Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/ipm

Investigating the role of in-situ user expectations in Web search

Ben Wang *, Jiqun Liu

School of Library and Information Studies, The University of Oklahoma, Norman, OK, 73019, USA

ARTICLE INFO

Keywords:
User search expectation
Interactive information retrieval
Online searching
User satisfaction

ABSTRACT

Pre-adoption expectations often serve as an implicit reference point in users' evaluation of information systems and are closely associated with their goals of interactions, behaviors, and overall satisfaction. Despite the empirically confirmed impacts, users' search expectations and their connections to tasks, users, search experiences, and behaviors have been scarcely studied in the context of online information search. To address the gap, we collected 116 sessions from 60 participants in a controlled-lab Web search study and gathered direct feedback on their *in-situ* expected information gains (e.g., number of useful pages) and expected search efforts (e.g., clicks and dwell time) under each query during search sessions. Our study aims to examine (1) how users' pre-search experience, task characteristics, and in-session experience affect their current expectations and (2) how user expectations are correlated with search behaviors and satisfaction. Our results with both quantitative and qualitative evidence demonstrate that: (1) user expectation is significantly affected by task characteristics, previous and in-situ search experience; (2) user expectation is closely associated with users' browsing behaviors and search satisfaction. The knowledge learned about user expectation advances our understanding of users' search behavioral patterns and their evaluations of interaction experience and will also facilitate the design, implementation, and evaluation of expectation-aware user models, metrics, and information retrieval (IR) systems.

1. Introduction

With growing research in cognitive psychology and behavioral economics, research on the role and impacts of cognitive bias attracts increasing attention in human–computer interaction (HCI) and information retrieval (IR) research community (Azzopardi, 2021; Gomroki, Behzadi, Fattahi, & Salehi Fadardi, 2021; Lau & Coiera, 2007; White, 2013). According to theories on *bounded rationality* (e.g., theory of satisficing), people usually make decisions not to maximize utility but to obtain satisfying results with the least efforts (Cancho & Solé, 2003; Kahneman, 2003; Tversky & Kahneman, 1992). In the context of information seeking, users often make decisions to continue or stop searching and browsing based on their anticipations and perceptions of information gain and search effort at both *whole-session* and *local* query levels, instead of pursuing the highest utility as it is assumed in most formal models (Agosto, 2002; Azzopardi, 2021; Pirolli & Card, 1999).

Before a search session is initiated, users usually behave under the impact of *pre-search expectations*, such as expectations or anticipations of task difficulty and task complexity, which represent their before-search impression regarding different aspects of the whole search task, and are usually triggered by pre-search task descriptions and study instructions (Kelly et al., 2009; Liu & Shah, 2019). Specifically, in contrast to pre-search expectations at the task level, users' *in-situ search expectations*, which is the focus

* Corresponding author.
E-mail addresses: benw@ou.edu (B. Wang), jiqunliu@ou.edu (J. Liu).

https://doi.org/10.1016/j.ipm.2023.103300
Received 10 August 2022; Received in revised form 2 January 2023; Accepted 26 January 2023

of our research, refer to their *within-session anticipations* of future gains and efforts in the next *query segment*.[1] Rather than connecting expectations to the general information gain or search efforts, in this study, we investigated user expectations in three behavioral features, including the expectation of useful page number, click number, and spending time at the query level, based on the previous parameter design on users' information gain and effort in formal user models (Azzopardi & Zuccon, 2016; McGregor, Azzopardi, & Halvey, 2021; Zhang, Liu, Mao, Zhang, & Ma, 2020). For instance, Azzopardi (2011, 2014), Azzopardi and Zuccon (2016) developed search cost-benefit models based on a series of observable actions, such as clicks and dwell time on examined pages. Our exploration of in-situ expectation starts with these widely employed behavioral features and solicits query-level expectation labels based on them.

Users' in-situ expectations and local search tactics usually change over time due to variations in in-situ search intentions, local search actions, and query-level search outcomes. Search intention can be regarded as a comprehensive expectation of search goals in different aspects. In this study, the in-situ expectations can quantify the search intention in terms of the useful page number, the click number, and spending time at the query level. In addition, the variations in in-situ search expectations across queries go beyond pre-search fixed impressions and better capture users' dynamic perceptions at within-session level. From a user-bias-aware perspective, users' search expectations could act as real-time *reference points* in their search process and affect their information behaviors (e.g. reference-dependence effect and the anchoring bias) (Abeler, Falk, Goette, & Huffman, 2011; Backus, Blake, Masterov, & Tadelis, 2017, 2022; Furnham & Boo, 2011; Liu & Han, 2020). In addition, user expectation was conceptualized and empirically examined as an essential component under expectation confirmation theory (ECT) (Bhattacherjee, 2001; Hossain & Quaddus, 2012). ECT claims that users' continuity intention or satisfaction is determined by their expectations and actual (perceived) performance of the product or service. According to studies of ECT in the information retrieval (IR) field, The combination effects of users' expectation and confirmation affect users' search behaviors as well as overall satisfaction (Cox & Fisher, 2004; Liu & Shah, 2019). Although these studies highlight the importance of studying the impacts of expectation disconfirmation on user behavior and experience, how users' in-situ expectations vary *during* the process of search interaction still remains an open question.

Searching under the impact of changing in-situ expectations regarding gains and efforts may lead to significant variations in search evaluation criteria and search tactics during information seeking and retrieval. These user-bias-related behavioral and perceptual variations could not be fully captured with oversimplified user models developed under the assumptions that individuals behave rationally (Evans, 2020). Given the critical role of user expectations, previous studies have made assumptions and simulated user expectations in building and evaluating new user models and search algorithm evaluation metrics (e.g., Chen et al., 2021; Cox & Fisher, 2009; Moffat, Bailey, Scholer, & Thomas, 2017). While simulated models helped incorporate the knowledge about user expectations and cognitive biases into IR evaluation studies and large-scale offline experiments, there were few studies or direct empirical evidence that revealed the way in which users' in-situ expectations and expectation disconfirmation status (i.e., exceed or below expectations) are connected to search tasks and affect their search interaction experiences. The simulated search expectations have not yet been examined empirically, and their actual impacts on users' search behaviors and experience at different stages of search sessions still remain understudied.

