of different things...the things I want to accomplish changes frequently." P125 described that "it's hard to get back into the groove and focus when you split things up."

5.6.2 Losing track of ideas and information

Participants reported it can be difficult to make connections between sessions because they often lose track of the information they found, their previous ideas, or what they have done in earlier sessions. Sometimes it can cause them to re-do the search, whereas sometimes, they found they were wasting time on revisiting what they had read before. For instance, P106 mentioned that "I think [the challenge] was trying to remember what I had done before and what I hadn't." P114 said that "it was harder to recall exactly what I was thinking at that moment."

5.6.3 Too much information to process

Another challenge for cross-session search is that it often involves large amounts of information to process – in some cases, too much information. This can especially be a problem when participants sought for information on the same subtopics but from different resources during multiple sessions. It can lead to problems in synthesizing what they found, difficulties in comparing options, and sometimes leads to missed opportunities. For instance, P108 noticed that "there are too many options [you have to] narrowing down [to] take the best and formulating something on its own." P103 mentioned that "sometimes I would find information after the fact, but I'm done so it is too late for [changing]."

5.6.4 Hard to keep motivated

The last challenge we identified related to cross-session search is that as their search activities spread over time, participants found it hard to keep motivated to find time and energy to continue their tasks. For instance, P110 described that "it would be difficult [because] crossing between each session...you gotta build up steam again." Similarly, P125 mentioned that " it can be hard to just sit down and actually do it, it does add [an] element like procrastination, struggle."

We also noticed that participants described the lack of motivation from two perspectives: 1) lack of motivation for keeping the task going because the task either turned out to be much more complicated or required much more work than they anticipated. For instance, P129 described that "I did not have an idea of how many pages and how many words per page would be most effective until after reading this section..." (P129-S2) 2) lack of motivation to keep a search session going because there was too much information that caused participants to be distracted during the session, or lose of their search directions. For instance, participant P120 described that "I'm less sure how much research I want to do on one particular policy since they are unique and the arguments for each one differ...I felt overwhelmed by my uncertainty about going too far in one direction." (P120-S2)

CHAPTER 6

Discussion and Implication

In this dissertation study, we conducted a diary study to investigate users' information behavior for tasks that involved cross-session search. Our analysis focused on five research questions:

RQ1. The stopping and resuming reasons leading people to search for multiple sessions and their relationships.

RQ2. The effects of session-level goal complexity on the types of information, the cognitive activities, and the stopping and resuming reasons during XSS.

RQ3. The effects of information types used and cognitive activities on the session stopping and resuming reasons.

RQ4. The effects on task stages on participants' perceptions of pre-task, post-task, and post-session difficulties.

RQ5. The challenges and difficulties caused by searching for multiple sessions during XSS.

We used both qualitative and quantitative methods for data analysis: Lin and Belkin (2005)'s MISE model was used to develop the coding frame for session stopping and resuming reasons. Anderson and Krathwohl (2001)'s taxonomy of knowledge types and cognition was used to analyze the types of information participants found during XSS and their mental processing of the information. We also used the open-coding method to analyze participants' motivation for conducting expected XSS and their non-search task activities.

In addition, we used quantitative methods to analyze the relationships between different factors: Chi-Square tests were performed to explore the relationships between session resuming and stopping reasons, the session goal complexity and information types found during sessions, session goal complexity and the cognitive activities conducted during the sessions, the relationships between the information types/cognitive activities and the session stopping and resuming reasons. ANOVA tests were performed to investigate the relationships among participants' perceptions of task/search difficulties at the pre-task stage, session stage, and post-task stage. Descriptive statistics were used to present the frequency distributions.

6.1 Cross-session search task characteristics

First, we presented an overview of the 25 cross-session work tasks from 25 participants. Our tasks are more diverse compared to previous studies on XSS from the following perspectives:

6.1.1 Task topics

The task topics ranged from creative writing (i.e., screenplay writing, book preparation), events planning (i.e., graduation ceremony trip, baby shower), personal hobby practice and learning (i.e., digital art training, sculpture DIY, [toy] collection curate), business event (i.e., company registration, property renting, shopping), to research projects (i.e., proposal, essay writing, presentation), etc. Our task coverage is much more diverse than previous studies on XSS, especially for those involving longitudinal observation of real tasks like in (Spink, 1996; MacKay & Watters, 2008a)

6.1.2 Task sources

Three task sources were identified: self-motivated tasks (n=12), tasks assigned by others (n=9), and tasks motivated through group discussions (n=4).

The results indicate that XSWTs could be tasks that are *motivated* by different sources. The task or the problem could be purely raised from one individual's information needs, being assigned by others in an organization, or based on discussion with others–who may or may not be involved in the tasks. There are more self-motivated tasks than the other two types. It could be caused by our screening processes or because of convenience–participants who want to complete self-motivated tasks have better control over how and when to complete their tasks to fit the study requirements. The organizations for the assigned tasks include college (i.e., school assignment), professional organizations (i.e., athlete institutions, companies, performance institutions), and NGO groups. A significant amount of tasks (n=9) in our sample are assigned by others, which may be because many of our participants are undergraduate or graduate students (n=16, 64%) and it is easy for them to come up with assignments that may take multiple sessions of searching to complete. The tasks are motivated through discussion with a group of people to achieve a common goal or for the same interests. However, there are only a few of the recruited tasks were motivated through group discussions. This might be because, compared to the other two types of tasks, tasks that are motivated by group discussion are generally less common. These tasks require the same interests among different people at the same time, and individuals in the group need to agree on the same task goal and are willing to complete the task together. Few previous research included the task resources for XSS task analysis. As an important task facet (Li & Belkin, 2008), the sources of tasks could potentially affect the task timeline, the decomposition of tasks into subtasks, the process and strategies about how to complete the tasks, and the allocation of task workload. It may also affect the evaluation of the final task outcomes. This is interesting for future investigations.

6.1.3 Task timeline and task doer

We compared the expected task time with the actual time that participants (partially) completed their planned task by the end of the study and found out that their tasks generally lasted over a longer time period than participants' expectations. For short-term tasks (i.e., 1-3 days), participants spent 5 days on average to complete. For middle-term tasks (i.e., 3-5 days), participants averagely spent 13 days; for long-term tasks (over a week), participants generally spent about 22 days on the tasks. From the perspective of the task-doers, 19 tasks were completed by the participants themselves, 5 tasks involved consulting others, and only 1 task required collaborative work between a group of people.

There might be several reasons why tasks lasted longer than participants' expected timelines: As some participants may have a more clear idea about the task timeline, some participants may not have a precise time point about when their tasks could be done. Meanwhile, even for tasks that have clear pre-set deadline– the tasks that were assigned by others (e.g., P108, P110, P112, P113, P121, P124, P125, P126, P127), most participants (7 out of 9) spent more days than they expected. This might be because of the following reasons:

1) the tasks participants chose to attend our study were less time-sensitive. Although some tasks do have specific deadlines, participants started earlier since they knew that they needed to spend

a significant amount of time and conduct multiple searches to complete the task, which left them plenty of time to work on the task before the deadline;

2) some participants were not clear about the time needed to complete their task because their task did not have concrete outcomes or clear boundaries that can be defined as the end of the task (e.g., learning about how to do digital art as a hobby (P114), collecting information about road safety issue (P120), self-training how to conducting statistical sampling (P128);

3) Participants' tasks spread out and new topics/directions emerged. Some participants spent more time than they expected on the task because their tasks spread out and appeared new factors that they were not aware of at the beginning cost more time and effort. For instance, P105 made a list of potential interviewees at first. But new topics emerged as he found more information and talked with people (from the list). So he wanted to continue to search for more information after the diary study.

4) It often took extra time when participants needed to consult other information sources or collaborate with others. It may take longer to get feedback from other people, achieve agreements through multiple rounds of communication, and make decisions about future directions, etc.

5)Some participants seemed to underestimate the efforts required to complete a task due to unfamiliarity or lack of experience. For instance, P129 had an idea of completing a book regarding parenting skills for communicating with children in two weeks. But she finally decided to stop the diary study after searched three sessions over 6 days because she found out that there were too many parts of it and she could not complete the book even if given another week.

As previous studies on XSS noticed that XSS work tasks often take a significant period of time to complete (MacKay & Watters, 2008a; Morris et al., 2008a), little has been discussed about the underlying reasons that cause the tasks to spread over time. Given that all tasks in this study are *expected* to involve XSS and expected to take multiple days to weeks to complete, tasks that are *unexpectedly* spread over time might involve more reasons we did not observe here.

The above reasons indicate that task timeline may be affected by task characteristics (e.g., task workload, task timeline, the priority of the task), characteristics of task doers (e.g., solo work or collaboration, prior knowledge about task topic and task skills), or the characteristics of the information sources required by the tasks (i.e., open-web resources, professional databases, information from other human beings), and how the task was planned to be completed and how people work on it.

6.1.4 Task complexity

The level of the XSS work task complexity for our data clustered into two groups: the *understand-apply* group (n=6) and the *evaluate-create* group (n=19). This result provides evidence that cross-session work tasks are not necessarily very complicated tasks: tasks at lower-level complexity could involve cross-session search because 1) some tasks contain multiple subtasks or components that require to be completed over time or hard to be completed within a one-sitting process; 2) tasks at lower-level complexity does not necessarily mean the steps/procedures are easier; 3) participants were not fully aware of the complexity of tasks before they actually work on it.

Furthermore, together with later analysis of the complexity of session goals (Section 5.3.1) and the complexity of the cognitive activities (Section 5.3.3) that participants conducted through the XSS process, our analysis indicates there are several different complexities related to XSS: a) expected work task complexity, b) session goal complexity, c) individual session cognitive complexity, and d) the experienced overall work task complexity. In this study, the work task complexity is mainly the evaluation of the *potential* most complicated cognitive activities required by the task based on participants' descriptions of how they *would* work to complete the task. Our distinctions of the different types of complexity are also related to the research of the relationships between work tasks, information-seeking tasks, and information-search tasks (Byström & Hansen, 2005) from the perspective of task complexity. Meanwhile, the *experienced overall work task complexity* could extend the previous research on search task difficulty (Arguello, Wu, Kelly, & Edwards, 2012) to a higher level from the perspective of work task.

Previous research often described the complexity of tasks involving XSS based on the number of subtasks or goals, number of search queries and sessions, and the length of tasks' time period (Czerwinski et al., 2004; Agichtein et al., 2012; MacKay & Watters, 2008a; Morris et al., 2008a). A few experimental research considered the relationships between subtasks (e.g., parallel or hierarchical) (J. Liu & Belkin, 2010) and cognitive activities (i.e., exploration) to design their tasks (D. Wu, 2018; Han et al., 2015). Relatively little has been discussed about the different types of complexities at different task levels, as we mentioned above. Recently, Urgo and her colleagues conducted a series of studies (Urgo et al., 2020; Urgo & Arguello, 2023) to investigate the effects of search goals at different complexity levels (i.e., knowledge types and cognitive activities, different types of subgoals) on users' perceptions of tasks and their search activities at different stages during single-session search. They found that search goals with different knowledge types can affect users' perceptions of expected information needs, task difficulties, satisfaction, their search behaviors, and final task outcomes.

Together, it indicates the effects of different levels of complexity during XSS have not been studied well. Based on what we have found here, there is ample scope for future research to delve beyond investigating the interconnections among the *Expected work task complexity, session goal complexity, actual session cognitive complexity, and actual work task complexity*, for instance, how these elements influence users' perceptions and dictate their search behaviors throughout the XSS process. Additionally, an exploration of the roles that other task-related factors may assume within these relationships, as well as the synergies that emerge under specific circumstances, could offer a more comprehensive understanding of the intricacies involved.

6.1.5 Non-search activities and XSS work modes

Our results identified seven types of non-search activities participants conducted during their task process: 1) setting up/adjusting task scope and goals, 2) creating task work plans, 3) working on the task using prior knowledge, 4) applying the found information to address task problems, 5) exchanging ideas with task collaborators, 6) consulting others for opinions and inspirations, and 7) monitoring task progress and ideas. Here, we note that in general: (a) the purpose of these non-search task actions was not to find the specific information required by the task, and (b) participants separated these actions from their interactions with search engines or other information systems.

Non-search information activities have long been noticed and often generalized as *information-related activities* (Byström & Hansen, 2005). For instance, research on human-information behavior often described information activities besides such as *presentation* in Kuhlthau(1991)'s ISP (Information Seeking Process) model, and *sense-making* behaviors in Dervin (1983)'s Sense-Making Theory. Marchionini (2006) described users' cognitive and metacognitive activities during exploratory searches such as discovery, planning, analysis, evaluation, etc.

Instead of treating the non-search information activities separately from the searching process, the commonality of the above examples of information use focuses more on users' mental processing of the information as part of the information-seeking/searching process rather than activities corresponding to the entire work task environment. As described by Wilson (2000):

> "Information use behavior consists of the physical and mental acts involved in incorporating the information found into the person's existing knowledge base. It may involve, therefore, physical acts such as marking sections in a text to note their importance or significance, as well as mental acts that involve, for example, comparison of new information with existing knowledge." (Wilson, 2000, p.50)

Previous research observed various types of non-search information activities in information tasks. For instance, task tracking, routing tasks like news browsing, meetings, emails, note-taking, information saving, search history reacquainting, collaboration, etc. (Czerwinski et al., 2004; Morris et al., 2008a; MacKay & Watters, 2008a). The challenge of studying non-search activities is to focus on the interplay and shifts between the non-search and search activities rather than isolating a particular activity, such as information retrieval tasks, and studying it alone.

Users' information activities could be affected by their work task types and task processes (Capra, Arguello, & Zhang, 2017; J. Liu et al., 2012; Vakkari & Hakala, 2000; Zhang et al., 2020). They may intertwine actions of (1) working on the main work task, with (2) searching for needed information, (3) applying found information, and (4) iterating these steps. From this perspective, users' information search activities and their information use activities could affect each other as the work task moves forward. Research on creative activities studied how people intertwine their information-searching behavior with other creative activities (e.g., find goals, look up, create ideas, combine ideas, execute) through the creative task process, which often involves XSS (Zhang, 2020). The findings show that participants iteratively searched for various information from various sources throughout their creative stages.

Our results observed three different patterns regarding how users' search sessions intertwined with non-search activities: *searching-doing-search-doing (SDSD)*, *Doing-Search-Doing-Search (DSDS)*, and *Search-Search-Doing (SSD)*. The *doing* represents the non-search task activities, as we identified earlier. We note that there are two main differences among these modes: (1) whether the process starts with a search activity or a doing activity, and (2) whether the search sessions are mixed with working sessions through the task process. These initial results can help us further understand how users' information search activities can be affected by other task activities. Conversely, these results also give insight into how non-search activities influence future search activities in cross-session search processes.

6.1.6 XSS motivations

Previous research on users' information-seeking motivation often concentrates on users' information needs about why people want/need to look for information, which is often traced back to the problem or situation people are faced with and their awareness of the information required (Wilson, 1997; Dervin, 1983). Meanwhile, research on motivations related to XSS focused more on the reasons for session resuming reasons (i.e., why people conduct successive searches after previous searches for the same task), which we will discuss in Section 6.2.1. Our analysis of XSS motivation here specifically focused on why people *expect* to search for multiple sessions from the overall task perspective.

Our analysis of the fifteen interviewed participants identified two groups of motivations: task-oriented and cognition-oriented motivations. The task-oriented motivations include 1) the work tasks often involved multiple parts that would need to be completed over time, 2) information related to the task was constantly changing, and 3) the work task had a large scope or timeline. The cognition-oriented motivations include 1) to understand the found information, 2) to explore new ideas, 3) to process found information for future use, 4) to organize the found information, and 5) to avoid mental fatigue by leveraging the time gaps between sessions. Some of these reasons align with previous research on task complexity that can cause search difficulty. For instance, Campbell (1988) says:

> "...any objective task characteristic that implies an increase in information load, information diversity, or rate of information change can be considered a contributor to complexity." (Campbell, 1988, p.43)

Our findings of multiple parts reflect the information load and diversity; the constant changing of the relevant information relates to the rate of information change. Previous research of task timeline often considers it as a *generic task characteristic* as a task takes place (i.e., how people complete a task) (e.g., short-term, long-term), frequency (e.g., unique vs. routine), or task stages (i.e., beginning, middle, final) (Li & Belkin, 2008). However, little has been discussed about the importance of why people conduct a task in a certain way. Our findings indicate that for XSS, task timeline plays an important as the reason that leads people to decide the way they want to allocate their time and energy to achieve their task goals.

In addition, the tasks with multiple parts motivation imply opportunities for search systems to aid users in breaking down their tasks and focusing on specific parts. It also helps explain why prior work found that cross-session searches often involve gathering information from different sources (Sellen et al., 2002; MacKay & Watters, 2008a).

As the task-oriented reasons are about why people want to search for multiple sessions, the cognition-oriented motivation focuses on why people want the time gaps between the sessions. There is much research about users' cognitive activities happening through the information search process, such as processing the information to solve the problem, organizing the found information, exploring information from different resources, etc. (Marchionini, 2006; Morris, Ringel Morris, & Venolia, 2008b). There is less research about users' task-related activities before or after their search sessions. This might be because a majority of previous research is about single-session searches. Or they treated each search session independently. But previous research on XSS rarely discussed users' task activities that happened between sessions. And to the best of our knowledge, little has been investigated from the perspective of motivations that people planned in advance to take the time gaps between the search sessions to conduct task-related activities. Our findings imply that 1) people would conduct task-related activities between search sessions as their task involve XSS, 2) their awareness of the need for the time gap between sessions is mainly because of their awareness of the cognition load required by the task and their own cognitive capability–self-awareness of learning capability are an important factor for self-regulated learning(Schunk & Greene, 2018), which is recently started to attract the attention of researchers in the *Search As Learning* community. Our findings indicate that participants' expectations about the cognitive activities have a profound influence on their search process and merit further exploration.

6.2 Session resuming and stopping reasons

6.2.1 Session resuming reasons

Overall, we found that spawning, transmuting, unanswered-incomplete, cultivated-updated, and anticipated were the main reasons that led participants to begin new search sessions during XSS. Further, we identified three ways that subtasks spawned (based on previous knowledge, earlier search sessions, and communication with others). These extend the original definition of spawning in (Lin & Belkin, 2005) and provide evidence to support prior work indicating that complex tasks with multiple subtasks are often associated with XSS (e.g.,(Donato, Bonchi, Chi, & Maarek, 2010b; Kellar et al., 2007; J. Liu & Belkin, 2010)).

Transmuting was also a common reason in our data, especially for initial sessions. This might be because many (76%) of our participants' tasks were self-generate and thus may have been less structured compared to assigned tasks. This is in line with Agichtien et al.'s (Agichtein et al., 2012) finding (based on search log data) that undirected tasks were more likely to be continued over multiple sessions. Prior research found that users often look for problem definition and background information to help structure less-determined tasks (Byström & Järvelin, 1995; Choi & Arguello, n.d.). Our findings show that in the XSS context, the information sought during transmuting sessions can lead to a change in the scope of the original problem since participants might narrow down the scope of the task or redefine their problems. Consequently, their work processes and the final outcomes could be different depending on the information users found and used for structuring the tasks.

Unanswered-incomplete was another important resuming reason, especially for sessions later during the XSS process. Our results suggest it may be more common for participants to continue a search that was unfinished instead of continuing a previous search that failed. This could have been because, during XSS sessions, participants often aimed to find multiple pieces of information. Therefore, there were more times when they did not finish finding all the information they sought rather than not finding any useful information at all. Another reason might be that for complex XSS tasks, there were often alternative solutions, so participants were likely to find at least some useful information during a session. We also note that Wu (D. Wu, Dong, Tang, & Capra, 2019) found that unanswered or unsatisfied search results may cause people to switch their search to a different device. Future work should explore how XSS search resuming reasons may be influenced by the use of multiple devices.

Our results also show that in the sessions resumed because of cultivated-updated and anticipated. Participants focused on looking for specific information. This differs from the MISE model's (Lin & Belkin, 2005) view that the purpose of these two reasons is to familiarize oneself with information that could be potentially useful for future tasks. Our findings suggest that it would be helpful to change this definition to include information that the user thinks might be helpful at some point in the future for the current work task.

Another contribution of our study is that we identified six additional session-resuming reasons that were not discussed by the MISE model and have been less explored in the context of XSS. Among these, exploring more topic aspects, finding inspirations/examples, and reviewing the information found earlier are related to existing MISE reasons but reflect nuances of users' information needs. For instance, similar to spawning, *exploring* can lead to the discovery of new information that may become subtopics for further investigation. However, in our classification, an exploring session is often less directed, whereas spawning sessions often have specific focuses and goals. The new resuming reason *finding inspirations/examples* relates to the MISE reason of transmuting when people search for background information or similar experiences on how to complete their tasks. A difference is that when finding inspiration, users' task goals were often more structured, and they focused more on the examples of the process and final outcome. In the original MISE transmuting, the focus is often to look for information to help structure or narrow down task scope. Our new resuming reason, "reviewing the previously found information" relates to multiple original MISE reasons. For instance, reviewing could happen in unanswered-incomplete sessions resumed when people want to reexamine what they had found to continue their previous search. It also might happen in cultivated-updated sessions when people revisit previous information resources to check for updated information. It might occur when people want to re-access information in the lost-treatment cases. However, a reviewing session aims to confirm what had been found previously. The information from the previous sources has been fully understood rather than looking for new information as in the cultivated, incompleted sessions.

Meanwhile, monitoring task progress, completing a search following a scheduled plan, and feeling in a mood/having the energy to search are non-information-need-related reasons. These reasons reflect users' metacognitive activities through the XSS process. Researchers have found that leveraging tools to support metacognitive activities can significantly improve users' interaction with the systems and affect their task outcomes during experimental tasks (Azevedo, Cromley, & Seibert, 2004; Stadtler & Bromme, 2008). Our findings show that users have the need for and often perform metacognitive activities in their real-world XSS tasks.

And the original "Anticipated" mode did not describe an XSS search situation. Other resuming modes describe the current search session as a continuation of the previous search and explain how the current search is related to the previous search. However, the anticipated mode describes that the current search is "anticipated" to be continued in the future—but it is not clear whether the "resuming" would happen or not: the resuming would happen only if the searcher actually comes back to continue this current search. Otherwise, the current search won't be resumed. Therefore, the "anticipated" as a resuming reason in XSS should be explained as that participants resumed a search session to look for information that they think could be helpful/useful for their task (the information is anticipated to be useful).

The reasons we did not find resuming reasons of transiting, rolling-back, and answer-lost possibly because: (1) All our recruited work tasks are expected to involve multiple search sessions. Participants were aware of their task and had some general ideas about the information that they needed to look for. On the other hand, transiting, rolling-back, and answer-lost imply situations that are unexpected, such as encountering interesting information that ....the found information did not work or the original information was no longer available (2) Another reason could be that during the diary study, we only ask participants to record their information activities related to the selected task. From this perspective, transiting is hard to capture during XSS since (1) it could be manifested by multitasking with multiple search tabs;

The original definition for anticipated could be easily confused with the "transiting" mode because both depict looking for information that is not necessarily needed for solving the current problem.

In addition, the reason we did not find rolling-back may be because most of our tasks are information-gathering tasks that most of our participants were collecting information through the cross-session search process. Or participants may not mention the information that was not useful for them. Instead, as they found the information found from earlier sessions did not work, they may switch to a new subtopic so that the new sessions started because of spawning.

For the five resuming reasons that are not covered by the MISE model, there might be two reasons: (1) the MISE resuming modes mainly focus on reasons related to the involvement of the problem, but these new reasons are not because of changes in the original information problem. On the contrary, in most of the cases, these session resuming reasons happened after the participants' information problem or their task requirement had already settled down. These reasons are more about the related information space and searchers' perspectives of the space and their expectations about the task outcomes. (2) the original MISE resuming modes are from a theoretical model developed based on the review of the literature. On the contrary, these extra five reasons are from the analysis of real users' search experiences.

6.2.2 Session stopping reasons

Our findings identified five major *session stopping reasons*: *goals achieved, satisfaction, interruption, cannot find the needed information, to consult others, and low task knowledge.* Among them, *Goals achieved* and *satisfaction* are the top two reasons that caused people to stop a session. Participants often described the results of their cognitive process of the found information during the session, such as inferring, evaluating, synthesizing, applying, and then decided to stop since their goals of the session were achieved or they felt satisfied with the current findings of the session. *Goal achieved* and *satisfied* has long been noticed as the common reasons that cause people to stop information-seeking behaviors at the conclusion of the tasks (W.-C. Wu & Kelly, 2014). Researchers found that people stopped their interactions with specific information systems when they reached a "frustration or satisfaction-point" (Cooper, 1968) when they encountered a certain number of desired search results (Bates, 1984), found enough information (March, 1994; Zach, 2005; Marchionini, 1997). This might also be the reasons why they were not included in Lin and Belkin (2005)'s original discussion of XSS stopping reasons since from the theoretical perspective of a longitudinal search process, these two reasons should not be considered before the entire task is completed.

The satisfaction and found enough information reasons are similar to *satisfied* reason here we found; however, in XSS, the *satisfaction* could happen both for 1) a session before the entire task is complete, and 2) after a task is completed, depending on the task stage. In other words, there might be two types of *satisfied* for XSS: interim satisfaction and final satisfaction. An interim satisfied session happens for sessions before a task is complete, while a final satisfied happens if the final session of an entire XSS process is completed. An interim satisfied session could be because a participant felt satisfied with a search task or multiple search tasks they conducted during a session, whereas a final satisfied would happen when participants decide that based on all the information they found through all sessions, they felt satisfied with the findings for the task. W.-C. Wu and Kelly (2014) found that there are differences in the factors that affect participants making different stopping points during the information search process: participants focused more on the properties of search results, queries used for search if they wanted to stop a specific search by using a specific query (query abandonment) during the search process; On the other hand, the content they had examined, the goal they wished to achieve, the subjective feelings, as well as the study (task) constraints are the main factors they considered for task stopping. These distinguish possible could be applied to XSS satisfied stopping points.

In addition, different from Lin and Belkin (2005)'s original discussion that people often stopped searching to conduct other cognitive activities to process the found information, our participants generally described they stopped the sessions after they cognitively processed the found information. This reflects two points: 1) participants considered the relevant cognitive activities as a part of their search process instead of separating them from searching. 2) It provides evidence that during information search sessions, searching and other cognitive activities are intertwined throughout the process. This might be because it is an automatic process in which people cognitively interact with the information they find to decide whether it is useful and how it can be used to complete their tasks. Another reason that few participants just collected the information and stopped searching to digest the information later might be because participants did not have pressure (i.e., time or economics). Instead, they had enough time to consume the information as they searched, and even if they could not finish, they knew that it would be easy to come back later to review the information in the next session.

Back in the early time (like at the beginning of 2000 when Lin and Belkin published their work on MISE), the internet was expensive and not so convenient to be accessed at any time or anywhere. It is not unusual that people might need to get as much information as possible within the limited time when they can access the internet and then process (i.e., assimilate, evaluate, synthesize) the found information offline–as a strategy to cut costs. Nowadays, users can easily search for information almost anytime and anywhere, and thanks to the stable internet, users can re-retrieve, re-find, and review the information they saw earlier very easily, so they can digest the found information as they still connect to the web and back to search whenever they needed. Sometimes people stopped to process the found information for more complicated cognitive activities, often because they sought a large amount of information and felt information overload or there were too many ideas/directions to be considered.

In Lin and Belkin (2005)'s original classification of session stopping reasons, they treated *all* stopping reasons as interruptions: self-motivated interruptions that related to internal factors of the task and imposed interruptions due to external factors induced by the environment including mental or physical fatigue. Participants in our study described both types of interruptions that caused them stopped sessions during their task process.

An important difference between self-initiated interruption and external interruption is whether the task-doer was aware of what happened-this is important as it affects participants' cognitive preparation and behaviors for preparing stopping and resuming (Darmoul, Ahmad, Ghaleb, & Alkahtani, 2015; McFarlane, 1998; Dabbish, Mark, & González, 2011). In the self-initiated interruption examples, participants knew about their mental/physical status, were exhausted, needed to take a break, or there was something else they needed to work on soon. On the contrary, when the interruptions were triggered by external sources, they often were less prepared or had no time to prepare for resumption when they were required to stop. The influence of these two types of interruptions on searching would be different: For the self-initiated interruptions, participants could predict and prepare for search stopping. They can choose when to stop, and which query to stop or stop for a specific search result; for instance, they could leave the search pages and tabs open, or they can save the found information. But in cases of external interruptions, they may have fewer opportunities to prepare for resumption or make a conclusion of the search. In the former situation, it is possible that the future information systems can suggest a stopping point or provide suggestions about the session goal and estimate the time that needs to complete the session goal at the beginning—this function is developed by some news reading applications that they suggest how long it might take to complete the reading of an article.

Compared to *satisfied, goal achieved* and *interruption*, we found a few examples of sessions that were stopped because of *cannot find the needed information, to consult others* or due to *the low task knowledge.* Little has been discussed by previous research either. Our results show that participants used different strategies to cope with the situations in XSS when they could not find something they needed. For instance, they tried to look for alternative solutions, or they decided *not* to look for the unfound information again but switched their attention to some other subtopics of the task. These different strategies may explain why there are only a few sessions that ended by *cannot find the needed information*: participants either did not think the unfound information was important that they had to find to complete their task or because there is always something else that could be used instead. Another underlying reason might be because *many of* these tasks were open-ended (i.e., learning, exploring) or self-assigned (i.e., no pre-set requirements), which made it rather flexible about the information needed to be found—which is another reason that can end by *satisfied.* Therefore, participants did not have to find certain information, and they didn't have to search again to find what they did not find. *To consult others* could cause participants to stop searching, especially when 1) their tasks involve others as task collaborators or people who will participate in the task, or 2) the opinions of others could help make further decisions since those people are more authority with expertise knowledge of relevant information. This aligns with our previous survey study finding that people play an important role in XSS and they are important information sources for XSS tasks (Li, Capra, & Zhang, 2020b). Our findings include one example when participants stopped a session because they felt unclear about their task requirements. This might be because as the participant found more and more information over time, their original understanding of the task changed, or their search directions drifted/distracted by the information they were not expecting. This is a common phenomenon during XSS, although it not always causes session stopping. But this might explain why many search assistant prototypes are developed to help people keep track of their information search activities, especially for complex tasks and tasks spread over time. But it should be careful about what to keep and how to assist searchers in navigating the information space and being aware of what they planned to do but also open to novel ideas as their searching spread over time.

In Lin and Belkin (2005)'s MISE model discussion, they indicated that during the information-seeking process, a searcher could stop a search session because they need to validate the information. In the current study, we did not find any participants who mentioned that they needed to stop searching to validate the information found during the process, either as the main stopping reason or as one of the multiple reasons for a session. This might be because of several reasons: participants could judge the authority of the information already when they process the information during the session. They may search across multiple information resources online to verify the information so that they don't need to check with other resources offline afterward. It may also be because current online resources provide enough details (e.g., URLs, different information sources for similar content, others' reviews) to help participants make their judgments; In addition, much of the information during XSS is required to help understand the task topic or gain inspirations and ideas.

The MISE model also includes a task restriction-related stopping reason, *deadline approaching*, which was also found as a stopping reason in our previous AMT study (Li et al., 2020a). However, from the diary study data, we did not find a deadline as a reason that directly led to a session stop, although several participants mentioned that they resumed their last search sessions because they wanted to conclude the search for the task. However, these were not considered as task deadlines were approaching. None participants explicitly said their sessions were stopped because of task deadlines. It suggests that the deadline could be a trigger to start a search session with the goal of completing the search. However, the search session would be stopped when participants achieved their goal (i.e., found the needed information), was satisfied with what they found (including found alternative information), or did not find what they needed. This is different from lab settings, where participants have to stop searching when they use up the time assigned to a task.

Regarding the relationships between the session resuming and stopping reasons, we were interested in exploring whether the reasons that people start a session would affect the reasons that make them stop the session later. We did not find statistically significant associations between the session resuming and stopping reasons. However, we observed some interesting trends. For instance, sessions resumed searching for potentially useful information (anticipated) often stopped because of interruptions than sessions resumed because of other reasons. Recently, Urgo and Arguello (2023) found that during learning oriented searching process, participants with high-quality self-set subgoals for their tasks led to best search outcomes compared to participants with assigned subgoals or no search goals. This could explain why the anticipated sessions were not associated with *goals achieved* or *satisfied* may be because participants during these sessions did not have clear search goals and just browsing for inspiration, for which they are more tolerant of being stopped in the middle without desired outcomes. The future design might consider helping searchers to set up subgoals if a session were detected to search for *anticipated* information. Another interesting observation is that the stopping reason *to consult other information sources* is most popular for sessions resumed because of *unanswered*. It echos the other fact that few *unanswered* sessions are stopped because of *goals achieved*, indicating that people did not find the needed information online and they turned to other information sources, usually people, to look for help. It would be interesting to continue investigating the purposes of the consulting: do they want to get information from others to help understand the task, or do they want to look for information that was hard to find online or do they just need to collaborate with others? *Cannot find the needed information* was only found for sessions resumed because of *spawn*, indicating that sessions with specific requirements may more easily lead to search fails. This would be contradicted to previous encouragement of helping searchers set high-quality search goals during a session since specific, detailed requirements might not be completed or satisfied by a session. Overall, our findings indicate that the relationships between session resuming reasons and stopping reasons are complicated. Future research may need to take more factors of XSS into consideration when analyzing their relationships. Another interesting direction of the resuming reason and stopping reasons is between sessions-whether the previous session's stopping reason affects the next session's resuming reason.

6.3 Session goal complexity, information types, cognitive activities

In RQ2, we used A&K's taxonomy of knowledge types and cognition as the framework analyzed the complexity of session goals at six different complexity levels, including retrieve, understand, apply, analyze, evaluate, and create; we also classified the types of information our participants retrieved during their XSS process into four categories including factual information, conceptual information, procedural information, and opinion/ideas. We then analyzed the associated cognitive activities about how they dealt with the found information to help their tasks at the same six cognitive complexity levels: retrieving, understanding, applying, analyzing, evaluating, and creating. We then selected the most complex cognitive activity to represent the complexity level of the cognitive activity for the associated session for further analysis.

For RQ2a, we analyzed the relationships between the session goal complexity and the information types. A Chi-Square test found that the relationship between the complexity of session goals and the information types is significant, meaning the session goal complexity can affect the information types participants found during sessions. Further analysis found that almost all types of information, except *conceptual* information, can be found across sessions with goals at all six complexity levels. *Factual* is very common for sessions with goals at the lowest complexity levels (*retrieve*) and for sessions with higher to highest complexity levels (*analyze, evaluate, create*). It is common for sessions with goals at the *understand* and *apply* levels but not as common as for others. *Conceptual* information is found for sessions with goals at *retrieve, understand, analyze*, and *evaluate* levels. It is more popular for the *understand, analyze* sessions than for the *retrieve and evaluate* sessions. *Procedural* information is more common for sessions with goals at the *apply, create* levels than for sessions with goals at other complexity levels. *Opinion* is more common for the *apply* sessions than it is for other sessions.

For RQ2b, we analyzed the relationships between the session goal complexity and the cognitive activities participants conducted during XSS sessions. The Chi-Square test found that there is a significant relationship between the two variables. The comparison of the proportion of the cognitive activities for sessions with goals at different complexity levels shows that: *Retrieving* is rarely the only cognitive activity conducted during a session. *Understanding* as the most complex cognitive activity is more popular for *understand* sessions than the *retrieve, analyze, evaluate* sessions and was not found as the most complicated cognitive activity for *apply, create* sessions. *Applying* as the most complicated cognitive activity can be found for sessions at all six complexity levels. It is very common for *apply* sessions and less popular for *evaluate* sessions. *Analyzing* can be found as the most complex cognitive activity for *retrieve, understand, analyze, evaluate* sessions but not for *apply, create* sessions. It is more common for *retrieve, evaluate* sessions than for the *understand, analyze* sessions. *Evaluating* can be found as the most complex cognitive activity for *retrieve, understand, analyze, evaluate, create* sessions but not in the *apply* sessions. It is more popular for *retrieve, analyze, evaluate, create* than for the *understand* sessions. *Creating* can be found as the most complex cognitive activity for all types of sessions. But it is more common for *apply, create* sessions than for the *retrieve, understand, analyze, evaluate* sessions.

For RQ2c, we analyzed the relationships between the session goal complexity and the session resuming and stopping reasons. We did not find significant associations between the variables. Based on the descriptive statistic data, we observed that, generally, the *spawning* is popular for sessions at all different complexity levels, and it is the most popular session resuming reason compared to other resuming reasons within the sessions at the same complexity levels. *Transmuting* also can be found across all types of sessions, and it is more popular for *understand, apply* sessions than for other sessions. *unanswered* not found for *apply* sessions and is slightly more common for *create and evaluate* sessions than for *retrieve, understand, analyze* sessions. *Cultivated* was found for *retrieve, analyze, evaluate* sessions and is more common for the two later types of sessions. *Anticipated* resuming sessions were associated with *retrieve, understand* sessions but not at all for any other types of sessions.

In terms of session stopping reasons, we found *goal achieved, interruptions* can be found for all different types of sessions and are the only two stopping reasons for *apply* sessions. *Goal achieved* stopping reason was least popular for *understand* sessions than for other sessions. Whereas *interruption* is more popular for *understand, apply* sessions than for others. *Satisfied* is common for *retrieve, understand, evaluate* sessions but less popular for *analyze, create* sessions, and was not found for *apply* sessions. *Consult, cannot find, and low task knowledge* was only found to be associated with *analyze, evaluate* sessions but not for other sessions.

6.3.1 Session goal complexity and information types

Task complexity has long been studied by previous research, ranging from its definition to its impact on searchers' perceptions of tasks, search systems, and their search behaviors. In RQ2, our analysis focused on studying the relationships between the complexity of session goals that participants set for individual sessions during their XSS processes, 1) the types of information they found useful during the search sessions, 2) the complexity of the associated cognitive activities about how users thought/dealt with the found information, and 3) the session resuming and stopping reasons, separately. Our quantitative analysis found that the session goal complexity is significantly associated with the types of information participants found useful during sessions, and the complexity of their cognitive activities during XSS, respectively. The quantitative analysis did not find significant associations between the session goal complexity and the session resuming and stopping reasons. We explained some interesting observations of these relationships based on descriptive statistics.

In terms of the relationships between the session goal complexity and the types of information, participants found useful during XSS, our results show that there are significant associations between the session goal complexity and the types of useful information participants found during XSS, which indicates that session goal complexity could affect the types of information people sought during XSS. In other words, it indicates that as participants set up their goals of a search session, they also consider the types of information they need to find for the session, as being said:

> "End users search for information to solve a problem or to make progress on the task at hand. They have a more or less developed mental model of the type of information required." (Serola & Vakkari, 2005, p.373)

The percentage distribution of different types of information for sessions with different complexity levels of goals shows that: besides sessions with goals at the *retrieval* level, sessions with goals at the levels of *evaluate, analyze* and *create* required a great amount of factual information than other types of information. This is contrary to prior studies that found users often need more factual information for simple tasks than that for complex tasks (Saastamoinen, Kumpulainen, & Järvelin, 2012; Choi et al., 2019). Our findings indicate that for factual information, not all participants were only satisfied by finding the information. Instead, they looked for factual information for more complicated goals that require analyzing, evaluating, and creating. This could inform future studies about how and for what purposes people interact with factual information for tasks with complicated goals.

We found that conceptual information was mostly needed for session goals at the *understand* level. This is in line with previous findings by (Choi et al., 2019). However, different from (Choi et al., 2019)'s conclusion that conceptual information was more useful for complex tasks than simple tasks, our results indicate that the use of conceptual information was not very popular for sessions whose goals were at a more complex level in our case: only a few participants mentioned sought for conceptual information for session whose goals at the *evaluate* level and no participants mentioned use conceptual information for *apply, create* sessions. This might be because 1) usually conceptual information is about definitions, which are the most basic, objective elements of a knowledge area, so there is less need for evaluation, and it cannot be directly applied for problem-solving for session goals at the *apply* level, 2) we did not find conceptual information for sessions whose goal is to create might be because, during the creative stages of a task, participants already have thorough knowledge about the topic areas and they no longer need conceptual information to help understand the basic elements related to their tasks.

In addition, our findings show that procedural information was especially popular for sessions whose goals at the *apply* and *create* levels, which reflects participants' needs for applicable information when they need to use the information to solve problems or generate new ideas by adapting the found information to their own situation. This aligns with (Urgo & Arguello, 2022)'s observation that people often engage in creative activities (e.g., modifying, combining) when they work on tasks involving procedural information. Byström and Järvelin (1995) found that problem-solving information is important for complex tasks. Zhang et al. (2020) found that people looked for procedural information during creative tasks to learn how to work on something, evaluate and select ideas and plan their projects. For instance, People who looked for procedural information that was not necessary would follow the procedure to actually work. Instead, they were more about learning the knowledge. at the beginning of a task, a person wants to gain an overview of how to do something, so they may search for general process information to help gain a sense of the work process. This is especially important for learning new skills or working on projects that users are not familiar with. As A&K's indicates that *apply* sometimes involves creative use of the found/learned information, together with the previous research, our findings suggest that future studies may merge the *apply* tasks with *create* tasks and investigate users' purposes of using procedural information at all complexity levels instead of just for more complex tasks.