To address the gaps above, our research (1) investigates the relationships between users' in-situ search expectations and contextual features (including both pre-search perceptions and task characteristics); (2) examines the impact of in-situ expectations on information search behaviors and user experience (in particular, search satisfaction). To achieve this, we conducted a user study to collect *direct evidence* (i.e., users' self real-time labeling) about the changes of users' in-situ expectations regarding information gains and search efforts under each individual query segment in Web search sessions. In addition, we adopted a mixed-method approach and utilized both statistical results and qualitative feedback to examine the characteristics of users' search expectations, contextual factors affecting in-situ expectations, as well as the impacts of search expectations on user behavior and satisfaction. The main contributions and implications include:

- Based on the search log and annotation data collected from 60 participants, we characterized users' in-situ search expectations regarding information gain and search efforts with explicit annotations. We then identified contextual factors affecting search expectations, including task type, previous search experience, and in-situ search experience under motivating tasks. Those results help explain why users have different expectations in various situations. Findings regarding these relationships advance our knowledge about how users form their in-situ expectations in searching unfamiliar topics at different stages. The results allow us to improve existing user models underpinning search interaction simulations and develop *expectation-aware* evaluation metrics.
- We also found that users' search behaviors and level of satisfaction are associated with and affected by their search expectations. The expectations mainly affect their browsing behaviors, including mouse-based and dwell time-based behaviors. These findings can enhance our understanding of the impact of in-situ user expectations on users' search decision-making in a specific search task, at different query positions, or in browsing results at lower ranks, and on evaluating the search experience.
- At the practical level, findings from our study can help incorporate user expectation as a key cognitive factor in personalizing user-centered IR algorithms (e.g., result ranking, query term recommendation, adaptive search interface layouts) and may allow systems to predict users' search tactics and interaction experience more accurately in real-time information search.

[1] A query segment starts with a query, includes all the browsing, clicking and evaluation actions associated with the query, and ends before the next query is issued (Rha, Mitsui, Belkin, & Shah, 2016).

2. Related work

2.1. Cognitive bias and reference dependence

In contrast to the assumption that individuals always pursue maximized utility, a series of cognitive psychology and behavioral economics experiments confirmed that people often seek to find satisfying or "good enough" options in decision-making scenarios rather than keep exploring the best options (Azzopardi, 2021; Cancho & Solé, 2003; Kahneman, 2003). This is because people are *boundedly rational* when facing decision-making tasks under uncertainty and cannot handle the complex calculation of potential benefits and costs or exhaust all possible alternatives (Simon, 2000; Tversky & Kahneman, 1991, 1992). Regarding people's behaviors and cognitive bias, behavioral economics has been widely applied in human–computer interaction to improve the design of technologies and assist users effectively (Caraban, Karapanos, Gonçalves, & Campos, 2019; Cockburn, Quinn, Gutwin, Chen, & Suwanaposee, 2022). These cognitive, perceptual, and contextual limits also affect people's behaviors and decisions in the information-seeking processes, such as query term selection, browsing search result lists and clicking retrieved documents, and continuing or abandoning search sessions (Pirolli & Card, 1999). One of the user biases caused by cognitive limits, which motivates individuals to make decisions based on gains and costs relative to reference points in mind, is *reference dependence* bias (Lieder, Griffiths, M Huys, & Goodman, 2018; Tversky & Kahneman, 1991). The reference-dependence effect acts on the reference points that people refer to when evaluating perceived gain and cost in the short term. Reference dependence bias can act in multiple forms, including anchoring, confirmation, and expectation. Anchoring bias motivates people to assign more weight to information they first observe, affecting their decisions when finding new information (Lau & Coiera, 2007). Confirmation bias describes peoples' tendency to find information that supports their hypotheses (Hergovich, Schott, & Burger, 2010; Mendel et al., 2011; White, 2013). In-situ user expectations also act as a common reference point and reflect people's estimation of future gain and cost based on past experience (Backus et al., 2022).

Theoretically, users will adjust their information needs and search intentions during the search process according to Berry picking model and information foraging theory (Bates, 1989; Pirolli & Card, 1999). In the evolving search process, users could make decisions based on their perceived gain and cost. These perceptions are not absolute values but are relative and dependent on their previous anticipations (Simon, 2000; Tversky & Kahneman, 1991, 1992). Therefore, users can form in-situ expectations before searching and have cognitive biases when they examine search results referring to these expectations. Azzopardi (2021) and Gomroki et al. (2021) further reviewed studies on cognitive biases in information retrieval, especially biases brought by users' pre-existing opinions or in-situ perceptions. In addition, these cognitive biases are interconnected and have complex effects on users' search decision-making (Azzopardi, 2011, 2021; Azzopardi & Zuccon, 2019). For example, the anchoring bias forms an information anchor or a reference point for the user. Then the user can make the decision by referring to the anchor or reference point when observing new information with the confirmation bias (Liu & Han, 2020). Users may also abandon the query or stop the task if they perceive low search result quality or less useful information (Wu & Kelly, 2014). Their perceptions are subjective and might relate to their search goals, in-situ reference levels, and expectations of the search result quality. With a pre-search expectation being established, a decrease in search result quality (especially when below the expectation level) may be perceived as a loss in search interaction, leading to an unpleasant search experience and a change of in-situ search tactics. Based on users' biased perceptions of gain and cost in the reference-dependence effect and the anchoring effect, previous studies implemented these biases into user modelings and investigated users' behaviors for information retrieval evaluation (Chen, Zhang, & Sakai, 2022; Liu & Han, 2020; Moffat et al., 2017; Zhang et al., 2017). These studies showed substantial improvements with assumptions of users' cognition and behaviors in the information-seeking process. However, not much research investigates how users form the reference points, estimate their gains and costs, or decide to keep or stop searching based on these perceptions. Therefore, it is important to investigate the role of user perceptions and explore how their perceptions of future gain and cost affect the information-seeking process. This attempt will also help researchers better understand the impacts of reference dependence bias and biased evaluation results in online information search.