Contrary to (Choi et al., 2019)'s work, our results show that participants looked for opinions across all levels of session goals rather than just for complex tasks. This might be because 1) opinions is not the main purpose of a search session; 2) opinions often come together with other types of information, such as opinion for factual information, opinion about specific definitions (concepts), opinions about a certain process, etc. Consequently, users found some opinions useful to help understand other associated information. The contrary also indicates the different use of different types of information between single-session search and XSS, which merit future investigation.

6.3.2 Session goal complexity and session's most complex cognitive activities

As for the relationships between the session goal complexity and the actual cognitive activities, we wanted to see if the session goal complexity affects users' cognitive activities, especially the most complicated cognitive activities during a search session. It means that, as we know that people may conduct different types of cognitive activities during the same search session and even with the same piece of information. Here we focused on the most complicated cognitive activity in that a participant described how they dealt with the found information during a session with the goal at a certain level of complexity. One of the reasons is that different from the information types that are exclusive from each other, human cognitive activities are not exclusive from each other. On the contrary, higher-level cognitive activity often contains lower-level cognitive activities. For instance, a participant uses the found information to generate novel ideas (i.e., create), has already found the needed information (retrieved), understood the found information and its relation with their own task needs, analyzed or evaluated the appropriateness of the information, although may not exactly follow the same way with all these steps.

As many prior studies focused on the relationships between task complexity and users' cognitive activities during controlled tasks with a single search session, little attention has been paid to users' cognitive activities during XSS. Some researchers studied users' query formulation and reformulation behavior during the cross-device search, which often involves cross-session search, and found common query activities include narrowing search, expansion search, translation, etc. (D. Wu et al., 2018; K. Huang, Zhang, Qin, & Guo, 2015) However, their analysis mainly focused on participants' interaction with the search systems and did not discuss how users interacted with the found information afterward.

In our preliminary AMT XSS study, we found that XSWT (cross-session work task) often involves multiple cognitive activities (ranging from 2 to 6) (Li et al., 2021a) during the *most recent search session*. This finding is generally saying that an individual search session (i.e., the most recent search session) includes multiple types of cognitive activities without taking the type or the amount of information found during the session. In the current study, we found a similar trend that individual search sessions often involve multiple types of cognitive activities *when participants deal with the information they found*—the thinking process a participant described how they thought about a specific piece of information found during a session often involve multiple levels of cognitive activities. Consequently, our intention is justified by using the most complicated cognitive activity to represent the sessions' actual cognitive complexity level for the analysis.

Our Chi-Square test results show that there are significant associations between the session goal's complexity level and the most complicated cognitive activity participants actually conducted during a session, although the post-hoc did not find pairwise significances. The descriptive statistics show that any type of cognitive activity could be found for session goals at any complexity level. In other words, there is little chance that a specific type of cognitive activity is associated with a specific level of session goal complexity. For instance, creative activities can be found across all complexity levels of session goals, including sessions with goals at the *retrieve* level. Meanwhile, creative activity was very popular for sessions whose goals were at the *apply* and *create* levels, indicating that during these sessions, participants more easily conducted creative thoughts. We also noticed that creative activities could be found during the early search sessions during XSS, indicating that sometimes, participants started to create new ideas at early task stages involving XSS.

Other possible differences probably can be found for *evaluating* activity between sessions whose goals at the *retrieve, analyze, evaluate, create* level and the sessions with the goals at the *understand* level. Sessions with goals at the *apply* level were not considered because it is also possible that there was not enough sample for session goals at the *apply* level in our study.

Understanding is another activity that is significantly more popular for sessions with goals at the same cognitive complexity levels compared to goals at other complexity levels, which may indicate that unlike *retrieve* sessions that participants may not quite be aware of how they want to deal with the found information, participants were more clear about their goals when they mainly looked for information to help them understand the task requirements or other background information.

Previous research found that task complexity has a significant impact on users' perceptions of search difficulty, the search activities such as clicks on different types of information, the time they spend with different types of information, their motivation to use information as well as the gains people learned from their interaction with information for single session searches (Choi et al., 2019; Capra et al., 2017; Brennan, Kelly, & Arguello, 2014). Urgo et al. (2020) suggests that the significances of different types of cognitive activities affected by task complexity possibly exist between "tasks at the extreme ends of A&K's cognitive process dimension" [p.6]. This should be taken into account for future studies on task complexity in users' cognitive processes during the XSS process. One of the interesting directions could be the study of the effects of the type of learning objective of the XSWT, in other words, the learning objective of the entire task, on the cognitive activities involved during a series of search sessions sequentially. For instance, for a creative work task, what is the sequence of the highest levels of cognitive activities for the individual search sessions as a participant make progress over time, what factors can affect the sequences of the cognitive activities for different sessions, and can future systems suggest general thinking process if the complexity level of a task is detected?

6.3.3 Session goal complexity and session resuming and stopping reasons

As mentioned earlier, the Chi-square tests did not find statistical significance for the relationships between the session goal complexity levels and the session resuming or the stopping reasons.

As to the relationships between the session goal complexity and session resuming reasons, the goals of sessions resumed because of *spawning* could be at any complexity level, and so do sessions resumed because of *transmute*. *Unanswered* was found for session goals at five complexity levels but sessions whose goals are at the *apply* level. *Cultivated* as the resuming reasons were found for both simple sessions (i.e., *retrieve*) and complex sessions (i.e., *analyze, evaluate, create*). *Anticipated* as the resuming reason was only found for sessions at the lower-end complexity levels (i.e., *retrieve, understand*), indicating that people less considered how to process the information found for future use–the main purpose of *anticipated* sessions was to find and understand the information.

For the relationships between session goal complexity and the session stopping reasons, the descriptive statistic shows that *goal achieved* and *interruption* could happen for session goals at any complexity level. *Satisfied* can be the stopping reason for most sessions but not sessions with goals at the *apply* level (again, small sample size). *Cannot find the needed information* was found for sessions with goals at the *analyze, evaluate* levels, and *low task knowledge* was only found for a session with goals at the *analyze* level.

Another possible reason why we did not find significance might be because of the uneven distribution of session goals at different levels. Future studies can investigate the relationship between session goal complexity and session resuming and stopping reasons from two perspectives:

first, session goal complexity could be analyzed by users' *search goal* complexity for the session because a session may sometimes contain multiple search goals or single search goals. The search goal often focuses on users' information needs. Meanwhile, the session resuming reasons are about how people cognitively process to fulfill their needs. Similar research has been done by (Capra et al., 2017), for which study they manipulated the task dimensions and specificity of the terms provided in the task description to evaluate the effects of the manipulation on users' perceptions of task difficulty and their search activities during single search sessions. The study found that sessions with more specific search terms were easier compared to searches without specified terms, and searching for subdimensions of a task is more challenging than for specific items. In the XSS process, this might indicate that *spawn* sessions were easier than *transmute* sessions, whereas *transmute* sessions could be easier to be *satisfied* than *spawn* sessions.

Another comparing perspective could be analyzing session goal complexity based on the task structures. For instance, for work tasks at the same complexity levels, whether the relationships between subtasks can affect the complexity of subtask goals and further affect the session resuming and stopping reasons if people search for information for subtasks over multiple sessions? J. Liu and Belkin (2010) designed tasks with different structures (i.e., parallel vs. dependent) to be completed by multiple sessions to observe users' search behaviors and found that different relationships between subtask structures could affect users' interactions with information and search systems. Their design of the tasks indicates that parallel subtasks could be at the same complexity level, whereas dependent subtasks could be associated with increasing complexity levels and extend users' dwell time and decision time as they search for information over time, but they did not study the effects of task structure on users' perceptions of task complexity or search difficulty. For the XSS process, future studies may design and decompose tasks into subtasks with different structures. For example, parallel subtasks may be similar to *spawn* sessions, and dependent subtasks may be similar to *transmute* sessions.

Last but not least, it should be noted that our analysis here has several limitations: 1) not all participants complete their tasks. As some participants completed their tasks within the diary study range, other participants only completed part of their tasks. For instance, some participants only collect the needed information to prepare for their next step (e.g., P105, P129). The cognitive activities participants conducted during a session are the cognitive activities with the found information, which are not the cognitive activities they do for the entire task. For instance, P129 needed to write a children's book, which is a creative task (at the whole task level), but she only collected information during the study session (understanding level for the work completed during the study session), and the cognitive activities she conducted during all these sessions are understanding—she did not actually apply any information she found to write the book yet. 2) another limitation is that we did not have a restriction on the types of participants' tasks. Different tasks require participants to do different things that involve different cognitive activities. 3) for apply activity, because this question was asked in the post-session questionnaire to capture participants' immediate thoughts about the information found from the session, participants may just think about how they *would* use the information but not actually *use* the information as they described in the responses. 4) what kind of cognitive activities people conducted to retrieve information depends on how much they told us, and to what extent, saying the details about the information, and what they thoughts about the information. Because many of the sessions focus on looking for information without little or no actual activities to apply the information to their tasks, the cognitive activities, therefore, are limited, especially when they involve the use of the information and how the use of information can impact their task progress. For instance, if a participant was looking for the taxi fee from the airport to Disneyland, the cognitive activity to find this information is retrieval. However, we could imagine that the price information may be used by the searcher later to book a shuttle–or he may decide not to book a taxi instead of taking a Uber if taxi fee is too expensive. But we don't know from the description here, except what they thought about the information at the current moment.

6.4 Information types, cognitive activities, and resuming and stopping reasons

6.4.1 Information types and resuming and stopping reasons

We analyzed the relationships between the information types and session resuming and stopping reasons. The purpose of this analysis is to find out whether 1) the reasons that motivated people to search for information are related to the useful information types they found during the session; 2) whether the types of information they found during the sessions relate to the reasons why people stopped a session during XSS process. It should be noted that here we distinguish the direction of the relationships: session resuming reasons come first before the useful information types since a participant would be motivated to search and then look for relevant information; On the contrary, the stopping reasons come after a participant found (or not found) some information during a session.

Regarding the relationships between the session resuming reasons and the information types, our Chi-Square tests found the relation between these two variables was significant, indicating for different session resuming reasons, there were significant differences between the types of information people searched for during the sessions. However, the posthoc tests were not able to find out the pairwise significances, which might be because of the small sample size of the data (many cells have expected counts that were less than 5.). The descriptive statistics show that factual information is popular for sessions resumed by all five reasons, especially for *spawn, cultivated* and *unanswered* sessions. Conceptual information was more popular for *transmute, anticipated* sessions than other sessions. Conceptual information was not found for *cultivated* sessions, whereas opinion was not found for *anticipated* sessions. The results indicate that: 1) no matter why a session was resumed, the factual information type is the most popular information participants searched for during XSS. It may also suggests that factual information could be used for various purposes compared to other information types, 2) *conceptual* information examples were not reported for cultivated information, which might be because conceptual information of a domain is unlikely to be changed frequently, they were used to help understand the basic concepts or relationships for a relevant task topic, so that once people learned about the concepts and the underlying rules they seldom look for the information again because of reasons for checking whether the definitions were changed. 3) *Opinion* information was popular for *transmute* sessions, indicating that people often looked for others' opinions when they needed to understand the task requirements. It was not found for *anticipated* sessions, suggesting that when people were looking for potentially useful information for future activities, others' opinions were less important compared to sessions resumed for other reasons.

Previous research has studied information types and users' motivations for information search from different perspectives. One common perspective is analyzing the relationships between users' information needs and the types of information. For instance, Wilson (1981) suggested that people perform information searching to fulfill their physiological, affective, and cognitive needs. Ingwersen (2000) distinguished different types of information needs based on users' information search intents, including "verificative needs" "ill-defined" or "muddled topic needs". Broder (2002) analyzed users' search queries and classified users' web search intentions into three categories: navigational intent to reach specific websites that users have in mind, informational intent for gaining relevant information that searchers assumed could be found on the web, while transactional intent for finding websites for further interaction like shopping, finding various services, etc. Kellar et al. (2007) built a more detailed typology of users' web task types and intentions: fact-finding, information gathering, browsing, and transactions. Byström studied professional workers' use of information and classified the information into three types based on how the workers use it: task information, task-solving information, and domain information. One of the commonalities among these studies is that they focused on *how the information used* or *for what purpose*, which could easily be affected by users themselves and different task situations.

Some recent studies distinguished users' search purposes and the information types they look for. For example, Zhang et al. (2020) identified six different information types (i.e., procedural knowledge, domain knowledge, finished examples, tips/opinions/recommendations, information about a specific topic, and inspiring information) and distinguished the information types from users' intents about the purposes of using the information like learn/figure out how to do something, seek inspirations, ideation, idea evaluation, plan a project, etc. This classification focuses more on the objective information characteristics than the relevant subjective situations compared to previous research. However, some of the information types and intents are unique since the study is about users' information activities for *creative tasks*, so not all the information types can be generalized to other search tasks. Through an analysis of the effects of task complexity on users' information needs for different types of information, Choi et al. (2019) found out that task cognitive complexity can significantly affect the types of information people expected to look for and the gains they learn from the found information. They found out that 1) facts information was significantly used more for retrieving and analyzing tasks than for understanding and creative tasks, concepts were often used for understanding tasks, and opinions were expected when searching for analyzing tasks, insights (e.g., others' suggestions, recommendations) often more useful for more complex tasks than retrieve tasks. Some of our observations are in line with these findings, such as conceptual information often used for transmuting sessions to help understand task requirements. In addition, Urgo et al. (2020) observed from a lab study and found out that users' need for different types of information is affected by the knowledge types involved in participants' task goals (i.e., factual tasks, conceptual and procedural tasks). For example, conceptual information tasks that asked participants to understand the basic concepts required for more concepts, whereas procedural information tasks required people to find more information on procedures. Together with these previous findings, our results suggest that during XSS, participants have expectations about the types of information that in need for sessions that are initiated/resumed because of different reasons, and these reasons are related to 1) the knowledge types that are required to solve the task problems, 2) the complexity level of the task goals. Furthermore, it may also indicate that for different information types, sessions resumed because of different reasons might be associated with different complexity levels. For instance, sessions resumed because of *transmute* might be easier compared to sessions resumed because of *unanswered* since *transmute* sessions required more conceptual information, whereas *unanswered* sessions required more procedural information as (Urgo et al., 2020) found out that conceptual information tasks were more difficult than procedural tasks. However, due to the complexity and various factors for XSS, this needs further investigation.

Regarding the relationships between the information types and session stopping reasons, similarly, a Chi-square test found that there were overall significant associations between the session stopping reasons and the information types. Since the stopping action happened after participants sought the information, it indicates that the information types could affect the reasons why people stopped their sessions during XSS.

The descriptive statistical analysis shows that *factual, procedural* information types were found across all six types of session stopping reasons. As to the *factual* information, this might be because factual information is so common. Therefore, it could be found for sessions stopped for any type of reason.

On the contrary, as *procedural* information can be found for sessions stopped for all six different stopping reasons, it seems it is more popular with sessions stopped by *interruption* and *low task knowledge*. This might be because 1) procedural information often contains a lot of details and require more domain background knowledge to help to understand. For instance, a person may find a recipe that could contain ingredients, cooking tools, and procedures to follow as she makes a dish, and she needs to stop frequently to check, compare, and confirm whether she did match what she read. Another person who is building a computer by himself needs to understand all the different names of the parts and know the fundamental ways which a computer functions. A lack of background knowledge would stop the person and require extra information to help. Previous research found that participants often experienced challenges like information overload when searching for procedural information. 2) applying procedural information could take more time and energy as people use the found information to solve their problems. (Urgo et al., 2020) found that when working on tasks involving procedural information, people often engaged in creative activities like modifying the protocol or combining procedures from several information sources to fit their own tasks. 3) Researchers also found that people often prefer visual content when they search for procedural information (Pardi, Kammerer, & Gerjets, 2019; Choi, Arguello, & Capra, 2023). It is possible that when watching videos, people often need to stop and replay some parts of the video to make sure they understand, or if the video contains too many steps, a participant may decide to finish it in several sessions. 4) use of procedural information involve "trial-and-error" (Choi et al., 2023), which would also cause interruptions and search for extra information to help fix problems.

Meanwhile, *factual and procedural* information are the only two information types for sessions stopped because of *consulting others, cannot find the needed information* and *low task knowledge.* There were a few sessions with procedural information that were stopped because participants wanted *to consult other information sources.* This is different from previous research that people like to consult other information sources for complex tasks, especially people (i.e., experts or others who have similar experiences) (Byström & Järvelin, 1995; Choi et al., 2023) It is also contrary to our previous AMT survey finding that many people recalled that they stopped a recent search session for an XSS task because they wanted to consult others (Li, Ward, & Capra, 2021b). This might be because, during XSS, people expect the time gaps between sessions to talk with others, so they may take it more as a natural stopping point rather than being forced to stop/interrupt, especially when they reach their goals of the sessions and wanted to talk with other people who participated in the task. It would be interesting to further investigate the differences between 1) sessions stopped unexpectedly to consult other human information sources and 2) sessions stopped expected to consult task collaborators to make decisions about future task steps.

Conceptual, opinion information were only found for sessions stopped because of *goal achieved, satisfied, interruption.* Besides the small sample size, it may also be because people often sought conceptual information and opinion information as peripheral information to help them understand the task and task requirements rather than core information that can be directly used as *factual,* *procedural* information. So that people may have less strict criteria for *goal achieved* and easily be *satisfied* instead of *cannot find the needed information*. In addition, since the information is used to help understand the task, and opinions are about others' views of specific events, people need less support by consulting or lack of knowledge to understand.

6.4.2 Cognitive activities and resuming and stopping reasons

Recall that for the cognitive activities our participants conducted during a session, we selected the highest-complexity level cognitive activity from the examples the participants provided to represent the session's cognitive activity.

Prior researchers introduced search stopping reasons related to cognitive activities first by (Nickles et al., 1995), for which they identified four types of rules/strategies that people use to help determine their stop of searching, for instance, stopping when a mental list with required information was found (i.e., mental list), stopping when they found enough information that reaches a pre-determined level (i.e., magnitude threshold), stopping when the information reaches a saturate status and no new information learned (i.e., difference threshold), and stopping when their mental model of the task is established (i.e., representational stability). However, their analysis did not explain how the actual cognitive activities users conducted during a session helped them achieve these goals.

Few studies investigated the cognitive reasons why people resume or stop search sessions during the XSS process. Lin and Belkin (2005)'s MISE model defined the session resuming reasons from the perspective of task status and stopping reasons as *interruptions*. Several *internal interruptions* indicate the relations between users' cognitive activities and session stopping reasons. For instance, they described that people would stop a session to evaluate–assimilate, synthesize, consult other information resources, validate the found information, or monitor (i.e., evaluate whether the collected information is enough) the found information to decide whether to continue their search. Other studies on XSS generally described that users stopped XSS because they found what they needed, their tasks were completed, or interruptions (MacKay & Watters, 2008a),

Our analysis of the relationships between users' cognitive activities aimed to find out 1) whether participants were aware of the cognitive activities and even set their session goals based on the cognitive activities they wanted/needed to conduct as the motivations to start a search session, 2) whether the cognitive activities (i.e., the highest-complexity level of cognitive activity) is associated with the reasons that cause participants to stop a search session.

Cognitive activities and session resuming reasons The Chi-square test results show that there was an overall significant association between the sessions' highest-complexity level cognitive activity and the session resuming reasons. The descriptive statistics show several interesting trends: 1) *evaluating* is the most popular highest-complexity level cognitive activity that was found across all five resuming reasons. 2) *retrieving* was rarely the most complex cognitive activity for a session. There was only one example when a participant did not describe any higher-complexity level cognitive activity other than retrieving. 3) *creating*, as the highest-complexity level cognitive activity, can be found for sessions resumed because most of the reasons except for sessions resumed because of *cultivated*. Spink, Bateman, and Jansen (1999) found out that people need to conduct successive searches to narrow down their previous search results or extend the previous search. To *narrow down* may require cognitive activities of understanding the information found through the search, analyzing the relationships between the found information and the task requirements, evaluate and making judgments about future search directions. In addition, *extending* the previous searches may involve comparing and evaluating the found information, meanwhile brainstorming future search ideas or creatively applying the found information to the tasks. *Retrieving* as the least popular cognitive activity representation of a session indicates that even for the most simple sessions whose goals might be described as finding certain information, higher complexity level cognitive activities are still involved—participants' cognition continually processes the found information after they found it from the web.

Consequently, it means that to further investigate the relationships between the session resuming reasons and the (expected) highest complexity level of cognitive activity, future studies can focus on the types of cognitive activities that may be involved in sessions resumed for a specific reason from the perspective of task status. For instance, for sessions resumed because of *spawn*, participants may need to decompose their tasks into multiple subtasks and set up sub-search goals, and need to analyze the relationships between the subgoals. *Transmuting* sessions may involve the understanding of a great amount of background knowledge and concepts. *Anticipated* sessions may require participants to creatively imagine how the information would be used in the future. And *clutivated* sessions may require participants to compare and evaluate the latest information to make sure what they found was updated.

Cognitive activities and session stopping reasons Chi-square tests also found statistical significance for the relationships between the highest complexity level of cognitive activity of a session and the session-stopping reasons. The descriptive statistics show some interesting trends: similar to many previous studies on users' (single) search session stopping behavior (Zach, 2005; Berryman, 2006), *satisfied* is a popular session stopping reason that can be found for sessions involving almost all types of cognitive activities as their most complex cognitive activities. In addition, Researchers found that users often stopped searching because the found information reached a pre-determined level of satisfaction or helped them establish a rather stable mental model when they were unfamiliar with the task topic areas (i.e., more complex tasks). Together, this indicates that the commonality of *satisfied* as the stopping reason might be because of the overall XSS task, which is often considered as *complex tasks* rather than activities involved in individual sessions.

Goal achieved was most popular for sessions involving mid-complexity-level cognitive activity *analyzing* as the most complex cognitive activity but was the least popular for sessions with their most complex cognitive activity for the simplest *understanding* sessions and for the sessions whose most complex cognitive active was *creating*, except for the one session whose highest complexity cognitive activity is achieving. Previous research (Brown, 2007; W.-C. Wu & Kelly, 2014) found that users often stopped by following the mental list stopping rule for tasks that can be easily decomposed into discrete elements. Similarly, for sessions that involved *analyzing* as the most complex cognitive activities, participants decomposed their information needs and the found information into smaller units to help understand. As they go through their list, their goals were achieved. On the contrary, sessions that involve *understanding* as the most complex activity may indicate that participants were less familiar with their tasks therefore, it would be hard to determine what to find and when it is enough; meanwhile, sessions involving *creating* as the most complex activity may require more critical thinking, and synthesize (instead of decompose) so that it is harder to decide when a goal was achieved.

Cannot find the needed information was found for *applying* sessions and *evaluating* sessions. Previous research defined *search failed* when people do not find any (at least one) useful result from a search (at query level) (Drabenstott & Weller, 2017; Wang, Liu, Mandal, & Shah, 2017). Wang et al. (2017) found that there were several barriers that why people cannot find useful information and cause search failure, including "too many irrelevant results, lack of topical knowledge, lack of patience, unconfident about finding information, unaware of information sources, and too much information"[p.443]. This suggests that 1) low task knowledge could be a sub reason for *cannot find the needed information*, 2) future studies are needed to find out whether some of the barriers that cause search failure are more easily to be found for *applying* and *evaluating* sessions than sessions involved other cognitive activities as the highest complexity level cognitive activities. For instance, it could be because applying the found information or making judgments about the found information has a higher level threshold about to what extent participants are familiar with their tasks (to avoid low task knowledge), or these sessions require more patience and confidence of finding information and overcome the information overload than other sessions.

The relationships between the cognitive activities and the session stopping reasons show that participants' cognitive activities during a search session can influence the reasons that lead them to stop a search session. The MISE model (Lin & Belkin, 2005) suggests users would stop a session to conduct cognitive activities like understanding, analyzing, evaluating, etc., which focused on *what participants wanted to do* after sessions were stopped. Our results extend the model by noticing that participants' cognitive activities before their search session stopping point can affect the stopping reasons. In other words, research should take both participants' *pre-stopping* cognitive activities and *post-stopping* cognitive activities into account for XSS studies.

6.5 Participants' perception of XSS difficulty

In RQ4, we focused on the relationships among participants' perceptions of different difficulties (i.e., overall task difficulty, search difficulty, difficulty finding enough information, and difficulty integrating the found information) across three task stages: 1) the overall pre-task stage when people had not yet started work on the task they wanted to do during the diary study, 2) session stage which was measured by averaging participants' post-session difficulties for all sessions they conducted for the entire task during the diary study, and 3) the post-task stage when participants completed the diary study after searched for multiple sessions. Our Repeated Measures ANOVA (RMA) analyses show that the task stage has significant effects on participants' perceptions of the overall task/session difficulties and their perceptions of the difficulty to find enough information, but no significance was found for their perceptions of the difficulty to search for information, nor the difficulty to integrate information.

The posthoc tests further identified that the mean of participants' rate of post-session overall difficulty is significantly lower than their pre-task overall task difficulty, and the post-session overall difficulty is also lower than the post-task overall difficulty. This means that participants' perception of the overall task difficulty is *always* significantly higher than the averaged, experienced search session difficulty. In other words, it means that participants felt the average session is less difficult than the overall task. This might be because the overall task contains multiple components and require finding much information. On the other hand, when working on a single session during the XSS process, participants may decompose the entire task into smaller units whose goals are more clear and require less information and are more manageable compared to the entire task. Pioneering researchers distinguished task complexity and task difficulty as two different concepts but noticed that task difficulty could be rooted in task complexity: Campbell described that " certain tasks can be difficult (i.e., require high effort), without necessarily being complex; in contrast, other tasks are difficult because they are complex." [p.45]. Although the original discussion focused on comparing information search tasks, our comparison of the overall difficulty at different XSS task stages extends the comparison to analysis of 1) task difficulty at the work task level and search session/task level; 2) the potential underlying reasons imply that the difficulty of search tasks (i.e., session) could be different from participants' perceptions of the difficulty for the entire work task that motivated the searches, which also reflects our analysis of task complexity at three different levels for XSS work tasks: the overall work task complexity, search task complexity, and search session complexity. The changing of the participant's perceptions of the overall difficulty at different task stages also suggests, especially for lab-controlled, assigned information search task studies, which often set up a *Semi-natural* (like real work) work task scenario for a search task in a single session search, it should be clear to both the researchers and the participants about which pre-/post-task difficulty they are measured, the perceived/experienced difficulty for the imaged work task scenario or the perceived/experienced difficulty for the search task that situated in the fake work task environment. Such kind of distinguish could be important to help future studies to predict task difficulty for XSS tasks as previous research has found that the users' behaviors at the "whole-session level" (e.g., task types, task features, users' characteristics) and "within-session level" features (e.g., document reading time, number of viewed pages) can help predict task difficulty and make adaptations to search results (J. Liu, Gwizdka, Liu, & Belkin, 2010).

The V-shaped overall difficulty indicates that breaking complex tasks into smaller units to complete may help reduce users' perceptions of task difficulty. It may not eliminate the perception of the overall task difficulty eventually (the post-task difficulty was the highest among the three task stages), but reducing the session-level difficulty may motivate participants and encourage them to take action. Urgo and Arguello (2023) found that in single session search, assigned sub-search goals for participants could significantly increase their perceptions of task difficulty compared to conditions when no sub-goals were assigned or self-set subgoals, and participants' perceptions of satisfaction were decreased for tasks with assigned subgoals, although not significant. In our diary study, all participants worked at their own pace to complete their tasks. This might explain the significance of the decreased session difficulty compared to the pre-task and post-task difficulty.

Unlike the V-shaped perception of the overall difficulty, participants' perceptions of the other three key difficulty measurements (i.e., difficulty to search for information, difficulty to find enough information, and difficulty to integrate the found information) all decreased from the highest point at the pre-task stage, although statistic significances were only found for the difficulty to find enough information. Except for the reason that breaking down tasks into smaller units may decrease the task difficulty levels, it is possible because as participants found more and more information for their tasks, they were more confident about finding enough needed information and putting them together to complete their tasks. Or on the other hand, it could be that as some participants become more and more understand their tasks and the available information and information sources, their criteria of relevance, the amount of information, and their goals for the tasks changed–they could become more easily being satisfied based on what they could find and how they could use the found information to achieve their goals. Because as prior researchers explained, setting, context, situation or a search system could influence the decision about how much information is enough (Marchionini, 1997). And Foster (2004) found that through the information-seeking process, users gradually learned when they had enough information through an iterative process of questioning whether they had acquired sufficient information to meet their information needs. It may also explain why among all the session-stopping reasons, *satisfied* is as popular as the *goal achieved* for the XSS process.

The last question we investigated in this section is about the relationships between participants' perceptions of post-session difficulties, including overall post-session difficulty, the difficulty to search for information, the difficulty to find enough information, the difficulty to integrate the found information, and the session resuming and stopping reasons, respectively.

The results of Chi-square tests show no significant association between the session resuming reasons and participants' perceptions of the post-session difficulties, meaning that based on the current analysis, the session resuming reasons have no effects on users' perceptions of search session difficulties. On the contrary, the chi-square test results show significant associations between the session-stopping reasons and all different types of perceived post-session difficulties.

Previous research found that users' perception of search task difficulty can be affected by users' characteristics and task features. For instance, through a comprehensive literature review, Wildemuth et al. (2014) summaries that studies of users' perceptions of task difficulty identified such factors including: "time limitations, specificity of the goals, users' knowledge, levels of clarity and uncertainty, as well as system and document features" [p.1131] Among these listed factors, the specificity of the goals, and levels of clarity and uncertainty closely relate to some features of session resuming reasons. For instance, sessions resumed because of *spawn, updated* may have more specific goals compared to sessions resumed because of *transmute, anticipated*. Meanwhile, other researchers found that users' knowledge could significantly affect one's evaluation of task difficulty, meaning that different people could evaluate task difficulty at different levels as the users' task knowledge are different (Byström & Järvelin, 1995). For instance, a normal decision to an expert could be a genuine decision task to a novice who has less knowledge of tasks (Byström & Järvelin, 1995). From this perspective, for sessions resumed because of the same types of resuming reasons, a participant who had more knowledge of the task topic may evaluate session difficulty differently from another participant who has less background knowledge or skills.

On the other hand, session stopping reasons have significant relationships with the different difficulty measurements. It indicates that the factors that lead people to stop searching could also be factors that affect their perceptions of task difficulties. For instance, Capra et al. (2017) found out that inducting uncertainty into tasks could make it harder. Taking session-stopping reasons into consideration, this might indicate that sessions that stopped for *satisfied* might be evaluated as more difficult than sessions that were stopped because of *goal achieved* as for sessions that were stopped because of *satisfied*, users' needs for information and the sources were less specified as their needs for sessions stopped because of *goal achieved*.

In addition, our previous analysis found that session stopping reasons also have significant associations with information types and the session's most complex cognitive activities. The significant associations between stopping reasons and difficulty measurements indicate that information types and the representative cognitive activities that people conducted during a session could be factors that contribute to task difficulty evaluations. Urgo and Arguello (2023) recently found out that participants' post-task measures, including difficulty to search and search satisfaction, were not affected by the designed task's cognitive processes. This might be because in Urgo's study, the task cognitive process was designed as a manipulated factor. However, in our study, the cognitive process was the *actual* cognitive process that participants conducted during a search session. One example is that a session with the goal complexity level at the *retrieve* level may involve a creative thought process. Thus, future research might distinguish the design/desired cognitive activities from the actual cognitive activities performed during the information search process. Because of the naturalistic features of the tasks for our diary study, we were limited to studying the effects of different task factors on users' perceptions of task difficulty as a subjective variable, and future research may further investigate in those directions.

6.6 Cross-session search challenges

In the last research question, we were curious about the challenges specifically associated with the XSS process. Our qualitative analysis of users' interview data identified four types of cross-session searching challenges users faced: *changing ideas between sessions, losing track of ideas and information, having too much information to process*, and *hard to keep motivated*.

As some of these challenges may be related to challenges people have in single-session searches, some differences are specifically associated with the XSS features. *Changing ideas* could be switching from a query to a different one in a single session search (i.e., query abandonment) (W.-C. Wu & Kelly, 2014). As to XSS, changing ideas could mean switching to a different search direction for the entire task for the following search sessions, it could mean changing by using different search strategies for the same search tasks, or it could mean changing the ideas about the final goals people wanted to accomplish. W. Jones et al. (2006) found that information fragmentation and lack of consistent maintenance of the information from different resources were major problems for long-term tasks.

In single-session searches, these alternatives have rarely been discussed. There are several reasons: 1) the natural complexity of XSS work tasks contains more possibilities about the approaches to the task, multiple dimensions and facets of the task that can be explored, and a certain degree of flexibility about how to complete a task, compared to search tasks for a single-session search. XSS tasks often contain multiple goals, whereas single-session search often focuses on a single goal (Agichtein et al., 2012). 2) In single session search, users "have to " complete the search within a single-sitting period of time. Searchers need to select their search focuses and concentrate on it within a limited time, especially for lab-controlled studies. Whereas for XSS, searchers have more time to take different trials to explore and step back if things do not work, which might cause other problems, such as tasks taking too long to complete or lack of motivation after the tasks have been spread to a long time period and cost both time and energy.

Losing track of ideas and information is common for the XSS process for two reasons: 1) users' cognitive process for XSS could be complicated. For instance, they wanted to explore many different subtask topics and methods to complete the task within and across the multiple sessions as their tasks spread out, so it could be challenging to focus on both exploring and keeping track of each direction of the search and thought processes; 2) not all participants would prepare for stopping to resume their searches. Preparation for stopping means that a person expects to stop ongoing behavior at a certain time point so that they would take actions before, during the interruption, or after returning to the stopped action to help resume the stopped activities. Researchers in Psychology and cognition areas have noticed that there are three main mental strategies people use to help recall/return to interrupted activities: rehearsing what has been done before being interrupted or during the interruption period, leveraging temporal cues to help retrieve where the task was stopped, and incidental encoding to cognitive structures into long-term working memory, which refers to user's ability to hold information actively (Diamond, 2013) (Oulasvirta & Saariluoma, 2006). Therefore, users' memory skills could affect the recall of ideas and information found in information search processes. Researchers of single-session search activities found that people with lower working memory evaluated their workload significantly higher and search engagement significantly lower than the evaluation of people with higher working memory for the same information tasks (Arguello & Choi, 2019).

Together, it implies that the great number of thoughts, ideas, and information during XSS requires a high level of working memory–short-term working memory to keep ideas within sessions and long-term working memory to keep ideas and information across sessions. Participants with different memory skills and their awareness of the use of different strategies to keep track of their ideas and thoughts would significantly affect the resumption of searching and other task activities during the XSS process.

Hard to keep being motivated is another common challenge related to XSS. Given that many of our sample tasks were self-motivated tasks than assigned tasks, there is more freedom about how and when to work on the tasks. On the other hand, it is possible that returning to a stopped search without external pressure could be hard. For instance, previous studies about interruptions during the information search process reported that users found that it is statistically significantly hard to return to complex tasks after interruptions, and these tasks were more easily interrupted than other tasks ironically (Czerwinski et al., 2004). In addition, even for tasks that participants were passionate about, users' search enthusiasms (i.e., flow (Csikszentmihalyi, 1990)) could be gone as the time gap between sessions extended. Search engagement has been a specific topic in IS area. O'Brien, Arguello, and Capra (2020) found that besides task topic, participants' evaluations about their interests and task difficulty can affect their engagement during the information search process for single-session search. Their results show that users' effort can negatively affect their engagement–indicating that more work on the task could decrease users' interest in working on the task during single search sessions. It would be worth for future studies to investigate: 1) whether the efforts that distributed overtime for long-term tasks could have the same effects on users' engagements for their information search activities for the task, 2) whether taking breaks between sessions could help reduce physical and mental fatigue so that people could refresh their mind so that they could be more engaged for the resumed search sessions during XSS.

6.7 Implications

Our findings shed light on many different aspects of cross-session search. Below we discuss the implications from the perspective of supporting the XSS process, task management, and methodology for future study on XSS.

6.7.1 Supporting search resuming and stopping behavior during XSS.

Our results provide contextualized examples of Lin and Belkin's resuming and stopping reasons, provide frequency data about how often our participants reported the reasons, demonstrate that the reasons were able to be understood by our participants, and provide quantitative analysis about the relationships among the resuming and stopping reasons. These findings could help search systems prioritize, rank, and present search results for XSS. For instance, for sessions that resumed because of *spawning*, search systems could show users related concepts and sub-topics related to the task (e.g., a knowledge map about the connected subtopics and important concepts in the task topic area); for sessions resumed because of *transmuting* when the problems get elaborated, systems could help users track how the problem evolved and the relationships of information found; for sessions resumed for *transiting* that the original problem swifts to a completely different problem, systems could help to avoid old information; as in the *rolling back* situations when participants' problem was not solved, systems could remind users of previous "leads" that they chose not to follow; In the *treatment lost* systems could remind user of previously found information and sources; in the sessions resumed because of *Cultivated* when user is trying to stay abreast of a topic, a system could send the user notifications when new, relevant information is available. Last but not least, in the *Anticipated* sessions when a user anticipates a future information need, systems may help the user save inspiring information in order to use it in the future.

Similarly, our findings of session stopping reasons provide insight into different types of support that a system could provide. For example, if the user ended a session because they needed to process the gathered information, the search system could help preserve the search results and session state so that the user could review them. If it ended based on a need to consult other sources, the system could summarize the information and sources that were found during the current session. If ended based on a need to validate information, the system could recommend validation sources. These types of assistance could be provided predicatively or could be provided as always-available tools that users could choose to engage with.

6.7.2 Task support and metacognition support

Our findings of participants' motivations identified two groups of motivations for *expected* XSS search behavior: task-oriented and cognition-oriented motivations. Together with the results of the quantitative analysis for RQ4 on task difficulties and RQ5 about the challenges participants encountered, our results suggest that supporting XSS search should go beyond providing search assistance within sessions and should consider participants' cognitive process and task engagement.

Supporting task management and cognition can go hand in hand: From the task perspective, systems could consider providing support to help users decompose the tasks into smaller units (e.g., from macrotask to microtask). Previous research found that prompting participants to set up their own goals during a single search session can significantly improve their task outcomes (Urgo et al., 2020). As many previous XSS studies assigned subgoals for participants, future research should consider providing the choice or comparing the effects of self-set subgoals and assigned subgoals during XSS. Breaking down complex tasks into manageable units can also help users better understand their task structures, brainstorm how they want to complete the task and develop task work plans. For instance, search systems could provide a workable area so that participants could record their thoughts about the found information, organize the found information and keep track of their ideas. It could also release their working memory workload so that they could focus more on searching and processing the found information to solve their problems. Some previous XSS support tools suggest keeping users' search trials and associated annotations over time (Morris et al., 2008b). But our preliminary study (Li et al., 2020a) found that only a small part of participants during XSS needed to find the information they saw previously. Similarly, our diary study suggests that few sessions were resumed because of *treatment-lost* that a participant needed to refind what they saw previously. Consequently, future search systems may consider providing more support for planning, organizing, brainstorming ideas, and keeping track of thoughts instead of saving search histories. Visualization of the task progress during XSS could be another direction for designing XSS assistant tools. For instance, systems could consider providing space, like a toolbar that contains different XSS work tasks. Within a taskbar, participants could set up subtasks and goals, as well as a timeline to indicate how long/how many sessions they would conduct to achieve the desired goals. As they make progress on a task, the taskbar can change colors or participants could manually mark their progress. Providing visual cues has multiple effects: it can visually remind users about their task process, remind them when they come back, and may motivate them to complete a task when they feel a lack of energy to work on an interrupted task.

6.7.3 Methodology implication

Finally, our research has methodological implications for designing XSS tasks in experimental studies. We have shown that different task resumption reasons imply different goals and search behaviors. In studies investigating task resumption, these differences should be considered. For instance, researchers could design search sessions with specific resuming purposes, such as designing tasks with vague task requirements or the need for searchers to narrow down the task scope. Sessions could also be interrupted before users complete the findings for the session, and in another session, the searcher could be asked to continue the search. Multiple reasons can be contrasted to study their impacts.

Due to the uncontrolled study design, the types of our tasks (e.g., task topics, task complexity, session numbers, session goals complexity) are unevenly distributed. Given that our Chi-square tests found overall significant associations between session goal complexity and the information types, session goal complexity and cognitive activities, information types, and session resuming and stopping reasons, future research on XSS could consider controlling some of the factors to future investigate how these variables affect each other. For instance, our results show significant associations ($p<0.001$) between session resuming reasons and information types. Given a group of successive sessions that were all resumed because of *subtask spawns*, will there be significant differences between participants' requirements for different types of information (i.e., factual, conceptual, opinion, procedural)? Or, given for sessions within the XSS process that require the same type of information, how can the system distinguish sessions that were resumed due to different reasons? Our results imply many potential research ideas in the XSS area, and we hope they can inspire more investigation in the future.

6.8 Limitations

This dissertation study does come with a few limitations worth noting. First and foremost, our participant recruitment was limited to those who expected to work on tasks involving XSS. This could potentially influence users' motivations to engage in multiple search sessions. For example, individuals might approach their tasks differently if they involve an unexpected XSS process compared to tasks that involve expected XSS. An unexpected XSS search could also influence their decisions on when to resume or halt sessions, differing from scenarios involving expected XSS work tasks. Future research could benefit from exploring the contrasts between expected and unexpected XSS processes.