2.2. Expectation confirmation effect

The role of in-situ expectation has been widely investigated in Economics research. Users' expectations can serve as a mediator affecting their satisfaction and technology adoption according to expectation confirmation theory (ECT) (Venkatesh & Goyal, 2010). ECT was first introduced to analyze consumers' satisfaction with the product and predict their intention to repurchase (Hossain & Quaddus, 2012; Qazi, Tamjidyamcholo, Raj, Hardaker, & Standing, 2017). Then the application of ECT was expanded from economic behaviors to information behaviors, and the expectation confirmation model (ECM) was developed based on ECT to better explain user behaviors and continuance in using information services (Bhattacherjee, 2001; Franque, Oliveira, Tam, & de Oliveira Santini, 2020). Particularly, ECT and ECM are prevalent approaches in human–computer interaction to investigate users' continuance intention of and satisfaction with using electronic devices and online services (Chen, Yen, & Hwang, 2012; Huang, 2019; Steelman & Soror, 2017).

Involving user expectations could help understand and predict user satisfaction in information retrieval. Improving user experience and satisfaction is one of the goals of information retrieval systems, and more researchers focus on evaluation metrics related to users' satisfaction beyond modeling their behaviors (Moffat, Thomas, & Scholer, 2013; Piwowarski, Dupret, & Lalmas, 2012; Zhang et al., 2017; Zhang, Mao, et al., 2020). Incorporating user characteristics and representing cognitive biases is a key step toward developing user models that better correlate with search satisfaction (Al-Maskari & Sanderson, 2010; Liu & Han, 2020).

As user satisfaction is complex and not statically based on the information retrieved by the system, it is more dynamic and dependent on the user's implicit state. Users' perceptions of future gain and cost influence their search experience and satisfaction through their search expectations. Cox and Fisher (2004) analyzed user expectation as the mediator of user satisfaction. They collected predefined query-result sets and asked participants about their expectations of the query performance and satisfaction with the search results. They found users' satisfaction was significantly associated with the difference between the quality of the search results and users' expectations. Based on the findings, they further investigated the expectation effect and developed an expectation-based user model to predict users' satisfaction levels (Cox & Fisher, 2009). Their model considered expectation as a mediator affecting users' judgment of the search results and their levels of satisfaction.

Recently, the expectation confirmation effect also raised researchers' awareness in the field of information retrieval. Liu and Shah (2019) investigated the expectation confirmation effect on users' expectations of task difficulty. They collected participants' annotations on pre-search expected difficulty and post-search perceived difficulty. They found that the disconfirmation of expected difficulty influenced users' search behaviors and success. This research demonstrated the importance of expectation confirmation in users' information-seeking process. However, other than task difficulty, more expectation factors such as information gain and cost are not investigated. Furthermore, the forming process of general search expectations and their specific impacts on search behaviors remain understudied.

3. Research questions

Since users' in-situ search expectations regarding gains and efforts are associated with reference dependence bias and may impact their search behaviors and satisfaction, this study aims to fill the gap by investigating the role of user expectations regarding information gains and search efforts. This work also serves as one of the initial steps toward building and evaluating broader bias-aware information search systems and system evaluation metrics. Specifically, we aim to answer the following research questions:

RQ1. What are the characteristics of users' in-situ expectations regarding useful page number, click number, and spending time?
RQ2. What are the contextual factors that affect users' search expectations?
RQ3. How do users' search expectations affect their search behaviors and levels of satisfaction?

4. Methodology

4.1. Participants and search task design

To answer the proposed research questions above, we conducted a user study to collect data on both search interactions and users' explicit feedback on their expectations and experience (e.g., expectation confirmations of gain and effort, and level of satisfaction). This user study recruited and assigned predefined complex search tasks of varying types to 60 participants (undergraduate students from the University of Oklahoma). When recruiting participants, several questions were given to them to collect their background information, including their year in college, major, and experience in searching for information beyond only finding facts. We adopted the four journalism tasks applied in previous studies (Liu, Mitsui, Belkin, & Shah, 2019; Rha et al., 2016) and recruited participants not from the journalism major to control prior experience in these task types.

Table 1 lists the task types, topics, and descriptions. The tasks are designed based on four work types in journalism: Copy editing (CE), Story pitch (SP), Relationship (RE), and Interview preparation (IP). These tasks can also be categorized by the task product (factual/intellectual), goal (specific/amorphous), level (document/segment), and named (explicitly named target or not), which cover typical task types in the information retrieval task classification (Li, 2009; Liu et al., 2010). These tasks have been scientifically demonstrated to be beneficial for motivating multi-round search interactions and regulating the possible impacts of a variety of contextual factors (e.g., variations in participants' topic familiarity and domain knowledge) (Cole, Hendahewa, Belkin, & Shah, 2015; Li & Belkin, 2008). Each of the four task types has two distinct and uncommon topics: coelacanths (CO), and methane clathrates and global warming (ME). The information in curly braces in Table 1, such as CO and ME, is the details related to the task topic provided to participants and replaced by an abbreviation here. These two topics are chosen to control the variable of participant familiarity with the issue, as neither is likely to be familiar to our participant pool. Thus, there are eight different tasks in total, combining four task types and two task topics. We allocated the eight tasks among 60 participants based on a rotation to balance task number by topic and task type. Each participant will be asked to complete two tasks of different topics and types (e.g., CE-CO and SP-ME) following the task instructions. The task instructions introduce the work scenario to the participants and ask them to search for useful information on the Web to complete the task. For example, they need to take notes about the statement correctness with supportive pages for the Copy Editing task or collect information and analyze the relationship with supportive pages for the Relationship task. Their notes with required information can be the task response, but they are not required to write a full-length article.

4.2. User study procedure

The user study process is demonstrated in Fig. 1. The study begins with instruction and a warm-up task for the participants to get familiar with the study process and read the task description. After reading the task description and requirements, participants

Table 1
Task descriptions in the user study (Liu et al., 2019).