Secondly, our study primarily concentrated on the overall task and session characteristics, encompassing aspects like complexity, difficulty, reasons for resuming or stopping, and cognitive activities. However, we didn't make task outcomes a focal point of our analysis. Owing to the fact that many XSWTs could not be completed within the set timeframe of the diary study (i.e., 2 weeks), many participants' actual work spanned much longer durations. Yet, numerous participants chose to stop recording their work and search processes once they had completed the minimum requirements stipulated by the diary study (i.e., at least three search sessions and task engagement for a minimum of two days).

Thirdly, while all tasks were contributed by our participants, they differed in their origin, with some being self-initiated and others assigned by supervisors. This variation in task types presents a challenge in evaluating their outcomes, given that participants worked at their own pace without any stipulations to complete their tasks during the study period. Moreover, our pool of participants, while diverse in task types, was recruited solely from our university, possibly skewing the representativeness of our sample. This limitation may constrain the generalizability of our findings to the wider population.

Besides, we also noted some limitations of the definitions employed in this study. For instance, a session is a concept that can be difficult to define. In this study, participants were instructed to define their own session boundaries. Defining sessions in different ways could impact the results. We also recognized potential constraints in our qualitative coding process. Although we differentiated between four exclusive types of information for the sake of study simplicity, not all information may be categorized in such a clear-cut manner. Occasionally, a single "piece" of information could straddle two categories, such as concept and opinion. Hence, the classification scheme we employed could be seen as another study limitation.

CHAPTER 7

Conclusion

This dissertation presents a diary study on users' cross-session search behavior in naturalistic settings. The study was grounded upon Lin and Belkin's (2005) theoretical framework on the resuming and stopping reasons for cross-session search. Our analysis focused on characteristics of cross-session search sessions, the information types and cognitive activities participants conducted during cross-session search, and their relationships. Specifically, we analyzed: the characteristics of the cross-session work tasks (RQ0), the session resuming and stopping reasons and their relationships (RQ1), the relationships between session goal complexity, information types, cognitive activities participants conducted during the search sessions, and session resuming and stopping reasons (RQ2), the relationships between information types people found during the sessions, their actual cognitive activities that are associated with the found information, and the session stopping and resuming reasons (RQ3), the effects of task stages on participants' perceptions of task difficulties during XSS (RQ4), and the challenges that are caused by XSS (RQ5).

In RQ0, we provided an overview of the characteristics of the cross-session work tasks involved in our diary study from the following aspects: task topics, task sources, expected task timeline and real time spent on the task, the number of task doers to work on the task, overall work task complexity. The main contribution of this research question is that we identified two groups of motivations that cause participants to conduct expected XSS: task-oriented motivations and cognition-oriented motivations. The task-oriented motivations described task characteristics and users' understanding of tasks that lead them to decide to search for multiple sessions; the cognition-oriented motivations describe users' self-awareness of their cognition capability and preferences as the reasons that made them search across sessions. We also distinguished different types of non-search task activities that participants conducted between their search sessions and proposed three XSS searching (S) and doing (D) patterns: SDSD, DSDS, and SSD. These patterns explain how participants initiated their activities for a task and whether and how they intertwined their searching activities with other non-search task activities.

In RQ1, we adapted Lin and Belkin's (2005) Multiple-Information Search Episode (MISE) framework of the sessions resuming and stopping reasons to analyze the main reasons that cause participants to (re)start a search session and to stop a search session during XSS. Our results show that, in our study, five out of the eight MISE session resuming reasons were found as the main reasons that caused users to start a session for an XSS work task: spawning, transmuting, unanswered/incomplete, cultivated, and anticipated. The other three reasons-answer lost, transit, and rolling back were found examples from participants' descriptions of their search processes but were not identified as major reasons that could cause users to start a new session. We also identified six reasons as side reasons that can cause people to search for multiple sessions that were not explicitly included in the original MISE model: to exhaust available information resources, to explore more topic areas, to find inspirations and examples, to review what they found earlier, consulting others triggers new sessions. As to session stopping reasons, we adjusted Lin and Belkin's (2005) classification of session stopping reasons into six different reasons: satisfied, goal achieved, interruption, cannot find the needed information, need to consult other information sources and low task knowledge. Our quantitative analysis did not find statistically significant associations between session resuming and stopping reasons. However, the descriptive statistics indicate that there were some trends for the distribution of the stopping reasons for different resuming reasons. For instance, *satisfied* and *goal achieved* as the two most popular stopping reasons can be found in sessions resumed for four main MISE resuming reasons. Whereas *consult other information sources, cannot find the needed information and low understanding of goal* were only found associated with some resuming reasons or only one type of resuming reasons. *Interruption* was found to be associated with four resuming reasons but not the updated sessions. It would be interesting for further studies to explore whether it is because those sessions end before interruption or participants were more willing to stop for those sessions as interruptions happen.

For RQ2 and RQ3, we focused our analysis on the relationships among 1) session goal complexity, 2) information types participants sought during their sessions, and 3) the cognitive activities they conducted during the XSS process, as well as the relationships between these factors and the session resuming and stopping reasons. We used Anderson and Krathwohl (2001) 's framework of cognitive complexity to analyze the cognitive complexity level of the session goals based on participants' responses to a question in the pre-session questionnaire at six levels: retrieve, understand, apply, analyze, evaluate, create. Using the same framework, we also analyzed participants' cognitive activities associated with the information examples they provided in the post-session questionnaire and selected the most complex cognitive activity of the session as the representation of the sessions' cognition. We also adopted a classification of information types from (Choi et al., 2019) to analyze the information types of the examples: factual, conceptual, procedural, and opinions/ideas. Our quantitative analysis results suggest that session goal complexity has statistically significant associations with the information types participants found during the sessions and also significantly associated with the most complicated cognitive activities participants conducted during the sessions, The descriptive statistics indicate that factual, procedural, and opinion information can be found for sessions whose goals are at all complexity levels, whereas conceptual information was found for sessions whose goals at the *retrieve, understand, analyze, evaluate* levels but not found for sessions whose goals at the *apply* and *create* levels. In addition, the *apply* and *create* levels involved the same types of information. RQ2 also analyzed the relationships between the session goal complexity and the actual most complicated cognitive activity participants conducted during XSS. The results show there were significant associations between the two variables. The descriptive statistic of the percentage of different cognitive activities that happened at different goal complexity levels indicates that evaluating was the most popular cognitive activity for session goals at the *retrieve, analyze, evaluate* level, *creating* was popular for session goals at the *apply, create* levels but can be found for goals at other levels as well. Understanding was most popular for session goals at the *understanding* level.

In RQ2, we did not find significant associations between the session goal complexity and the session resuming or stopping reasons. But in RQ3, our analysis found significant associations between the session resuming and stopping reasons and the information types. Meanwhile, the resuming and stopping reasons were also found to be significantly associated with the session's most complicated cognitive activities.

Together, these results imply that participants' session goals may not be directly associated with the session resuming and stopping reasons. On the other hand, the session goal complexity does seem to affect users' needs for different types of information and how they cognitively process the found information to achieve their goals. Meanwhile, the latter two factors (i.e., information types and the most complicated cognitive activities of sessions) are significantly associated with the session resuming and stopping reasons. Therefore, the session goal complexity and session resuming and stopping reasons might be associated indirectly through the information types and cognitive activities participants conducted through the search process. Future research can investigate how these factors are directly/indirectly related to each other by more strict controlled study designs.

For RQ4, we compared participants' perceptions of task difficulty (pre-task, post-task) and session difficulty and the relationships between the session difficulty and the corresponding session resuming and stopping reasons. Through these comparisons, we investigated the effect of task stages on participants' perceptions of task/session difficulty and whether the levels of difficulties related to the reasons that cause people to start or stop a session. We compared the difficulties from four aspects: the overall difficulty, the difficulty in searching for information, the difficulty in finding enough information, and the difficulty in integrating the found information. The Repeated Measures ANOVA tests show that task stages significantly affect users' perceptions of overall levels of difficulties and the difficulty to find enough information. Post-hoc tests found that for the overall difficulty, participants' perception of post-session difficulty was significantly lower than the pre-task difficulty and the post-task difficulty. In other words, participants' perception of session-level difficulty was significantly lower than their perceptions of the entire task difficulty. We also found that participants' perception of the post-task difficulty to find enough information was significantly decreased compared to their pre-task perception of finding enough information, which means that they felt the experienced difficulty of finding enough information was significantly lower than the difficulty they expected before they started the searching for the tasks. Overall significances were also found for the comparison between the difficulty to find enough information. In summary, RQ4 found that participants experienced overall difficulty, and the difficulty to search for information for the entire task was higher than for the individual search sessions. On the other hand, their perceptions of difficulties to find enough information and to integrate the found information generally decreased as their tasks moved forward. These findings provide evidence to support helping users decompose complex tasks into smaller units. It also implies that users of XSS tasks may overestimate the difficulties of their tasks from multiple aspects, which could potentially negatively affect their motivations to work on the tasks. Future research may consider helping users appropriately evaluate the difficulty levels and provide tips to encourage users and keep them motivated to complete their tasks.

In RQ5, we briefly discussed XSS challenges that participants experienced which were specifically caused due to the work over time. We summarized five main challenges participants mentioned during the interview about why they thought were caused by searching for multiple sessions, including changing ideas between sessions, losing track of ideas and information, too much information to process, and hard to keep motivated. Compared to the two groups of motivations that cause people to search for multiple sessions, we could find that these challenges are all related to users' cognition workload—how participants' thoughts vary through the XSS process, their regulations (e.g., evaluation, monitoring) of ideas and process of information, and how to keep oneself self-motivated for tasks spread over time. Some of these aspects are closely related to recent topics discussed by the Search As Learning community and could be further investigated by future research to help the learning and exploring process that involves XSS.

This work has implications for multiple subareas in IR and IS: Our analysis of the session resuming and stopping reasons during XSS provides a different perspective to help search engines refine and predict search results compared to previous research that was mainly focused on the used queries or other search histories. In terms of task management, our findings and analysis of the relationships among the complexity and difficulty of the overall XSS work task, and those at the session levels, and their relationships with the information types, cognitive activities, and the session resuming and stopping reasons provide a comprehensive, detailed description about how these aspects were associated with each other based on observations of real-life task examples. We identified cognition challenges that were specifically caused by searching across sessions. These findings provide insights into the design of assisting tools, such as task management and XSS process visualization. Our results also have methodology implications that future design of XSS studies could consider manipulating the tasks from subtask goal levels and task/session complexity and difficulty levels, the desired cognitive activities levels, and design sub-sessions that can be motivated and stopped by different reasons.

APPENDIX A

Diary Study Recruiting and Screening Email

A.1 Recruiting letter

To: UNC mass mailing list (and possibly to other graduate student mailing lists)

Cc: yuanli@live.unc.edu

Subject: Participants needed for a research diary study about cross-session search

—–**Message Text**——–

We are recruiting participants to join a diary study focusing on why and how people search for information online over time to complete a task or a project.

There are some complex tasks or projects that often take a couple of days to complete in our daily work and life. To finish those tasks or projects, we often need to search for information online over time. For instance, a student who looks for a summer internship searches multiple times online to find and apply for job openings. Another person looks for information across days to make a road trip plan. Somebody needs to find information from different sources to complete a research report.

To join our study, your project or task should be one that:

(1) you have not started the task yet,

(2) you need to search for information online to complete the task,

(3) you plan to spend several days (less than two weeks) to complete the task, and

(4) you anticipate to do Internet searches on different times or days as part of completing the task.

Your project or task could be a school assignment, research report, a project related to personal hobbies, or other tasks that have concrete outcomes when you complete it. We recommend you do **NOT** select a task that you may be uncomfortable talking about with others.

During the diary study, we will ask you to:

(1) use a pre-set google study account on your computer when you search information for the task,

(2) fill out pre-/post-session questionnaires and keep a work diary,

(3) share your search history for this task with the researcher (you will be able to remove any private or sensitive information),

(4) search online for information for your task or project for at least three times (across multiple days) over two weeks,

(Optional): By the end of the study, we will recruit participants who completed the diary study for an additional 45 minutes interview about your task experience. You can choose whether or not to participate in the interview.

The diary study will last until you (1) report that you complete your task and study requirements, or (2) decide to withdraw from the study before completion. You can withdraw from the study at any time.

You will receive an Amazon Gift Card with $80 for completing the study, and we will offer an extra $40 to participants who provide high-quality responses for questionnaires and diary forms that contain rich details related to information search and task process and genuine thoughts about their cognitive and other task behaviors. In addition, participants who participate in the interview will receive another $20.

To join this study, you need to be: (1) affiliated with UNC (e.g., part-time/full-time worker at UNC, students, research fellows), (2) over 18 years old, (3) a native English speaker.

If you are interested, click the **link** below to fill out a short screening questionnaire to tell us some basic information about the task or project you select and your plan for the task/project.

We will email the selected participants within two business days after we received your responses. We anticipate the study running through early fall semester 2021 till we complete data collection with 30 participants.

Any data you provide will not be shared outside of our research team.

Principal Investigator: Yuan Li

Principal Investigator Department: School of Information and Library Science at University of North Carolina at Chapel Hill

You will not be offered or receive any special consideration if you take part in this research; it is purely voluntary. This study (IRB #21-1428) has been reviewed by the UNC Institutional Review Board and determined to be exempt from Federal regulations.

A.2 Screening questions

Q1: Please briefly describe the task or project you will work on. The task should be one that your haven't started yet but plan to spend several days and complete it recently (e.g., making a road trip plan, complete a complex research project, looking for a summer job, etc.). Please do **NOT** select any tasks that you may feel uncomfortable talking about with others.

Q2: What will be the final outcome(s) of your task?

Q3: What is the timeline for your task?

Q4: Please briefly tell us why you think the task will take the time (you selected above) to complete?

Q5: How and why will your task involve searching for information online over time?

Q6: Which of the following web browser do you usually use?

Q7: Once being selected, you will receive a scheduling email for a one-hour introduction meeting, please provide three 1-hour time slots you will be available within the next week.

Q8: Thank you for your response, we will send you a follow-up email with 24 hours if you are selected. Please enter your name and email address below:

Your name:

Email:

APPENDIX B

Scheduling Email

To: [participant's email address]

cc: yuanli@live.unc.edu

Subject: introduction meeting invite for Cross-session search diary study

—– **Message Text** —–

Thank you for being interested in our study!

In preparation for your participation in the cross-session search diary study, I am sending the Zoom connection information and a copy of the informed consent form in case you would like to review it ahead of time.

For our session on [DAYOFWEEK, MONTH, DAY starting at TIME], please connect to Zoom with audio and video using the computer that you will use for the study following the information below.

During this initial Zoom meeting, we will go through the study consent form and sign it if you agree to participate in the study. We will also guide you through the process to set up a new Google study account for you to use during the diary study and work on a practice task. In addition, we will ask you to fill out a pre-task questionnaire about the task you selected to participate in this study.

[Zoom information]

[copy information from Zoom invitation similar to below]

[Name] is inviting you to a scheduled Zoom meeting.

Topic: Cross-session search diary study introduction meeting

Time: [Time]

Join Zoom Meeting

[URL]

Meeting ID: [ID #]

Password: [Password]

Please do not share the Zoom information or make it available on any public-facing website or calendar.

If you have any questions or encounter any difficulties connecting, please contact the researcher at yuanli@live.unc.edu or 919-360-6310.

Many thanks,

Yuan Li

You will not be offered or receive any special consideration if you take part in this research; it is purely voluntary. This study (IRB #21-1428) has been reviewed by the UNC Institutional Review Board and determined to be exempt from Federal regulations.

APPENDIX C

Study Introductory Meeting

Thank you for participating in our study!

In today's meeting, we will go through three parts: (1) read and sign the study consent form; (2) set up the Google study account on your computer; (3) walk through the diary study process and questionnaires. You are welcome to ask any questions at any time.

To make it easier to understand, I am going to ask you to share your screen with me. Is that okay for you to share your screen with me?

Wait for the participant to confirm

Before you start to share your screen, please close any windows on your desktop that you won't like to share. Once you are done, you can start to share your screen with me.

Wait participant to start screen share

C.1 Consent form

First, I will send you the consent form link in the Zoom chat, please open the link in your browser and read it, you can ask me any questions if you have, let me know when you are done.

https://unc.az1.qualtrics.com/jfe/form/SV_0capO4OREpF9XPo

C.2 Set-up Google study account

To protect your personal Google account, and make sure the study data won't be mixed with your personal information, we will use a new Google account for your study.

Next, I am going to walk you through the steps to set up the Google study account. Before you start to share your screen, please feel free to close any windows on your desktop that you wouldn't like to share. When you are ready, please start to share your screen.

Set up a new profile in Google browser:

Step 1: Open your Google Chrome browser,

Step 2: Click the profile circle at the top right corner (sometimes it may be a circle with your name initials if you already logged in) and click "Add" at the bottom from the dropdown list to add

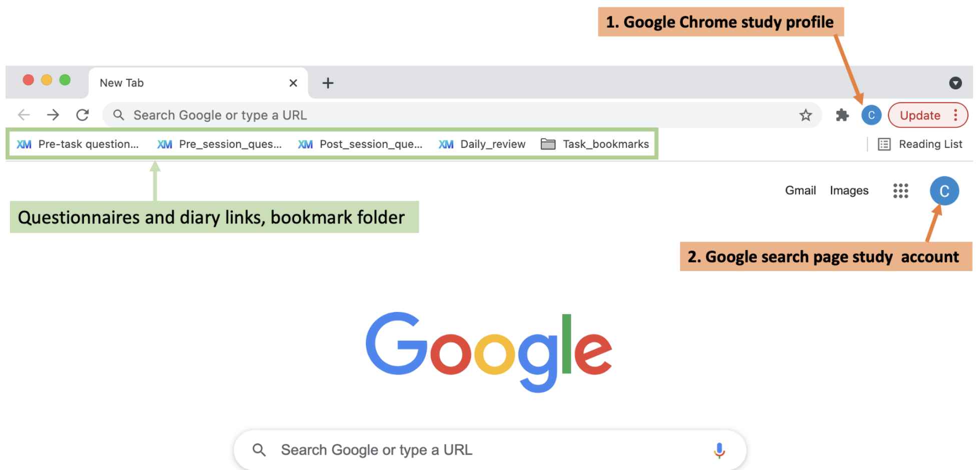


Figure C.1: The Google browser interface logged in with a pre-set study account

a new profile.

Step 3: In the pop-up window, type a name you want to use for this study profile (you can also select a profile picture if you want), and then click "Add" at the bottom.

Now you can see a new browser window.

Log in the study account under the new profile:

Step 5: In the new browser window, click on the new profile circle, and click "turn on sync...", this will allow the researcher to access your search activity for this account remotely.

Step 6: In the popped-up dialogue box, log in by using the study account and password, I am going to send you the account and password by chat:

Your Google study account is: xssdiaryp101@gmail.com, password: XXXXXXXXXX (send by chat)

Please remember this account and password and do not change it.

Now your new profile and study account are set up. On the new Google Chrome window, make sure you also log into the study account (Figure: C.1).

Double check that participant turn the sync on–they can see the pre-set bookmarks Step to turn on sync: (1) click on the "update" button behind the profile; (2) click setting on the drop-down list; (2) turn on sync.

C.3 Study procedure

Once you log into the study profile and study account, you can see several existing bookmarks on the bookmark bar (C.1).

Now I am going to ask you to fill out a pre-task questionnaire, this questionnaire will ask you more details about the task you are going to do over following days for the diary study. You only need to fill out this questionnaire once. You can skip questions if you feel not comfortable to answer. This also applies to any questions in other questionnaires and diary forms through the study. You can stop and ask me any questions if you are not sure.

Now please click the first bookmark link on your browser toolbar to complete the pre-task questionnaire.

You can stop sharing your screen if you want, let me know when you are done.

Wait participants till they complete the pre-task questionnaire.

Do you have any questions about the pre-task questionnaire?

Since the pre-task questionnaire only need to be filled once, now you can go ahead and delete the pre-task questionnaire bookmark so you won't see it in the future.

During the diary study period, you need to follow the following steps to complete the study:

1. Every time when you decide to search for information online for this task, please make sure that you shift to the Google study profile and login with the assigned account,

2. Fill out the *Pre-session questionnaire*,

3. Start search for information for your task as usual,

4. During your search sessions, remember to bookmark useful web pages and save them into the "Diarystudy_bookmark" folder at the bookmark bar,

5. When you complete a search session, click and fill the *Post-session_ qustionnaire.* You need to fill this questionnaire every time when you complete a search session for this task. A session is a continuous period of time that you search for information for this task. One search session may contain multiple search queries,

6. You need to repeat Step 1 to 5 for each search session you conducted,

7. By the end of a day when you worked on the task, you need to fill the "*Daily_ review diary*" (the third bookmark). You need to fill this diary questionnaire at the day if you searched for information or worked on other parts of your task,

8. After you complete your daily diary, please review your search history under the Google study account, you can remove any records that are not relevant to the study or you don't want to share,

9. Repeat Step 7 and 8 for each day you worked on the task till you complete the task.

In the daily review diary, there is a question asking you whether you have completed your task, if you selected yes, I will know that you are done with the diary study, and will connect you for further steps.