Task type	Instructions
Copy Editing (Factual Specific Segment Named True)	Assignment: You are a copy editor at a newspaper, and you have only 20 min to check the accuracy of six italicized statements in the excerpt of a piece of news story below. Task: Please find and save an authoritative page that either confirms or disconfirms each statement.
Story Pitch (Factual Amorphous Segment Named False)	Assignment: You are planning to pitch a science story to your editor and need to identify interesting facts about {CO}{ME} Task: Find and save Web pages that contain the six most interesting facts about {CO}{ME}.
Relationship (Intellectual Amorphous Document Named True)	Assignment: You are writing an article about {CO}{ME} You have found an interesting article about {CO}{ME}, but in order to develop your article, you need to be able to explain the relationship between key facts you have learned. Task: In the following, there are five italicized passages, find an authoritative Web page that explains the relationship between two of the italicized facts.
Interview Preparation (Intellectual Amorphous Document Named False)	Assignment: You are writing an article that profiles a scientist and their research work. You are preparing to interview {a researcher in CO}{a researcher in ME} Task: Identify and save authoritative Web pages for the following: Identify two (living) people who likely can provide some personal stories about {the researcher} and his work. Find the three most interesting facts about {CO}{ME} Find an interesting potential impact of {the researcher} 's work.

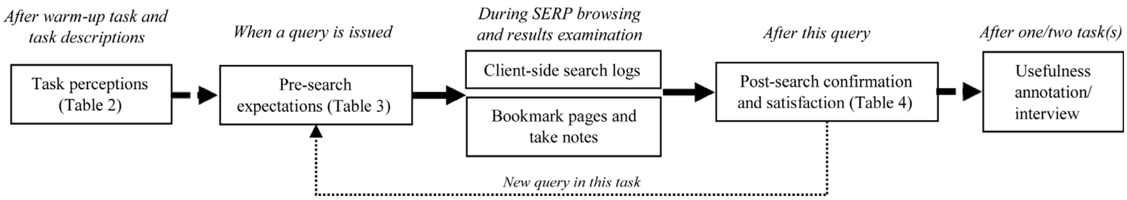


Fig. 1. User study and data collecting procedure.

Table 2
Pre-task questionnaire (assigned before each task started).

Variable	Question and scale
Topic familiarity	How familiar are you with the topic of this task? 1 (not at all) – 4 (somewhat) – 7 (extremely)
Task experience	How much experience do you have with this kind of task? 1 (not at all) – 4 (somewhat) – 7 (extremely)
Task difficulty	How difficult do you think it will be to find the information for this task? 1 (not at all) – 4 (somewhat) – 7 (extremely)

will take a survey about their initial perceptions of the task, including topic familiarity, task experience, and task difficulty. The survey questions and scales are listed in Table 2. These perceptions can reflect users' previous experience with the task type and pre-knowledge about the task topic (Liu et al., 2019). We collected data on those perceptions as contextual factors and investigated their impacts on users' in-situ expectations in the analysis step.

For each task session, participants are asked to input queries. After entering the query and before browsing the retrieved pages, users were directed to a page of the pre-search questionnaire about their expectations of the useful page number, the click number, and spending time. The survey questions are listed in Table 3. We designed these questions based on the assumptions about users' search expectations and behaviors made in existing formal models and simulation-based experiments (Backus et al., 2022; Moffat et al., 2017; Zhang et al., 2017). These behaviors (i.e., finding useful pages, total clicks, and spending time) are frequently studied behavioral measures and used for calculating users' information gain and effort (Azzopardi & Zuccon, 2016; McGregor et al., 2021; Zhang, Liu, et al., 2020). Therefore, we assume that when users answer questions about their expectations of those behaviors, they can estimate the information gain and search effort, which can serve as reference points during the search. The measurements of the expectations of useful pages and clicks are numeric. For the expectation of spending time, we provide five ordinal options instead of asking users to estimate the exact time they want to spend. Each option is connected with a description of how users may perceive the time cost (e.g., instantly, quickly, need more time). These descriptions are simple and intuitive and may help users measure their expectations of time cost. Besides the three types of expectations, we also asked users if they had any other expectations during post-search interview sessions. Here and in the remainder of this paper, we use (*expectation of*) # Useful pages, # Clicks, and spending time to specifically represent expectation features, which differ from users' actual behaviors of clicks and dwell time (the actual time users spend on reading content pages).

After completing the short questionnaire of pre-search expectations, participants can examine the retrieved documents on search engine result pages (SERPs) and collect information according to the task requirements. They need to take notes and bookmark/save supportive pages to fulfill the task requirements. To capture their behavioral variations of search activities, we will gather behavioral data from the search log (e.g., querying behaviors, mouse-based browsing behaviors, dwell time-based browsing behaviors). The

Table 3
Pre-query questionnaire (assigned before each query issued).

Expectation type	Question
# Useful pages	How many useful pages do you expect to find? (Numeric)
# Clicking results	How many results do you expect to click before obtaining the expected number of useful pages? (Numeric)
Spending time	How much time do you expect to spend on this search? (Ordinal) • fewer than 30 s (I can find the useful result instantly) • 30 s to 1.5 min (I can find the useful result quickly after inspecting it) • 1.5 min to 3 min (I need some time to read the results, but it will not take so long.) • 3 min to 5 min (I need some time to read the results.) • more than 5 min (I need more time to read the results carefully.)

Table 4
Post-query questionnaire.

Variable	Question and scale
Gain confirmation	How much useful information did you obtain through this search/all the queries? 1 (much less than expected) – 4 (just as expected) – 7 (much more than expected)
Effort confirmation	How much effort did it take for you to collect useful information during this search/task? 1 (much less than expected) – 4 (just as expected) – 7 (much more than expected)
Satisfaction	Rate your satisfaction level with this search/task. 1 (very unsatisfied) – 2 (unsatisfied) – 3 (neutral) – 4 (satisfied) – 5 (very satisfied)

Table 5
Interview questions.

Question	Sub-questions
What is the basis for your pre-search expectations?	• How did you determine your expectation of useful information/spending time for the first query? • After browsing results from the first query and getting some experience, how did you change your expectation of useful information/spending time in following queries? • Other than useful information/spending time, what did you use to represent your expectation of gain/cost?
What do you feel if retrieved search results exceed/do not meet your expectations?	• How did you change your following search actions when useful information exceeded/do not meet your expectations? • How did you change your following search actions when spending efforts exceeded/did not meet your expectations? • After adjusting your expectation, how did you change your search activities accordingly?
What search obstacles and challenges did you encounter during the search sessions?	• How did these problems affect your subsequent search expectations and actions? • How did you address these problems in your previous experience? • Other than the problems you mentioned in the post-query questionnaire, what problems did you meet in this study?

behavioral data will be collected through a browser extension based on an open-source user study toolkit (Chen et al., 2021) with essential modifications for this study. Those behaviors will be further processed as behavioral features (i.e., behavioral features in Table 6) for the analysis.