Or you can send me an email tell me whenever you decide to withdraw, stop, or complete the task.

Do you have any questions so far?

C.4 Practice task

To introduce you to this process, next, we are going to do a practice task:

Ask participants to share their screens if they haven't yet.

Imagine that you want to buy a 4k television, and would like to find some reviews online:

Switch to the new profile and study account on Google Chrome:

At any time when you want to search for information online for your task, you need to:

(1) Open the Google Chrome browser,

(2) Click the profile circle,

(3) Select the study account under "Other people",

(4) Log into the study account (You can allow the page to remember your account automatically).

Now please close this new account. Open your Google Chrome as usual.

Now please switch to the new study account.

Searching, bookmarking, and recording diary questionnaires: Now you are going to start a search session for this task,

(1) Please click and fill out a pre-session questionnaire for this session,

(2) Okay, now try to find a couple of interesting web pages,

(3) Now, add a bookmark to the bookmark folder,

(4) Next, let's imagine that you are done with this session, please practice filling out the post-session questionnaire based on the television task,

(5) Image that you have completed your search for today. Now, click the daily review diary to fill it.

Please keep in mind that if you need to purchase something, please use your own Google account, since we don't want to log any personal information on this study account.

Examples of high-quality response and poor-quality responses:

During the process when you fill out any of the questionnaires or a diary form, please remember to describe your answers with rich details and genuine thoughts happened during your search process. For instance, a high-quality description of your thoughts during search could be like this:

> "I needed a new computer mouse, so I researched one by comparing suggestions from different websites and reviewing ratings on places like Amazon and Newegg before making my decision. I changed my search terms only a couple of times and relied more on highly reviewed Amazon items when I made a decision. I was able to find something sufficient and cheap by the end and have already ordered it."

On the other hand, a poor-quality response of a search does not include much details, one example could be like this:

> "to check the weather"

in the above examples, the high-quality response described why the person need to look for information, how s/he conducted the search, the examples of websites they consulted, and the search keywords that were used as well as the final search results. But in the poor example, it did not provide any details about why the person need to search information about weather, where the location is about and what is the results. Remember that the more details you provide, the better it can help us understand your search process and your task.

C.5 Clean browser history and log out

After you complete the daily diary questionnaire, remember to (1) review your search history records to make sure that are no irrelevant links to the study task, (2) close the Chrome window or shift back to your personal Google account.

When you use the web for other tasks that are not relevant to this task, please remember to switch back to your normal Google account. Since this account using a profile different from your original ones, it will not remember any of your passwords, or search history you did with your other Google accounts.

In case if you searched irrelevant information using this study account, you can always clean your search history and delete your data from this account. We recommend you to scan your diary study search history everyday when you fill the daily review diary questionnaire.

To clean irrelevant search history records, you can:

(1) Click the menu button (the three vertical dots) and click on "History" and select "History" at the second drop-down list,

(2) In the newly opened history page, you can scan through your browser history to find the ones you don't want to share, or those that are not related to this task,

(3) Click the three vertical dots behind a link you want to delete, and click "Remove from history".

(4) Delete multiple history links: you can check the boxes in front of the links you want to delete, and then click "Delete" at the top of the page to delete all the selected pages together.

To shift back to your personal profile/account, you can:

Close the entire study account browser.

Your personal Google account won't be affected by this new account. Next time, when you open your Google Chrome, it should be your personal account by default, unless you manually change it to the study account.

Or, you can:

(1) On the top right corner of the browser, click on the profile cycle to show the drop-down list,

(2) Select your personal account under "Other people".

Now the study profile will close and your personal account will open in a new Chrome window.

We except that you start the task the day after this introductory meeting, like from tomorrow, and you will have two weeks to complete your task. The minimum requirements include: (1) complete at least three search sessions during the two weeks; (2) complete the pre- and post-sessions questionnaires for the three search sessions; (3) provide at least two daily review diaries; (4) keep your search history for the three search sessions.

You can withdraw or stop the study at any time. But as we said in the consent form, if you stop before your task is complete, or did not provide the minimum required data, we will pay you based on the days you worked on the task for $15 per day, and no more than $45 for an incomplete study.

If you have any questions, please feel free to send me an email: **yuanli@live.unc.edu**. We will respond to your email as soon as possible.

APPENDIX D

Consent Form

The consent form is created using the UNC IRB template and attached as an external Pdf document starting from the next page.

IRB TEMPLATE Version 2.1 - 1/17/2020 - Do not alter this text box

Consent to Participate in a Research Study

IRB Study # 21-1428
Title of Study: How and Why do people search across sessions for complex tasks: A diary study
Principal Investigator: Yuan Li
Principal Investigator Department: School of Information and Library Science
Principal Investigator Phone number: (919) 360-6310
Principal Investigator Email Address: yuanli@live.unc.edu
Faculty Advisor: Robert Capra
Faculty Advisor Contact Information: 919-962-9978
Funding Source and/or Sponsor: National Science Foundation (NSF)

CONCISE SUMMARY

The purpose of this study is to investigate how and why do people search for information online across multiple sessions overtime to complete a task. In our daily work and life, people often need to search for information at different times or stages when they work on a complex task that they cannot complete in one-sitting. To investigate the reasons that lead people to conduct multiple session searches, we are recruiting 30 participants for an online diary study. Participants will provide information about a task or a project of their own, which they plan to work on across multiple days and search for information for multiple times online to complete the task. Once recruited, we will first schedule a one-hour Zoom meeting to introduce the study and set up the diary materials. Participants will use an assigned Google study account on their own computer when they search for information online for their task. They will need to fill out a pre-session questionnaire and a post-session questionnaire each time they conduct a search session for the study task. By the end of a day when they worked on the task, participants are asked to record their work process in a daily work review diary. Their search history will be recorded under the assigned Google study accounts and shared with the research team. Each participant is expected to provide questionnaire responses for at least three search sessions, and at least two daily work review diaries. The diary study period for each participant will last till they complete their tasks but no longer than 2 weeks. We will also select 10 to 12 participants from those who completed the diary study for an additional 45-minute interview via Zoom. Participants who participated in the interview will be paid with additional compensation.

We are recruiting participants who are UNC students or employees. We do not anticipate that this study involves more than minimal risks. However, suppose the data collected about a participant's cross-session search process was breached and associated with a particular participant, it could lead to social, psychological, or economic risks for the participant. To minimize the risks, all of data collected will be stored confidentially in the research team's UNC OneDrive folder. The data for this study will not be shared with anybody outside of the research team. We also ask participants not to select a task for this study that they would be uncomfortable talking about with other.

This study is designed to benefit society by gaining new knowledge. Participants will not benefit personally from being in this research study.

If you are interested in learning more about this study, please continue to read below.

What are some general things you should know about research studies? You are being asked to take part in a research study. To join the study is voluntary. You may choose not to participate, or you may withdraw your consent to be in the study, for any reason, without penalty. Details about this study are discussed below. It is important that you understand this information so that you can make an informed choice about being in this research study.

What is the purpose of this study? The purpose of this research study is to understand how and why people search across multiple sessions to complete complex tasks in their work and daily life. Cross-session searches are very common. It describes when people search online for information over several days or multiple times in a day to complete one task or project.

We have designed this diary study to investigate the reasons that lead people to search across session, how the reasons relate to the information they seek, and how they cognitively process the found information. This study asks people to take write down notes (electronic diary entries) about their search sessions when they happen. This can help us better understand the reasons for each search and how they relate to aspects of the overall task. We hope that the results of this study can provide insights into how to design search systems that are more helpful for complex tasks.

How many people will take part in this study? If you decide to be in this study, you will be one of approximately 30 people in this research study.

What will happen if you take part in the study? It is expected that the diary study will start 1 to 2 workdays after the introduction meeting. You are expected to work and record your task experiences for at least two days but no longer than 2-weeks. During the 2-week study period, you will search for information for your selected tasks across multiple sessions overtime and fill out the required questionnaires and diaries based on your information search and work process for the task.

First, you will be assigned with a pre-set Google study account. This account is created for this study specifically. The account assigned to you is protected with a unique password. Please note that the research team can also access this Google account remotely. During the study, you need to log into this account on your own computer when you work on your selected task. By doing so, the account will only record your search history for this task. If you searched any irrelevant information you can review and delete your search history at any time. In this Google account, there are four bookmarks in the browser toolbar: the pre-task questionnaire, pre-session questionnaire, post-session questionnaire, and a bookmark folder.

You will fill out the pre-task questionnaire during the introduction meeting. You only need to fill out this questionnaire once. And then the link will be deleted after you complete the questionnaire.

Once you start an information search session on a specific time in a day during the study period, you are required to follow the following steps:

1. You need to log into the Google study account,
2. You should fill out the pre-session questionnaire about the goal you want to achieve for that session,
3. You can start to search for information as usual,
4. You can bookmark useful pages for your tasks into the bookmark folder which can help you answer questions later in the post-session questionnaire,
5. Once you decide to stop the search session, you should fill out the post-session questionnaire about your search experience (e.g., the useful information you found during the session, how you will use the information to complete the task, any difficulties and challenges you had during the session),
6. You can log out the Google study account after you stop searching for the task.

By the end of a day when you worked on the task (e.g., sought for information online, or worked on other non-information searching parts of the task), you need to:

1. Log into the Google study account,
2. Complete the daily review diary about you work experience for the task during that day,
3. Check your search history to make sure there is no irrelevant search logs to the task, or you can delete any search history that you don't want to share with the research team,
4. Log out the Google study account.

In the daily review work diary, there is a question asking you whether you complete the entire task or not, if you answer yes, the researchers will be notified that your task is complete and you would no longer work on it any more. The researcher will review your questionnaire responses and daily diaries within a few days and decide whether you are qualified for the retrospective interview. If you are selected, we will send an invite email for the interview and schedule a 45-minute interview at your convenience. At the end of your participation in the study, we will send you the study compensation using an electronic Amazon gift card via email.

The interview will be on Zoom. During the interview, we will ask questions about your search experience and process to complete the task. We will also ask you questions about your responses to the questionnaires and diaries. The interview will be recorded: audio recordings will be made along with a recording of the shared computer screen that will be show your diary entries and browsing history organized in an Excel spreadsheet that exported from the Qualtrics and Google search history. During the Zoom interview, your name on Zoom will be replaced by your participant's ID (the assigned Google study account name) so that your real name won't be recorded. We will NOT record your webcam/face images either. The interview is completely voluntary. Your total payment for the parts you already completed won't be affected if you decide not to participate the interview.

What are the possible benefits from being in this study? Research is designed to benefit society by gaining new knowledge. You may not benefit personally from being in this research study.

What are the possible risks or discomforts involved from being in this study? Since this study will ask you to record your search history and answer questions about a task you are doing in real life, you should select a task that you feel comfortable talking about to other people. For example, you should not select a task that is private in nature or that is a confidential part of your work.

You may feel uncomfortable or embarrassed if you are unable to complete the task or to provide answers to the questionnaires.

What if you are a UNC employee? Taking part in this research is not a part of your University duties, and refusing will not affect your job. You will not be offered or receive any special job-related consideration if you take part in this research.

What if you are a UNC student? You may choose not to be in the study or to stop being in the study before it is over at any time. This will not affect your class standing or grades at UNC-Chapel Hill. You will not be offered or receive any special consideration if you take part in this research.

What is a Certificate of Confidentiality? This research is covered by a Certificate of Confidentiality. With this Certificate, the researchers may not disclose or use information, documents or biospecimens that may identify you in any federal, state, or local civil, criminal, administrative, legislative, or other proceedings in the United States, for example, if there is a court subpoena, unless you have consented for this use.

The Certificate cannot be used to refuse a request for information from personnel of a federal or state agency that is sponsoring the study for auditing or evaluation purposes or for information that must be disclosed in order to meet the requirements of the federal Food and Drug Administration (FDA).

The Certificate of Confidentiality will not be used to prevent disclosure as required by federal, state, or local law, such as mandatory reporting requirements for child abuse or neglect, disabled adult abuse or neglect, communicable diseases, injuries caused by suspected criminal violence, cancer diagnosis or benign brain or central nervous system tumors or other mandatory reporting requirement under applicable law. The Certificate of Confidentiality will not be used if disclosure is for other scientific research, as allowed by federal regulations protecting research subjects or for any purpose you have consented to in this informed consent document.

You should understand that a Certificate of Confidentiality does not prevent you from voluntarily releasing information about yourself or your involvement in this research. If an insurer, employer, or other person obtains your written consent to receive research information, then the researchers may not use the Certificate to withhold that information.

How will your privacy be protected? All of the search data you provide will be stored confidentially. This means that there will be no way for anybody, except the research team, to ever link your data or the results of the study to your identity.

We will collect your name and email address for communication for this study. Your name and email address and the assigned Google study account and a participant ID will be linked and stored in a separated document. All your questionnaire responses and diary entries will be labeled by your assigned Google study account name. We will only share the password to the study account with the research team members. All these Google study account will be deleted after the study is complete. The linkage document will be deleted as well after this study is complete. During the retrospective interview, your name will be replaced by your study account ID. You webcam/face image won't be recorded. Only the audio and shared screen will be recorded. The recording data will be stored in the research member's UNC OneDrive account and on password-protected computers used by the researchers.

Participants will not be identified by name in any report or publication about this study. We may use de-identified data and/or specimens from this study in future research without additional consent.

What if you want to stop before your part in the study is complete? You can withdraw from this study at any time, or skip any question for any reason when answering questions in the questionnaire and during the interview. There will be no penalty at all. The investigators also have the right to stop your participation if you have an unexpected reaction, have failed to follow instructions, etc.

Will you receive anything for being in this study? Will it cost anything? You will be receiving an Amazon Gift Card of $80 for completing the diary study. Any payment provided for participation in this study may be subject to applicable tax withholding obligations. Participants who provide high-quality data will receive an extra $40 bonus. Here high-quality data is defined as your responses to the questionnaires and diary forms include rich details of your information search process and genuine thoughts about your cognitive and other task behaviors. Participants who participated in the additional interview will receive an additional $20 gift card for their participation. If you are not invited to the interview, you will receive the Amazon Gift Card code via email sent by the research team with 1-2 days after you complete your study. If you attend the interview, you will receive the email with your Amazon Gift Card Code right after you complete the interview. In any case, if you decide to withdraw from the study before you complete the task or failed to provide the required questionnaires/diaries/search history data, you will be paid based on the number of days you worked on the task for $15 per day (no more than $45).

What if you have questions about this study? You have the right to ask, and have answered, any questions you may have about this research. Contact the principal investigator listed above with any questions, complaints, or concerns you may have.

What if you have questions about your rights as a research participant? All research on human volunteers is reviewed by a committee that works to protect your rights and welfare. If you have questions or concerns, or if you would like to obtain information or offer input, please contact the Institutional Review Board at 919-966-3113 or by email to IRB_subjects@unc.edu.

Participants will type their names into a box by the end of the consent form to indicate their acknowledgment of the consent.

APPENDIX E

Pre-task Questionnaire

Thank you for participating in our study!

Below are some important definitions that can help you answer the questions during the diary study:

A search query is a specific term or phrase that you entered into a search engine.

A search session is a continuous period of time during which you searched online. A search session may include multiple search queries.

A multi-session search is a search process that involved multiple search sessions (e.g., at different times in a project, or across different days).

E.1 Task description

Q1: During the recruiting email, you provided a task for attending this study, please describe what your task is about in detail, make sure that:

(1) You haven't started the task yet,

(2) The task requires you to search for information on the Web, and

(3) You plan to work across multiple days to complete the task.

NOTE: For all the questions, please provide full and complete responses. However, please do not include any personally identifying information (e.g., names, addresses, etc.) in your responses.

Q2: Please describe how you will work to complete this task. Your answer will help us understand the process of the task.

Q3: Which of the following options can best describe why you are undertaking this task?

Q4: Will you require anyone's assistance to complete the task?

Q5: How much time do you expect to spend to complete this task?

Q6: What are your goals for this task? Please list the main goal and any sub-goals.

Q7: What kinds of information do you want to find for this task? If you think you will need different types of information, please list as many as you can think of.

Q8: Please indicate your level of agreement with the following statements about this task (1 for not familiar at all, 7 for extremely familiar).

- How much do you already know about the task topic?
- How much do you already know about the skills to complete this task?
- How much do you already know about your solution to this task?

Q9: Please indicate your level of agreement with the following statements about this task (1 for strongly disagree, 5 for strongly agree):

- I think it will be difficult to complete this task.
- I think it will be difficult to search for information for this task.
- I think it will be difficult to integrate the information I find to complete this task.
- I think it will be difficult to decide when I have enough information to complete this task.

E.2 Demographic information questions

Q1: Your participant ID (your assigned Google study account No.):

Q2: What is your age in years?

Q3: What is your gender?

Q4: What is the highest degree or level of school you have completed? (If you're currently enrolled in school, please indicate the highest degree you have received.)

Q5: If you are employed, what is your current occupation?

APPENDIX F

Pre-Session Questionnaire

Please fill out this questionnaire every time before you start a search session (A search session is a continuous period of time when you are searching online. A search session may include multiple search queries.)

If you conducted multiple search sessions within one day, please fill this questionnaire for each search session separately.

Q1: Your assigned Google study account No.:

Q2: What activities did you do related to your task before you started this search session?

Q3: What are your goals for this search session? Why did you start this search session?

Q4: From the list below, select **all** that apply to your goals for this session.

- I need to find some specific factual information.
- I need to learn about important concepts and their definitions related to my task.
- I need to find instructions and follow them to solve some problems for my task.
- I need to find information that can help me differentiate between related ideas.
- I need to find and evaluate different alternatives to make informed decisions for my task.
- I need to find some examples and inspiration to generate my own ideas.
- Other (please specify)

Q5: From the same list, select the one **best** reason that describes your **main goal** for this session.

- I need to find some specific factual information.
- I need to learn about important concepts and their definitions related to my task.
- I need to find instructions and follow them to solve some problems for my task.
- I need to find information that can help me differentiate between related ideas.
- I need to find and evaluate different alternatives to make informed decisions for my task.
- I need to find some examples and inspiration to generate my own ideas.

APPENDIX G

Post-Session Questionnaire

Please fill this questionnaire every time when you complete a search session (A search session is a continuous period of time during which you searched online. A search session may include multiple search queries.)

This questionnaire is important for us to capture your information search behaviors. To avoid problems caused by misremembering, please fill this questionnaire as soon as you complete your search session.

If you conducted multiple search sessions within one day, please fill this questionnaire for each search session separately.

Q1: Your participant ID (your assigned Google account No.):

Q2: Please indicate your starting time and stopping time for this search session.

Q3: What did you want to achieve in this session?

Q4: Why did you start this search session? Please list as many reasons as you can remember.

Q5: In many cases, people's perception of the task or their information needs for the task change over time. Based on your own task progress, which one of the following descriptions best matches the most important reason that you started this search session?

- My task requirements were not clear.
- My task had subconcepts (or subtopics) that I needed to understand.
- Information I found previously did not work.
- I needed to re-find information I had seen before.
- I wanted to continue a previous search that ended without finding satisfactory information.
- I wanted to find updated (e.g., the most recent) information related to the task.
- I did not have a specific goal about the task, but thought that looking for the information might be useful in the future.

Q6: Please describe your search session process in detail (e.g., what information you were looking for, what you found, how it can help with your task.)

Q7: How is the search direction of this session different from your previous search sessions?

Q8: During this search session, you may have encountered information relevant to your task (helpful or not helpful). In the following questions please provide up to three examples of the information you found in this session and answer the following questions:

- Describe the information that you found
- URL
- How do you think the information helpful or not for your task?

Q9: Based on your search experience during this past session, indicate your level of agreement with the following statements:

- It was difficult to search for information in this session.
- It was difficult to understand the information I found in this session.
- It was difficult to integrate different information I found in this session.
- It was difficult to decide when I had enough information to complete this session.
- Overall, this search session was difficult.

Q10: Besides searching for information online, what other activities did you do for your task during this search session? Please describe in detail. For example, you may have tried to apply some of the information to your task, or you may have taken notes or discussed the project with a collaborator as you were searching. If you did not do anything except search online, please explain why.

Q11: What are the reasons that made you stop this search session? Please list as many reasons as apply.

Q12: Below are some reasons that are directly related to your task that may have led to you stopping your searching session. Select **the most important task-related reason** that made you stop this search session:

- I needed to consult other sources (e.g., another person, organizations) for more information about the task.
- I could not find the information I was looking for.
- I needed to process/think about the gathered information before taking next step.

- I needed to validate the found information by applying it to my task.
- I found all the information I needed.
- Task deadline was approaching.
- I could not complete the task.
- Other reasons (Please specify)

Q13: Describe any difficulties that you had searching for information during this session. What did you do to address the difficulties?