After each query and before formulating the next query or stopping the search, participants are asked to complete a survey about their gain confirmation, effort confirmation, and query satisfaction. The questions are listed in Table 4. After completing one task, participants need to review the clicked pages for each query and evaluate each page with a 3-point scale usefulness score. We collect the usefulness scores and calculate the search result quality (i.e., the search result quality features in Table 6) for further analysis.

After participants complete two search tasks, there is a post-search interview. We designed 9 open-ended questions to collect more detailed qualitative information regarding participants' search expectations and experience in the tasks, including the formation of in-situ expectations, reasons behind (changes of) search behaviors, and problems encountered during the search. Questions regarding search obstacles were developed and modified based on the questionnaire from Sarkar, Mitsui, Liu, and Shah (2020). The interview questions are listed in Table 5. We asked these questions according to our observations of their search behaviors and in-situ feedback from the questionnaires above.

In this study, participants are searching on modified Bing pages. We chose Bing as one of the largest commercial search engines to simulate the natural search scenario. In addition, we removed ads, sponsored results, videos, and Q&A but kept only title-snippet style documents to control the vertical type of the retrieved documents on SERPs. The whole process of prompting questionnaires and search log collecting is automatic. Each search task will take participants about 20 min to complete, and one study session takes about 1 h and 30 min, including two search tasks, usefulness annotations, and an interview. Each participant will receive $25 as compensation for this study.

4.3. Analysis

After the data collection, based on the users' search logs, in-situ survey data, and post-search interviews, we conducted three analyses to answer the research questions. For each research question, we did both a qualitative analysis from the interview and

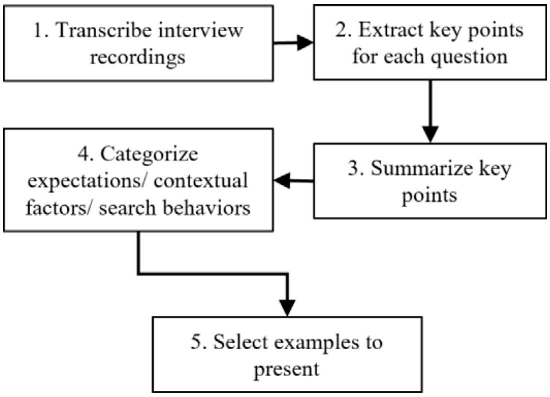


Fig. 2. The coding process of the qualitative analysis.

Table 6
Contextual factors and search behavior features.

Features	Description
In-situ search experience - continuity and feedback	
QueryOrder	Index of the query in a session.
LastGain	Gain confirmation of the previous query.
LastEffort	Effort confirmation of the previous query.
LastSat	Satisfaction feedback of the previous query.
In-situ search experience - search result quality of the previous query	
LastDCG	Discounted cumulative gain.
LastRBP	Rank biased precision.
LastERR	Expected reciprocal rank.
LastMaxUse	Maximum usefulness score of clicked results.
LastAvgUse	Average usefulness score of clicked results.
LastUseNum	Number of useful pages.
Querying behaviors	
QueryLength	Number of terms used in an issued query.
UniqueTerm	Number of unique terms used in an issued query.
NewTerm	Number of new unique terms used in an issued query.
QuerySim	Similarity between the current query and the previous query.
Browsing behaviors - Mouse based	
ClickCount	Number of clicks.
Clicks@3	Number of clicks between ranks 1-3.
Clicks@5	Number of clicks between ranks 1-5.
Clicks@5+	Number of clicks below rank 5.
ClickDepth	The deepest or lowest rank of the clicked result.
AvgClickRank	The average rank of clicked results.
UniquePage	Number of unique clicked pages
ScrollDist	Total scrolling distance.
Browsing behaviors - Dwell time based	
QueryDwellTime	Total dwell time within a query segment.
TimeFirstClick	Time delta between the start of a session and the first click.
TimeLastClick	Time delta between the start of a session and the last click.

quantitative analyses based on features from the questionnaires and search log data. The qualitative analysis is based on a coding process shown in Fig. 2 by two researchers participating in observation and interview during the user study. The two researchers compiled users' responses in the interview data to summarize and categorize the key points related to users' expectations, contextual factors, and search behaviors. Then, we chose two to three examples from users' responses to illustrate each identified category.

According to the research questions and features extracted from the qualitative analysis, we did the quantitative analysis in three aspects. First, we analyzed the characteristics of expectation features by visualizing the distributions of expectation features and investigated the relationship between expectations of # Useful pages/ # Clicks and spending time. Second, we investigated the contextual factors including features of task characteristics (i.e., task types and topics), users' previous experiences (features in Table 2), and in-situ search experience. The in-situ search experience listed in Table 6 is represented by users' feedback and the search result quality of the previous query in the session, including confirmations, satisfaction, and effectiveness metrics. The effectiveness metrics are usefulness-based discounted cumulative gain (DCG), rank biased precision (RBP), expected reciprocal rank

Table 7
Hypotheses of the statistical tests (only alternative hypotheses are listed).

H1: There are correlations between user expectation features (Table 3)
H2: Users have different expectations among task types and between task topics
H3: There are correlations between user expectations and previous task experience (Table 2)
H4: There are correlations between user expectations and in-situ search experience at the query level (i.e., in-situ experience features in Table 6)
H5a: There are correlations between user expectations and search behaviors at the query level (i.e., querying and browsing behaviors in Table 6)
H5b: There are correlations between user expectation confirmations of the previous query (i.e., LastGain and LastEffort) and search behaviors of the current query.
H6a: There are correlations between user expectations and satisfaction at the query level.
H6b: There are correlations between user expectation confirmations and satisfaction at the query level (Table 4).

(ERR) (Chapelle, Metlzer, Zhang, & Grinspan, 2009; Dupret, 2011; Moffat & Zobel, 2008) and other measures widely implemented in previous research (e.g., MaxUse, AvgUse, UseNum) (Liu & Han, 2020; Mao et al., 2016). We also involved the query order as the in-situ search experience factor to represent the search continuity. In addition, we analyzed those factors in different tasks and query positions (the first/end queries and queries with the peak/bottom values of expectations) to investigate how contextual factors would affect users' expectations in different tasks or at different search stages. Third, we investigated the influences of expectations on users' search behaviors. We extracted behavioral features from the search log data and divided them into querying behaviors, mouse-based browsing behaviors, and dwell time-based behaviors listed in Table 6. These behavioral features can reflect how users formulate queries and browse the search results.

Regarding the quantitative analysis, we examined the normality of data distribution and chose statistical tests accordingly to test the differences in expectations among task types and between task topics and find the associations between expectation features and contextual factors or behaviors. As the data contained ordinal variables and was not normally distributed (which did not fulfill the distribution requirements of Pearson's correlation tests or ANOVA), we selected Kruskal–Wallis nonparametric test to examine the differences among task types/between task topics and employed Spearman's rank correlation to calculate the correlation coefficients (Hollander, Wolfe, & Chicken, 2013). The hypotheses are listed in Table 7. There are eight main hypotheses, and in each hypothesis, there are multiple features involved (e.g., experience features and behavioral features). We did exploratory analysis and statistical tests on the relationship between these features and our proposed expectation/confirmation features. As the number of statistical tests is high, a correction for the multiple comparisons problem is required for obtaining robust results. Thus, we implemented the false discovery rate (FDR) correction (Benjamini & Hochberg, 1995) and reported the corrected p-values for the statistical tests.

5. Results

5.1. Descriptive characteristics

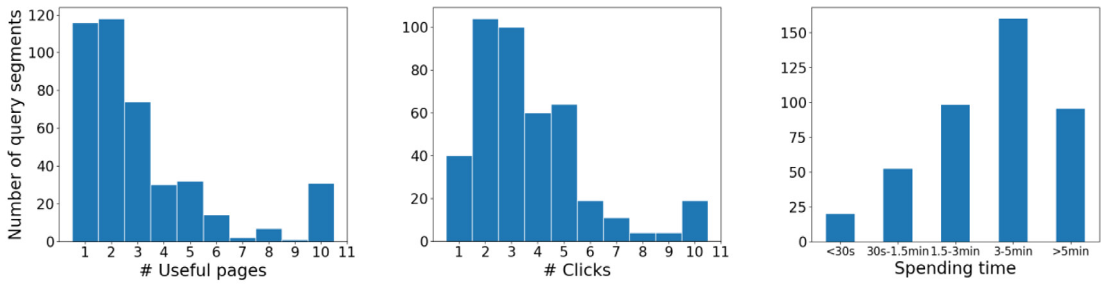
As a result of the user study, we recruited 60 undergraduates as participants. Based on the year in college, there are 12 freshmen, 14 sophomores, 17 juniors, and 17 seniors. Each participant had two tasks different in both the task type and the topic. Therefore, we collected 116 task sessions after removing 4 sessions with essential data missing. As we have four task types and two task topics, there are 30 sessions for each task type and 60 for each topic. For the familiarity with the task topic, the third quartile familiarity is 2 (on a scale of 7), indicating that 45 out of 60 participants were unfamiliar with the task topic before the study. For the experience on the task type, the third quartile is 4 (on a scale of 7), indicating that 45 of them have some experience or less on the task type. For participants' search activities for the task, we collected 418 queries in total, with 3.5 queries on average for each session. 93 out of 116 sessions have two and more queries. Participants viewed 529 SERPs and 1955 documents in total and 16 documents on average per task session. The long search sessions and high interactions demonstrate the task complexity as we expected. Regarding the three research questions, the three subsections below answer the research questions, respectively.

5.2. RQ1: Expectation features

For RQ1, to investigate users' in-situ query-level expectations, we first analyzed the interview data and extracted key points from participants' perspectives about expectations in terms of information gain and search efforts and summarized four expectation features. The four main expectation features are listed in Table 8. We selected examples from participants' responses in the interview, and [SXX] denotes the participant's ID. For the first three pre-defined expectation features, more than half of the participants could describe their estimations of useful pages, clicks, or spending time to represent their expectations. Their perspectives on useful pages and spending time are aligned with our definitions of gain and effort. However, for the number of clicks, it is difficult for participants to distinguish between the gain or effort. For example, about eight participants thought getting more useful pages required more clicks, but the effort has more weight on the reading effort than the click itself. In addition, participants who could estimate the spending time preferred to consider less time as an indicator of high efficiency.

Table 8
Expectation features and examples from the interview.

Expectation feature	Description	Example
Number of useful pages	Predefined gain factor, the number of useful pages the user expects to find.	Ï knew that I was going to be doing more work in making formulating my own answer, so I thought that there would be more information to look through, and I picked higher numbers (for the expectation of useful pages).” [S04] “I just figured that there would be three to four to five useful results just on that first initial page of the search.” [S06]
Number of clicks	Predefined effort factor, the number of clicks before reaching the expectation of useful pages.	“Amount of pages that you have to visit is how much effort you’re putting in.” [S10] “At the beginning (I) expected to find or to go through more pages, (but) I didn’t realize that usually.” [S19] “The less time you do (read), the more (clicks) you can do in a quicker time frame, (so) I don’t think that the more clicks you do define (gain or effort of) the results.” [S22]
Spending time	Predefined effort factor, the time the user expects to spend in this query under the 20 min constraint.	“...I expect to eventually find everything without much trouble, ..., (it is) the expectation of finding something fairly quickly.” [S10] “...I figured the time will be much longer because it was more reading involved in more technical terms rather than reading facts, which is pretty much like just scanning through.” [S01]
Answer for the task	Users’ general gain, the user expects to find information according to the query and the task.	“...Since I put (the query) in the quotation, I thought I would get the exact answer.” [S08] “...As I came in here being focused, I wanted to find the correct right information.” [S12]



(a) (b) (c)

Fig. 3. Distribution of expectation features.

During post-search interviews, we asked participants if they had any other expectations during the search. 20 participants reported difficulty giving an accurate number of useful pages, clicks, or spending time because they were unfamiliar with the task topics. Instead, they could offer a rough estimation of expected gains and efforts and aim to find the answer for the task. Although they mentioned their intention to find specific information to complete the task, they found it difficult to clearly estimate how many results they needed to click or how much time they needed to spend. In addition, most of these participants only issued one query in the session, so they did not get the chance to form specific expectations with knowledge learned from the first and only one query. In contrast, for users who issued more than one query in their sessions, even though they did not have exact expectations in the first query due to the unfamiliar topics, they became clearer about their information needs and had specific expectations in the following queries.