Q14: What kind of assistance did you wish the search systems could provide to help you complete your task in this session? What kind of extra help do you wish the search engine could provide during this search session?

Q15: What do you plan to do next for this task after this search session?

Now you can close this window or switch back to your own Google account.

Please remember to log into the study account next time when you need to search again for this task.

APPENDIX H

Daily Review Diary

This questionnaire is a daily review of your everyday process on the task. It focuses on your actions about how you used the information after the search session to complete your task.

This questionnaire needs to be filled out each day when you worked on the task. If you did not work or search for information for this task throughout the whole day, you can skip this questionnaire.

Q1: Your participant ID (your assigned Google study account No.):

Q2: Have you filled the pre-/post- search session questionnaires for all your search sessions today?

Q3: Please describe your today's work process for the task (Please include both your online search and other task activities):

Q4: Besides searching for information online, which of the following sources did you use to look for information related to this task today? (Select all that apply)

Q5: What kind of information you were looking for from the information source(s) you selected above? And why?

Q6: How did the information you found online today help you with your task? If not, please explain why.

Q7: What were the difficulties or problems for you to use the found information today?

Q8: What kind of help do you wish the computer or the search systems to provide for supporting yourr task activities today?

Q9: By now, are you finished your entire task or do you still plan to do more work on it?

(if yes for Q9) **Q10:** Based on your information searching experience of the task, please indicate your level of agreement with the following statements:

- It was difficult to search for information to complete this task.
- It was difficult to understand the information I found to complete this task.
- It was difficult to integrate different information I found to complete this task.
- It was difficult to decide when I have enough information to complete this task.
- Overall, this task was difficult.

(if yes for Q9) **Q11:** Please provide three possible time slots that you are available for an interview within next week. The interview will take about 45 minutes.

Q12: Next, please open and scan through your study account's search history today, and remove any history links or bookmarks that are not relevant to the task or anything that you don't want to share with the research team.

APPENDIX I

Interview Invitation

To: [Participant's email address]

Cc: yuanli@live.unc.edu

Subject: Invitation: Cross-session search diary study interview

—–**Message Text**——–

Thank you for completing the diary study. We are glad to invite you to participate in the retrospective interview about your search experience for the task you completed.

The interview will last about 45 minutes, and you will receive an extra 20 dollars for the interview participation.

If you are interested, please provide us three time slots that you are available for a Zoom meeting next week if you are interested in participating in the interview.

If you decide not to participate in the interview, please reply to this email at your most convenient time, and we will send you the payment of the diary study once we receive your confirmation.

This interview invitation will be expired after three days [by DD/MM/YYYY]; if we did not receive any information from you, we will cancel the interview invitation and send you your final payment via email.

Many thanks,

Yuan Li

School of Information and Library Science

University of North Carolina at Chapel Hill

APPENDIX J

Interview Confirmation

To: [Participant's email address]

Cc: yuanli@live.unc.edu

Subject: Confirmation: Cross-session search diary study interview

—–**Message Text**——–

Thank you for your confirmation of participating in the interview. We have reserved this interview session for you on DD/MM/YYYY at TIME.

[Interview Zoom information]

[copy information from Zoom invitation similar to below]

[Name] is inviting you to a scheduled Zoom meeting.

Topic: Cross-session search diary study retrospective interview

Time: [Time]

Join Zoom Meeting

[URL]

Meeting ID: [ID #]

Password: [Password]

Please do not share the Zoom information or make it available on any public-facing website or calendar.

If you have any questions or encounter any difficulties connecting, please contact the researcher at yuanli@live.unc.edu or 919-360-6310.

Many thanks,

Yuan Li

School of Information and Library Science

University of North Carolina at Chapel Hill

APPENDIX K

Interview Reminder

To: [Participant's email address]

Cc: yuanli@live.unc.edu

Subject: Reminder: Cross-session search diary study interview

—–**Message Text**——–

This is a reminder that you will participate in the diary study interview. We have reserved this Zoom interview session for you on DD/MM/YYYY at TIME.

This session has been reserved specifically for you. If you are unable to make it to the study session at this time, please let us know as soon as possible.

[Interview Zoom information]

[copy information from Zoom invitation similar to below]

[Name] is inviting you to a scheduled Zoom meeting.

Topic: Cross-session search diary study retrospective interview

Time: [Time]

Join Zoom Meeting

[URL]

Meeting ID: [ID #]

Password: [Password]

Please do not share the Zoom information or make it available on any public-facing website or calendar.

If you have any questions or encounter any difficulties connecting, please contact the researcher at yuanli@live.unc.edu or 919-360-6310.

Many thanks,

Yuan Li

School of Information and Library Science

University of North Carolina at Chapel Hill

APPENDIX L

Interview Cancellation

To: [Participant's email address]

Cc: yuanli@live.unc.edu

Subject: Cancellation: Cross-session search diary study interview

——**Message Text**———

Sorry we did not receive your confirmation about participating in the interview. If you are still interested to participate in it, please reply this email within next 24 hours with three time slots for a 45-minute Zoom interview.

If we did not receive your reply, your interview session will be cancelled. And we will send you your study payment with next 2 business days.

This interview is purely voluntary. Your study payment for the parts you already completed won't be affected if you decide not to participate the interview.

If you have any questions or encounter any difficulties connecting, please contact the researcher at yuanli@live.unc.edu or 919-360-6310.

Many thanks,

Yuan Li

School of Information and Library Science

University of North Carolina at Chapel Hill

APPENDIX M

Daily Review Questionnaire Reminder

To: [Participant's email address]

Cc: yuanli@live.unc.edu

Subject: Reminder: Complete questionnaires and daily work diary for cross-session search task

—–**Message Text**——–

Hi [Particiipant's name],

This email is to remind you that you are currently participating in our diary study. If you plan to work or have already worked on your project today, please remember to complete the pre- and post-session questionnaires for each search session you completed today, and also remember to keep a daily work review diary for your task process. You can log into the assigned Google study account to find the links of these questionnaires.

Remember to review your search history today and make sure to remove links that are irrelevant to the task or any other links you don't want to share with the researcher.

If you have any questions or encounter any difficulties connecting, please contact the researcher at yuanli@live.unc.edu or 919-360-6310.

Many thanks,

Yuan Li

School of Information and Library Science

University of North Carolina at Chapel Hill

APPENDIX N

Retrospective Interview

Thank you for participating in this interview. Today, I will ask you some questions about your experience when you work on the task you completed during your cross-session search.

To help you remember some details of your task process, we create an Excel spreadsheet with your questionnaire responses, diary recordings, and search history on the Google Study account. Next, I am going to ask you some questions about your task process for each session.

During the interview, you can skip any questions if you feel uncomfortable answering.

Do you have any questions?

I will record our interview conversation. It only records our voice but not our webcam image.

First, you can now change your Zoom display name to your study number,

Change Zoom Name to NNN

Now I will start the recording. You can see the blinking red dot on the left up corner on the Zoom app, indicating the recording is on-going.

Start Zoom recording

—–Interview Questions—–

Q1: First, could you briefly describe the whole process of your project, and what are the final outcomes of the task?

Q2: Here is a excel spread sheets about your search sessions queries, thinking about the different stages as you work on this task, how do you describe the relationships/connections between these sessions?

Ask the following questions (Q3-Q7) for each session

Q3: Based on your search history for NN session, what made you search focus on XXX during this session?

Q4: During you NN session, were there any unexpected information that you did not think of before you start this task? Can you describe how did you find it? And how did the session help your task? [Show the spreadsheet and point to the queries and useful info logged]

Q5: During the NN session, what were the most difficult part you experienced? How did the difficulties affect your task progress? How did you overcome the difficulties?

Q6: During the NN session, were there any unexpected challenges you encountered? What were they?

Q7: In your NN daily diaries, you mentioned the following extra assistance that you wish the system can provide help with, can you explain more about them (Point participants to the column about all the functions they wish system has).

Q8: After review all your search sessions, now thinking about the entire task/project process, how do you describe your task stages for this project, which stage does each session belongs to?

[Q9] (optional): comparing the recorded search session diaries with search log data, in cases when people searched but did not recorded a session questionnaire, show them the search queries and links, ask them to recall why they started that search and how the stopped.

Q10: Do you have any questions for me?

Now I am going to stop the recording.

————-End of interview————-

That's all my questions today. And thank you so much for your answer. Do you have any questions you want to ask?

Next, I am going to **send you your Amazon gift card by email**.

Now I am going to send you **a link for the receipt information**, please open and sign the receipt with your name and today's date.

Thanks!

APPENDIX O

Incentive Email

To: [Participant's email address]

Cc: yuanli@live.unc.edu

Subject: Reminder: Diary study incentive and final instruction

—–**Message Text**——–

Hi [Particiipant's name],

Thank you for participating in the study. Below is the information for your Amazon Gift Card. You can apply the gift card directly to your account.

Since your participation in this diary study is finished now, we will close the Google Study Account. Therefore, you will no longer be able to access it. In this email, we attached an exported version of your search history (including any URLs and bookmarks you saved) if you want to use it later for your task.

You can also delete the Google study account from your computer. To delete the account:

(1) open your Google Chrome browser,

(2) click your profile circle on the right corner to open the drop-down list of your account,

(3) click the setting gear icon next to the "other profiles",

(4) select the Google study profile account on the newly opened page,

(5) click the three-dot menu icon on the top right corner of the Google study account and click "delete",

(6) click "Yes" to confirm that you want to delete all the data associated with this account.

If you have any questions or encounter any difficulties connecting, please contact the researcher at yuanli@live.unc.edu or 919-360-6310.

Many thanks,

Yuan Li

School of Information and Library Science

University of North Carolina at Chapel Hill

APPENDIX P

Debriefing Form

The consent form is created using the UNC IRB template and attached as an external Pdf document starting from next page.

IRB Study # 21-1428
Title of Study: How and Why do people search across sessions for complex tasks: A diary study
Principal Investigator: Yuan Li
Principal Investigator Phone number: 919-360-6310
Principal Investigator Email Address: yuanli@live.unc.edu

Thank you for your participation in this research study. For this study, it was important that we withhold the criteria of final payment about some aspects of your participation. Now that your participation is completed, I will describe the criteria to you, why it was important, answer any of your questions and provide you with the opportunity to make a decision on whether you would like to have your data included in this study.

What you should know about this study
All participants that complete the diary study with the required three sets of data: (1) at least three search sessions' search histories, (2) corresponding pre-/post-session questionnaires for the threes sessions mentioned earlier, (3) at least two daily diary reviews, will receive the $40 bonus plus the basic $80 for their participation. We said that $40 bonus was for participants who provide higher-quality data with rich details and genuine thoughts to encourage people to provide details about their cross-session information process and motivate them to keep recording as many diary/questionnaires as possible. This is very important for us to gain knowledge about the real-life cross-session process. As long as you complete the minimum requirements of the three sets of data described above, you will receive $120 as your compensation that including the $40 bonus.

If you did not complete the study for the following reasons: (1) you withdrew from the study before you complete it; or (2) you did not provide **ALL** the required three sets of data as described above, you will be paid based on the days that you worked the study according to the dates on the records from your completed questionnaires and diaries for $15 per day.

Right to withdraw data
Although the collected responses and diary data are only about the task you described but not include any of your identifying information, you may choose to withdraw the data you provided prior to debriefing without penalty.

If you have questions
Please ask any questions you have (now or later) by sending emails to Yuan Li (yuanli@live.unc.edu) or at the contact information provided above. If you have any questions or concerns regarding your rights as a research participant in this study, you may contact the Institutional Review Board at 919-966-3113 or by email to IRB_subjects@unc.edu.

Add any of the following, if applicable:
Please do not disclose research procedures and/or purpose to anyone who might participate in this study in the future as this could affect the results of the study.

Final Report: If you would like to receive a copy of the final report of this study when it is completed, please feel free to contact the researcher. Below are some more related research that are published before by the same research team:

Li, Y., Ward, A. R., & Capra, R. (2021, March). An Analysis of Information Types and Cognitive Activities Involved in Cross-session Search. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* (pp. 313-317).

Li, Y., & Capra, R. (2020). Exploring factors affecting renewal and stopping reasons in cross-session search. *Proceedings of the Association for Information Science and Technology*, *57*(1), e267.
Li, Y., Capra, R., & Zhang, Y. (2020, March). Everyday Cross-session Search: How and Why Do People Search Across Multiple Sessions?. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval* (pp. 163-172).

REFERENCES

Agichtein, E., White, R. W., Dumais, S. T., & Bennet, P. N. (2012). Search, interrupted: Understanding and predicting search task continuation. In (pp. 315–324). New York, NY, USA: Association for Computing Machinery.

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of bloom's taxonomy of educational objectives*. Boston, MA: Allyn & Bacon.

Annett, J. (2003). Hierarchical task analysis. *Handbook of cognitive task design, In E. Hollnagel (Ed.)*, *2*, 17–35.

Arguello, J. (2014). Predicting search task difficulty. In M. de Rijke et al. (Eds.), *Advances in information retrieval* (pp. 88–99). Cham: Springer International Publishing.

Arguello, J., & Choi, B. (2019). The effects of working memory, perceptual speed, and inhibition in aggregated search. *ACM Transactions on Information Systems*, *37*(3), 1–29.

Arguello, J., Wu, W.-C., Kelly, D., & Edwards, A. (2012). Task complexity, vertical display and user interaction in aggregated search. In *Proceedings of the 35th international acm sigir conference on research and development in information retrieval* (pp. 435–444).

Aula, A., Khan, R. M., & Guan, Z. (2010). How does search behavior change as search becomes more difficult? In (pp. 35–44). New York, NY, USA: Association for Computing Machinery.

Azevedo, R., Cromley, J. G., & Seibert, D. (2004, 7). Does adaptive scaffolding facilitate students' ability to regulate their learning with hypermedia? *Contemporary Educational Psychology*, *29*, 344-370.

Backhausen, D., Klas, C.-P., & Hemmje, M. (2015). Adaptive information retrieval support for multi-session information tasks. In *International conference on theory and practice of digital libraries* (pp. 317–320). Springer, Cham.

Bates, M. J. (1984). The fallacy of the perfect thirty-item online search. *RQ*, *1*(24), 43–50.

Bates, M. J., Wilde, D. N., & Siegfried, S. (1993). An analysis of search terminology used by humanities scholars: the getty online searching project report number 1. *The Library Quarterly*, *63*(1), 1–39.

Bell, D. J., & Ruthven, I. (2004). Searcher's assessments of task complexity for web searching. In *European conference on information retrieval* (pp. 57–71). Berlin, Heidelberg: Springer.

Berg, B. (2001). *Qualitative research methods for the social sciences*. Boston, MA, USA: Pearson Boston.

Berryman, J. (2006). What defines" enough" information? how policy workers make judgements and decisions during information seeking: Preliminary results from an exploratory study. *Information Research: An International Electronic Journal*, *11*(4), n4.

Bharat, K. (2000). Searchpad: Explicit capture of search context to support web search. *Computer Networks*, *33*(1-6), 493–501.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual* *review of psychology*, *54*(1), 579–616.

Bouidghaghen, O., Tamine, L., & Boughanem, M. (2011). Context-aware user's interests for personalizing mobile search. *IEEE International Conference on Mobile Data Management*, *1*(1), 129–134. doi: 10.1109/MDM.2011.51

Brennan, K., Kelly, D., & Arguello, J. (2014). The effect of cognitive abilities on information search for tasks of varying levels of complexity. *IIiX '14: Proceedings of the 5th Information Interaction in Context Symposium*.

Broder, A. (2002). A taxonomy of web search. In *Acm sigir forum* (Vol. 36, pp. 3–10). ACM New York, NY, USA.

Brown, C. (2007). *Cognitive psychology.* SAGE Publications Ltd.

Byström, K. (n.d.). Information and information sources in tasks of varying complexity. *Journal of the American Society for Information Science and Technology*, *53*, 581-591.

Byström, K., & Hansen, P. (2005). Conceptual framework for tasks in information studies. *Journal of the American Society for Information science and Technology*, *56*(10), 1050–1061.

Byström, K., & Järvelin, K. (1995). Task complexity affects information seeking and use. *Information processing & management*, *31*(2), 191–213.

Campbell, D. J. (1988). Task complexity: A review and analysis. *Academy of management review*, *13*(1), 40–52.

Capra, R. (2006). *An investigation of finding and refinding information on the web* (Unpublished doctoral dissertation). Virginia Tech.

Capra, R., Arguello, J., Crescenzi, A., & Vardell, E. (2015). Differences in the use of search assistance for tasks of varying complexity. In *Proceedings of the 38th international acm sigir conference on research and development in information retrieval* (pp. 23–32).

Capra, R., Arguello, J., O'Brien, H., Li, Y., & Choi, B. (2018). The effects of manipulating task determinability on search behaviors and outcomes. In *The 41st international acm sigir conference on research & development in information retrieval* (pp. 445–454).

Capra, R., Arguello, J., & Zhang, Y. (2017). The Effects of Search Task Determinability on Search Behavior. *ECIR2017*, 108–121.

Capra, R., Marchionini, G., Velasco-Martin, J., & Muller, K. (2010). Tools-at-hand and learning in multi-session, collaborative search. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 951–960).

Carter, S., & Mankoff, J. (2005). When participants do the capturing: the role of media in diary studies. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 899–908).

Charmaz, K. (2000). Grounded theory: Objectivist and constructivist methods. *Handbook of qualitative research*, *2*, 509–535.

Chen, L., & Qi, L. (2010). A diary study of understanding contextual information needs during leisure traveling. In *Proceedings of the third symposium on information interaction in context* (pp. 265–270). New York, NY, USA: Association for Computing Machinery.

Choi, B., & Arguello, J. (n.d.). A qualitative analysis of the effects of task complexity on the functional role of information. *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, 328-332.

Choi, B., Arguello, J., & Capra, R. (2023). Understanding procedural search tasks 'in the wild'. *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval*, 24-33.

Choi, B., Capra, R., & Arguello, J. (2019). The effects of working memory during search tasks of varying complexity. In *Proceedings of the 2019 conference on human information interaction and retrieval* (pp. 261–265).

Church, K., Smyth, B., Bradley, K., & Cotter, P. (2008). A large scale study of european mobile search behaviour. In *Proceedings of the 10th international conference on human computer interaction with mobile devices and services* (pp. 13–22).

Cooper, W. S. (1968). Expected search length: A single measure of retrieval effectiveness based on the weak ordering action of retrieval systems. *American documentation*, *19*(1), 30–41.

Crabtree, B., & Miller, W. (1992). A template approach to text analysis: Developing and using codebooks. *Doing qualitative research*, 93–109.

Crandall, B., Klein, G., Klein, G. A., & Hoffman, R. R. (2006). *Working minds: A practitioner's guide to cognitive task analysis.* MIT Press.

Creswell, J. W., & Poth, C. N. (2007). *Qualitative inquiry and research design: Choosing among five approaches.* Thousand Oaks, CA: Sage publications.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience.* NY: Harper Perennial.

Czerwinski, M., Horvitz, E., & Wilhite, S. (2004). A diary study of task switching and interruptions. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 175–182).

Dabbish, L., Mark, G., & González, V. M. (2011). Why do i keep interrupting myself? environment, habit and self-interruption. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 3127–3130).

Darmoul, S., Ahmad, A., Ghaleb, M., & Alkahtani, M. (2015). Interruption management in human multitasking environments. In (Vol. 28, p. 1179-1185).

Dervin, B. (1983). An overview of sense-making research: concepts, methods and results to date..

Diamond, A. (2013). Executive functions. *Annual review of psychology*, 135-168.

Donato, D., Bonchi, F., Chi, T., & Maarek, Y. (2010a). Do you want to take notes?: identifying research missions in yahoo! search pad. In *Proc. www* (pp. 321–330).

Donato, D., Bonchi, F., Chi, T., & Maarek, Y. (2010b). Do you want to take notes?: identifying research missions in yahoo! search pad. In *Proc. www* (pp. 321–330).

Drabenstott, K. M., & Weller, M. S. (2017). Failure analysis of subject searches in a test of a new design for subject access to online catalogs. *Journal of the American Society for Information Science*, *54*, 436-445.

Dupuis, M., Endicott-Popovsky, B., & Crossler, R. (2013). An analysis of the use of amazon's mechanical turk for survey research in the cloud. In *Iccsm2013-proceedings of the international conference on cloud security management: Iccsm* (p. 10).

Exploring social context in mobile information behavior. (2014). In *Proceedings of the american society for information science and technology* (Vol. 51, pp. 1–10). Seattle, USA: Wiley Online Library.

Foster, A. (2004). A nonlinear model of information-seeking behavior. *Journal of the American Society for Information Science and Technology*, *55*, 228-237.

Golovchinsky, G., Diriye, A., & Dunnigan, T. (2012). The future is in the past: designing for exploratory search. In *Proceedings of the 4th information interaction in context symposium* (pp. 52–61).