Fig. 3 shows the distribution of expectations. Figs. 3(a) and 3(b) are the distributions of the number of useful pages (# Useful pages) and the number of clicks (# Clicks). For most queries, participants expected fewer than five useful pages or clicks, and for about 20 queries, participants set 10 or more for their expectations. For the expectation of spending time, Fig. 3(c) shows for more than half of queries, participants chose 1.5 to 5 min, which are the median to longer time options, given the 20-min constraint.

The distributions of expectations also reflect participants’ difficulties in explicitly describing their expectations of useful pages or clicks. Among participants with expectations of more than ten useful pages or clicks, only two or three of them did reach the number with such high expectations based on our observation. The high expectations may come from their irrational estimation or confusion. In addition, although the average of # Clicks is higher than # Useful pages according to Figs. 3(a) and 3(b), 26 participants set the # Click lower than # Useful pages at the first time. Even though we carefully explained the relationship between useful pages and clicks, about five participants only increased # clicks, but it was still lower than # Useful pages. They were confused about the expectation of the number of useful pages and clicks until we further explained these concepts in the interview. As reasons for their confusion, they might just want to find such more useful pages with fewer clicks or estimate the number of useful documents retrieved by the search engine instead of the useful documents they would inspect. This confusion also indicates that these participants might not have clear expectations of an exact number of useful pages or clicks in real-life situations.

To further analyze the relationships between the expectation features, we also calculated the correlation coefficient between the spending time and # Clicks, respectively. The correlations are significant between the spending time and # Useful pages and #

Table 9
Contextual factors affecting expectation.

Contextual factor	Description	Example
Task characteristics	The impact of the task itself, including the task topic and the task type.	"For famous (topics) like climate change, I would expect to find more things. Not like the fish (coelacanth). I wouldn't expect to find much because we do not talk as much about that fish." [S05] "I assumed (looking for a single fact) would be a little easier than trying to make a connection between two ideas, ..., I could find that one fact at least without having to look through multiple things." [S13]
Previous experience and knowledge	Users' prior knowledge of the task or general search experience, including topic familiarity, task type experience, perceived difficulty, year in college, and other search habits.	"(The unique information) is extremely specific. It's extremely hard to find, so I expect I spend more time finding that small thing." [S05] "Because I'm a senior, ..., I have a lot of experience using the search engine and doing research for projects, ..., I know how to put good queries in, and so from the beginning, I assume that I could complete the tasks pretty quickly and without much effort." [S04] "I would expect the more and double the number of pages that I would actually click onto the number of pages I'm actually going to use. I personally like to check a bunch of different pages really quickly." [S07]
In-situ search experience	If the results in the last query exceeded or did not meet the user's expectation.	"I found the wrong thing, and I couldn't find it (the relevant information), (but) my expectation for the next search was higher, I was expecting more." [S26] "(Searching a specific number) is not going to work, so then I had to switch gears to something else." [S07]

Clicks, and the Spearman's rank coefficients are 0.211 and 0.331 with corrected $p << 0.01$, respectively. The result indicates that the expectation of spending time is more associated with the expectations of clicks than related to the expectation of useful pages. This is because participants with specific search intentions expected to find fewer pages about specific information with more clicks and more time.

5.3. RQ2: Contextual factors affecting user search expectations

For RQ2, to investigate why users have these expectations in the search process and what contextual factors affect their in-situ expectations, we extracted and categorized these factors from the interview and analyzed the relationships between these factors and expectation features.

Table 9 shows three categories of common contextual factors affecting users' expectations. First, in experimental settings, the task itself may affect users' expectations. For example, different task types required participants to find general or specific information accordingly. For the task topic, participants generally mentioned that global warming (SE) was more popular than the coelacanth (CO), although they did not have intensive knowledge of either. As they had little knowledge about the task topic in this study, they could refer to experience in similar task types and general search experience. Almost all participants mentioned they measured their expectations based on previous search experience. Besides, their background, such as their year in college, may be related to their search skills and ability to handle unfamiliar tasks. In addition, users' expectations are also influenced by their in-situ experience during the search process. Whether the results in the previous query exceed or do not meet users' expectations will also affect their search actions and expectations for the current query.

For the quantitative analysis, Figs. 4, 5, Table 10, 11, and 12 are the results of relationships between these factors and expectation features. We first investigated the expectations under different tasks in experimental settings. Fig. 4 shows the expectation distributions among the four task types and between the two topics. We used the box-and-whisker plot to visualize the distribution with the green line inside the box as the median, the upper and lower boundaries as the first quartile Q1 and the third quartile Q3, and the minimum and maximum bars as extensions to 1.5*(Q3-Q1), and outliers. Table 10 lists the results of the Kruskal–Wallis test for investigating the difference in the previous experience factors (e.g., task type experience, task topic familiarity, and perceived difficulty) and expectations among the task types and between the task topics. According to the test results, there are significant differences in # Useful page, # Click, spending time, and perceived difficulty among different task types. For example, from Figs. 4(a), 4(b), and 4(c), the participants had lower expectations of useful pages for the task types CE and RE because these two tasks have explicitly named information targets as search results. On the other hand, participants in the task type IP had a higher variance in # Useful pages, # Clicks, and spending time, probably because this task type asked participants to have intellectual task products with amorphous goals (facts and impacts about a person). For the task topic, although there are significant differences in the topic familiarity between the two topics, there are no significant differences in expectation features. Accordingly, in Figs. 4(d), 4(e), and 4(f), the expectation distributions are basically the same between the two topics. This result further declares that we controlled the variation of topic familiarity, so it has no significant impact on users' search expectations. However, the four task types have quite different impacts on users' search expectations, so we further differentiated the task types when analyzing the relationships between expectations and other factors.