Gomes, S., & Hoeber, O. (2021). Supporting cross-session cross-device search in an academic digital library. In *Proceedings of the 2021 conference on human information interaction and retrieval* (pp. 337–341).

Gunthert, K. C., & Wenze, S. J. (2012). Daily diary methods. *Handbook of research methods for studying daily life*, 144-159.

Han, S. (2018). *Understanding, modeling and supporting cross-device web search* (Unpublished doctoral dissertation). University of Pittsburgh.

Han, S., He, D., & Chi, Y. (2017). Understanding and modeling behavior patterns in cross-device web search. *Proceedings of the Association for Information Science and Technology*, *54*(1), 150–158.

Han, S., Yue, Z., & He, D. (2015, April). Understanding and supporting cross-device web search for exploratory tasks with mobile touch interactions. *ACM Trans. Inf. Syst.*, *33*(4), 16:1–16:34. Retrieved from \http://doi.acm.org/10.1145/2738036 doi: 10.1145/2738036

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative health research*, *15*(9), 1277–1288.

Huang, K., Zhang, L., Qin, T., & Guo, J. (2015). A study on the adjustment model of university library users' queries based on opac log-taking the library of beijing normal university as an example. *Library and Information Service*, *59*(23), 51-58.

Huang, M. (1992). Pausing behavior of end-users in online searching (unpublished doctoral dissertation). *College Park, MD, University of Maryland.*

Ingwersen, P. (2000). Users in context. In *European summer school on information retrieval* (pp. 157–178).

Ingwersen, P., & Järvelin, K. (2005). The integrated IS & R research framework. In *The turn:*

Integration of information seeking and retrieval in context (pp. 259–311).

Jansen, B. J., Booth, D., & Smith, B. (2009). Using the taxonomy of cognitive learning to model online searching. *Information Processing & Management*, *45*(6), 643–663.

Jansen, B. J., Koshman, S., & Spink, A. (2005). Repeat visits to vivisimo. com: Implications for successive web searching. *Proceedings of the American Society for Information Science and Technology*, *42*(1).

Jansen, B. J., Spink, A., Blakely, C., & Koshman, S. (2007). Defining a session on web search engines. *Journal of the American Society for Information Science and Technology*, *58*(6), 862–871.

Järvelin, K., Vakkari, P., Arvola, P., Baskaya, F., Järvelin, A., Kekäläinen, J., ... Savolainen, R. (2015). Task-based information interaction evaluation: The viewpoint of program theory. *ACM Transactions on Information Systems (TOIS)*, *33*(1), 1–30.

Jhaveri, N. (2004). Intermediate and post-session web page revisitation techniques and tools. *Master Thesis*.

Jhaveri, N., & Räihä, K.-J. (2005). The advantages of a cross-session web workspace. In *Chi'05 extended abstracts on human factors in computing systems* (pp. 1949–1952).

Jones, R., & Klinkner, K. L. (2008). Beyond the session timeout: automatic hierarchical segmentation of search topics in query logs. In *Proceedings of the 17th acm conference on information and knowledge management* (pp. 699–708).

Jones, W., Bruce, H., Foxley, A., & Munat, C. F. (2006). Planning personal projects and organizing personal information. *Proceedings of the American Society for Information Science and Technology*, *43*(1), 1–24.

Kellar, M., Watters, C., & Shepherd, M. (2006). A goal-based classification of web information tasks. *Proceedings of the American Society for Information Science and Technology*, *43*(1), 1–22.

Kellar, M., Watters, C., & Shepherd, M. (2007). A field study characterizing web-based information-seeking tasks. *Journal of the American Society for information science and technology*, *58*(7), 999–1018.

Kelly, D., Arguello, J., Edwards, A., & Wu, W.-C. (2015). Development and evaluation of search tasks for iir experiments using a cognitive complexity framework. In *Proceedings of the 2015 international conference on the theory of information retrieval* (pp. 101–110).

Kieras, D. E. (1988). Towards a practical goms model methodology for user interface design. In *Handbook of human-computer interaction* (pp. 135–157). Elsevier.

Kim, J. (2006a). *Task as a predictable indicator for information seeking behavior on the web.*

Kim, J. (2006b). Task difficulty as a predictor and indicator of web searching interaction. In *Chi'06 extended abstracts on human factors in computing systems* (pp. 959–964).

Kim, K.-S. (2001). Information-seeking on the web: Effects of user and task variables. *Library & Information Science Research*, *23*(3), 233–255.

Kotov, A., Bennett, P. N., White, R. W., Dumais, S. T., & Teevan, J. (2011). Modeling and analysis of cross-session search tasks. In *Proceedings of the 34th international acm sigir conference on research and development in information retrieval* (pp. 5–14).

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.

Li, Y., & Belkin, N. J. (2008). A faceted approach to conceptualizing tasks in information seeking. *Information Processing and Management*, *44*(6), 1822–1837. doi: 10.1016/j.ipm.2008.07.005

Li, Y., Capra, R., & Zhang, Y. (2020a). Everyday cross-session search: How and why do people search across multiple sessions? In *Proceedings of the 2020 conference on human information interaction and retrieval* (p. 163-172). New York, NY, USA: Association for Computing Machinery.

Li, Y., Capra, R., & Zhang, Y. (2020b). Everyday cross-session search: how and why do people search across multiple sessions? *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, 163-172.

Li, Y., Ward, A. R., & Capra, R. (2021a). An analysis of information types and cognitive activities involved in cross-session search. In *Proceedings of the 2021 conference on human information interaction and retrieval* (pp. 313–317). New York, NY, USA: Association for Computing Machinery.

Li, Y., Ward, A. R., & Capra, R. (2021b). An analysis of information types and cognitive activities involved in cross-session search. *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*, 313-317.

Lin, S.-J. (2001). Modeling and supporting multiple information seeking episodes over the web. *PhD Dissertation*.

Lin, S.-J., & Belkin, N. (2005). Validation of a model of information seeking over multiple search sessions. *Journal of the American Society for Information Science and Technology*, *56*(4), 393–415.

Lin, S.-J., & Belkin, N. J. (2000). Modeling multiple information seeking episodes. In *Proceedings of the asis annual meeting* (Vol. 37, pp. 133–47).

Liu, C., & Song, X. (2018). How do Information Source Selection Strategies Influence Users' Learning Outcomes'. *Proceedings of the 2018 Conference on Human Information Interaction&Retrieval - CHIIR '18*.

Liu, J., & Belkin, N. J. (2010). Personalizing information retrieval for multi-session tasks: The roles of task stage and task type. In *Proceedings of the 33rd international acm sigir conference on research and development in information retrieval* (pp. 26–33). New York, NY, USA: ACM.

Liu, J., Gwizdka, J., Liu, C., & Belkin, N. J. (2010). Predicting task difficulty for different task types. *Proceedings of the American Society for Information Science and Technology*, *47*(1), 1–10.

Liu, J., Liu, C., Cole, M., Belkin, N. J., & Zhang, X. (2012). Exploring and predicting search task difficulty. In *Proceedings of the 21st acm international conference on information and knowledge management* (pp. 1313–1322). New York, NY, USA: Association for Computing Machinery.

Liu, J., Liu, C., Yuan, X., & Belkin, N. J. (2011). Understanding searchers' perception of task difficulty: Relationships with task type. *Proceedings of the American Society for Information Science and Technology*, *48*(1), 1–10.

Liu, P., & Li, Z. (2012). Task complexity: A review and conceptualization framework. *International Journal of Industrial Ergonomics*, *42*(6), 553–568.

MacKay, B., & Watters, C. (2008a). Exploring multi-session web tasks. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 1187–1196). New York, NY, USA: Association for Computing Machinery.

MacKay, B., & Watters, C. (2008b). Understanding and supporting multi-session web tasks. *Proceedings of the American Society for Information Science and Technology*, *45*(1), 1–13.

MacKay, B., & Watters, C. (2009). Providing support for multi-session web tasks. *Proceedings of the American Society for Information Science and Technology*, *46*(1), 1–26.

MacKay, B., & Watters, C. (2012). An examination of multisession web tasks. *Journal of the American Society for Information Science and Technology*, *63*(6), 1183–1197.

March, J. G. (1994). *Primer on decision making: How decisions happen.* Simon and Schuster.

Marchionini, G. (1997). *Information seeking in electronic environments* (No. 9). Cambridge university press.

Marchionini, G. (2006). Exploratory search: from finding to understanding. *Communications of the ACM*, *49*(4), 41–46.

McFarlane, D. (1998). Interruption of people in human-computer interaction. *A Dissertation Submitted to the Faculty of The School of Engineering and Applied Science of The George Washington University*, 308.

Mehrotra, R., & Yilmaz, E. (2017). Extracting hierarchies of search tasks & subtasks via a bayesian nonparametric approach. In *Proceedings of the 40th international acm sigir conference on research and development in information retrieval* (pp. 285–294).

Montanez, G. D., White, R. W., & Huang, X. (2014). Cross-device search. In *Proc. cikm* (pp. 1669–1678). ACM.

Morris, D., Ringel Morris, M., & Venolia, G. (2008a). Searchbar: A search-centric web history for task resumption and information re-finding. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 1207–1216). New York, NY, USA: ACM.

Morris, D., Ringel Morris, M., & Venolia, G. (2008b). Searchbar: A search-centric web history for task resumption and information re-finding. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 1207–1216). New York, NY, USA: ACM.

Morse, J. M. (1994). Designing funded qualitative research. *Handbook of qualitative research (2nd Ed)*.

Nickles, K. R., Curley, S. P., & Benson, P. G. (1995). *Judgment-based and reasonging-based stopping rules in decision making under uncertainty* (Vol. 7285). University of Minnesota.

Obendorf, H., Weinreich, H., Herder, E., & Mayer, M. (2007). Web page revisitation revisited: implications of a long-term click-stream study of browser usage. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 597–606).

O'Brien, H. L., Arguello, J., & Capra, R. (2020). An empirical study of interest, task complexity, and search behaviour on user engagement. *Information Processing and Management*, *57*(3).

Oulasvirta, A., & Saariluoma, P. (2006, oct). Surviving task interruptions: Investigating the implications of long-term working memory theory. *International Journal of Human Computer Studies*, *64*(10), 941–961. doi: 10.1016/j.ijhcs.2006.04.006

Pardi, G., Kammerer, Y., & Gerjets, P. (2019). Search and justification behavior during multimedia web search for procedural knowledge. *Companion Publication of the 10th ACM Conference on Web Science*, 17-20.

Pickard, A. J. (2013). *Research methods in information*. Facet publishing.

Radford, M. L. (2006). The critical incident technique and the qualitative evaluation of the connecting libraries and schools project. *Library Trends*, *55*(1), 46–64.

Reid, J. (2000). A task-oriented non-interactive evaluation methodology for information retrieval systems. *Information Retrieval*, *2*(1), 115–129.

Riss, U. V., Rickayzen, A., Maus, H., & van der Aalst, W. M. (2005). Challenges for business process and task management. *Journal of Universal Knowledge Management*, *2*, 77–100.

Robertson, S. E., & Hancock-Beaulieu, M. M. (1992). On the evaluation of ir systems. *Information Processing & Management*, *28*(4), 457–466.

Saastamoinen, M., Kumpulainen, S., & Järvelin, K. (2012). Task complexity and information searching in administrative tasks revisited. *IIiX 2012 - Proceedings 4th Information Interaction in Context Symposium: Behaviors, Interactions, Interfaces, Systems*, 204-213.

Saracevic, T. (1991). Interaction between users and intermediaries in online searching. In *Proceedings of the 12th national online meeting* (pp. 329–341).

Savolainen, R. (2012). Expectancy-value beliefs and information needs as motivators for task-based information seeking. *Journal of Documentation*.

Schunk, D. H., & Greene, J. A. (2018). *Handbook of self-regulation of learning and performance*. New York: Routledge, Taylor & Francis Group.

Scollon, C. N., Kim-Prieto, C., & Diener, E. (2003). Experience sampling: Promises and pitfalls, strengths and weaknesses. *Journal of Happiness Studies*, *4*(1), 5–34.

Sellen, A. J., Murphy, R., & Shaw, K. L. (2002). How knowledge workers use the web. In *Proceedings* *of the sigchi conference on human factors in computing systems changing our world, changing ourselves - chi '02.*

Serola, S., & Vakkari, P. (2005). The anticipated and assessed contribution of information types in references retrieved for preparing a research proposal. *Journal of the American Society for Information Science and Technology*, *56*, 373-381.

Silverstein, C., Marais, H., Henzinger, M., & Moricz, M. (1999). Analysis of a very large web search engine query log. In *Acm sigir forum* (Vol. 33, pp. 6–12).

Singer, G., Norbisrath, U., & Lewandowski, D. (2012). Ordinary search engine users assessing difficulty, effort, and outcome for simple and complex search tasks. In *Proceedings of the 4th information interaction in context symposium* (pp. 110–119).

Singer, G., Norbisrath, U., & Lewandowski, D. (2013). Ordinary search engine users carrying out complex search tasks. *Journal of Information Science*, *39*(3), 346–358.

Song, Y., Ma, H., Wang, H., & Wang, K. (2013). Exploring and exploiting user search behavior on mobile and tablet devices to improve search relevance. *WWW 2013 - Proceedings of the 22nd International Conference on World Wide Web*, 1201–1211.

Spink, A. (1996). Multiple search sessions model of end-user behavior: An exploratory study. *Journal of the American Society for Information Science*, *47*(8), 603–609.

Spink, A., Bateman, J., & Jansen, B. J. (1999). Searching the Web: A survey of EXCITE users. *Internet Research*, *9*(2), 117–128. doi: 10.1108/10662249910264882

Spink, A., Griesdorf, H., & Bateman, J. (1999). A study of mediated successive searching during information seeking. *Journal of Information Science*, *25*(6), 477–487.

Spink, A., Wilson, T. D., Ford, N., Foster, A., & Ellis, D. (2002). Information seeking and mediated searching study: Part 3. Successive searching. *Journal of the American Society for Information Science and Technology*, *53*(9), 716–727. doi: 10.1002/asi.10083

Stadtler, M., & Bromme, R. (2008). Effects of the metacognitive computer-tool met.a.ware on the web search of laypersons. *Computers in Human Behavior*, *24*, 716-737.

Stanton, N. A. (2006). Hierarchical task analysis: Developments, applications, and extensions. *Applied ergonomics*, *37*(1), 55–79.

Teevan, J., Adar, E., Jones, R., & Potts, M. A. (2007). Information re-retrieval: repeat queries in yahoo's logs. In *Proceedings of the 30th annual international acm sigir conference on research and development in information retrieval* (pp. 151–158).

Teevan, J., Karlson, A., Amini, S., Brush, A. B., & Krumm, J. (2011). Understanding the importance of location, time, and people in mobile local search behavior. In *Proceedings of the 13th international conference on human computer interaction with mobile devices and services* (pp. 77–80).

Toms, E. G., Villa, R., & McCay-Peet, L. (2013). How is a search system used in work task completion? *Journal of information science*, *39*(1), 15–25.

Tonn-Eichstädt, H. (2006). Measuring website usability for visually impaired people-a modified goms analysis. In *Proceedings of the 8th international acm sigaccess conference on computers and accessibility* (pp. 55–62).

Trenck, A. v. d., Neben, T., & Heinzl, A. (2014). The impact of self-determination on the information-stopping behavior of professionals: An exploratory study in the software industry..

Urgo, K., & Arguello, J. (2022). Understanding the 'pathway' towards a searcher's learning objective. *ACM Transactions on Information Systems (TOIS)*, *40*, 1-43.

Urgo, K., & Arguello, J. (2023). Goal-setting in support of learning during search: An exploration of learning outcomes and searcher perceptions. *Information Processing and Management*, *60*(2), 103158. Retrieved from \https://doi.org/10.1016/j.ipm.2022.103158 doi: 10.1016/j.ipm.2022.103158

Urgo, K., Arguello, J., & Capra, R. (2019). Anderson and krathwohl's two-dimensional taxonomy applied to task creation and learning assessment. In *Proceedings of the 2019 acm sigir international conference on theory of information retrieval* (pp. 117–124).

Urgo, K., Arguello, J., & Capra, R. (2020). The effects of learning objectives on searchers' perceptions and behaviors. In *Proceedings of the 2020 acm sigir on international conference on theory of information retrieval* (pp. 77–84).

Vakkari, P. (2001). A theory of the task-based information retrieval process: a summary and generalisation of a longitudinal study. *Journal of documentation*, *57*(1), 44–60.

Vakkari, P. (2003). Task-based information searching. *Annual review of information science and technology*, *37*(1), 413–464.

Vakkari, P., & Hakala, N. (2000). Changes in relevance criteria and problem stages in task performance. *Journal of Documentation*. doi: 10.1108/EUM0000000007127

Van Lingen, F., Evans, D., Metson, S., Wakefield, S., Wilkinson, R., Jackson, J., ... Kuznetsov, V. (2010). Job life cycle management libraries for cms workflow management projects. In *Journal of physics: Conference series* (Vol. 219, p. 042024).

Vuong, T., Saastamoinen, M., Jacucci, G., & Ruotsalo, T. (2019). Understanding user behavior in naturalistic information search tasks. *Journal of the Association for Information Science and Technology*, *00*(0), asi.24201. doi: 10.1002/asi.24201

Wang, Y., Huang, X., & White, R. W. (2013). Characterizing and supporting cross-device search tasks. In *Wsdm 2013 - proceedings of the 6th acm international conference on web search and data mining* (pp. 707–716). doi: 10.1145/2433396.2433484

Wang, Y., Liu, J., Mandal, S., & Shah, C. (2017). Search successes and failures in query segments and search tasks: A field study. *Proceedings of the Association for Information Science and Technology*, *54*, 436-445.

Wang, Y., Sarkar, S., & Shah, C. (2018). Juggling with information sources, task type, and information quality. In *Proceedings of the 2018 conference on human information interaction & retrieval* (pp. 82–91).

Wildemuth, B. (2002, 12). Effective methods for studying information seeking and use. *Journal of the American Society for Information Science and Technology*, *53*(14), 1218.

Wildemuth, B., Freund, L., & Toms, E. G. (2014). Untangling search task complexity and difficulty in the context of interactive information retrieval studies. *Journal of Documentation*.

Wilson, T. D. (1981). On user studies and information needs. *Journal of documentation*, *37*(1), 3–15.

Wilson, T. D. (1997). Information behaviour: an interdisciplinary perspective. *Information processing & management*, *33*(4), 551–572.

Wilson, T. D. (2000). Human information behavior. *Informing Science*, *3*(2), 49–55. doi: 10.28945/576

Wu, D. (2018). Identifying and Modeling Information Resumption Behaviors in Cross - Device Search. In *Sigir '18* (pp. 1189–1192). Ann Arbor, MI.: ACM. Retrieved from \https://doi.org/10.1145/3209978.3210126 doi: 10.1145/3209978.3210126

Wu, D., Dong, J., Tang, Y., & Capra, R. (2019). Understanding task preparation and resumption behaviors in cross-device search. *Journal of the Association for Information Science and Technology*.

Wu, D., Dong, J., Tang, Y., & Capra, R. (2020). Understanding task preparation and resumption behaviors in cross-device search. *Journal of the Association for Information Science and Technology*, *71*(8), 887–901.

Wu, D., Liang, S., & Bi, R. (2018). Characterizing queries in cross-device OPAC search: a large-scale log study. *Library Hi Tech*, *36*(3), 482–497. Retrieved from \https://doi.org/10.1108/LHT-06-2017-0130 doi: 10.1108/LHT-06-2017-0130

Wu, D., & Yuan, F. (2018). Exploring dynamic change of users' cross-device search performance. *Proceedings of the Association for Information Science and Technology*, *55*(1), 931–932.

Wu, W.-C., & Kelly, D. (2014). Online search stopping behaviors: An investigation of query abandonment and task stopping. *Proceedings of the American Society for Information Science and Technology*, *51*(1), 1–10.

Xie, H. (2002). Patterns between interactive intentions and information-seeking strategies. *Information processing and management*, *38*(1), 55–77.

Xie, I. (2009). Dimensions of tasks: Influences on information-seeking and retrieving process. *Journal of Documentation*, *65*(3), 339-366.

Ye, C., Wilson, M. L., & Rodden, T. (2014). Develop, implement, and improve a web session detection model. In *Proceedings of the 5th information interaction in context symposium* (pp. 336–338).

Zach, L. (2005). When is "enough" enough? modeling the information-seeking and stopping behavior of senior arts administrators. *Journal of the American Society for Information Science and Technology*, *56*(1), 23–35.

Zhang, Y. (2020). Beyond searching: Understanding how people use search to support their creative endeavors. *PhD Dissertation*.

Zhang, Y., Capra, R., & Li, Y. (2020). An in-situ study of information needs in design-related creative projects. In *Proceedings of the 2020 conference on human information interaction and retrieval* (pp. 113–123). New York, NY, USA: Association for Computing Machinery.