Besides differentiating the task type, we also investigated the impacts of contextual factors on expectations at different query positions because the query positions influence users' expectations (shown in Fig. 5). We chose four query positions from each search session to represent the query sequence, including the first/end queries and the queries with peak/bottom values of expectations. We further analyzed the impacts of query position on expectations in the next subsection, and here we focused on the contextual

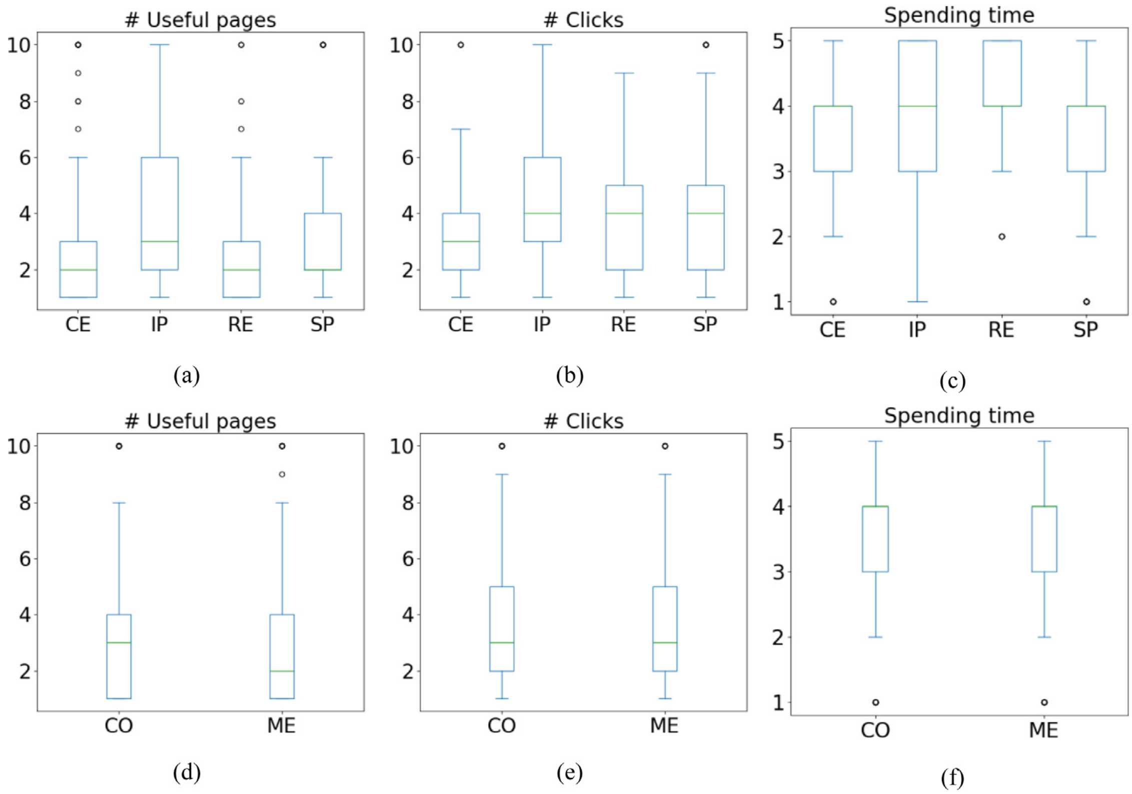


Fig. 4. Expectation distributions among task types and task topics.

Table 10
Kruskal–Wallis's H test on the previous experience factors and expectations among task types and between task topics.

Kruskal–Wallis's H	Experience/Familiarity	Difficulty	# Useful pages	# Clicks	Spending time
Task type	3.22	**39.43****	**28.84****	**26.97****	**18.78****
Task topic	**12.77****	6.05	1.41	0.01	0.05

Values in **boldface** indicate the statistical significance: corrected $p < 0.05$, ** corrected $p < 0.01$.

factors. To make the results concise, we only reported significant results for contextual factors in different task types or at different query positions if the contextual factor was not significant in general.

Table 11 are the correlation coefficients of previous experience and expectation features. We used participants' feedback on task topic familiarity, task type experience, perceived difficulty, and the year in college to represent their previous experience. Based on the results, the task topic familiarity has no significant relationship with all expectation features, which is consistent with the results in Table 10. For the task type experience, although there is no significant difference among task types, it has positive correlations with the search expectations of spending time. For # Useful pages, it is negatively related to task type experience for the task type SP but has no significant correlation among all tasks. The perceived difficulty negatively correlates with # Useful pages but has different impacts on # Clicks among task types. It has a positive correlation in task type CE but negative correlations in RE and IP. In addition, it positively correlates with spending time in task type SP, but no significant correlation in the four task types overall. these relationships indicate that users will have a lower expectation of the number of useful pages if they have more experience in the task type SP or higher perceived difficulty overall. Their expectation of spending time will increase with more task type experience overall or perceived difficulty in task type SP. These relationships indicate that previous experience may allow users to have a more conservative expectation of their search gain and effort by lowering the expectation of useful pages and raising the spending time. However, the perceived difficulty has opposite impacts on # Clicks, indicating adaptive search expectations and behaviors for different task types. For the year in college, although participants mentioned the year in college contributed to their search experience, it is not significantly related to any expectations here in individual task types or four task types together. This may be related to the small sample of participants or the difference in majors requiring different levels of search skills. One freshman participant mentioned he had developed much search experience during the research projects in high school. Therefore, it suggests recruiting more participants, controlling the major, or investigating the research-related search experience in further studies.

We also investigated the impacts of in-situ experience during the search on users' expectations and used the query order, feedback of the previous query (i.e., LastGain, LastEffort, and LastSat), and the result quality of the previous query (i.e., LastDCG/RBP/ERR,

Table 11
Correlation between previous experience and expectation.

Spearman's r	# Useful pages	# Clicks	Spending time
Task topic familiarity	−0.029	0.064	0.021
Task type experience	−0.083 (SP:**−0.390***)	0.108	**0.189****
Perceived difficulty	**−0.202****	−0.055 (CE:**0.247***, RE:**−0.368***, IP:**−0.380****)	0.044 (SP:**0.379***)
Year in college	−0.082	−0.039	−0.011

Results in parentheses are separated by task types;
Values in **boldface** indicate the statistical significance: * corrected $p < 0.05$, ** corrected $p < 0.01$.

Table 12
Correlation between in-situ search experience and expectation.

Spearman's r	# Useful pages	# Clicks	Spending time
In-situ search experience - continuity and feedback			
QueryOrder	**−0.232****	**−0.241****	**−0.180****
LastGain	0.028	0.004	−0.020
LastEffort	−0.041	−0.042	0.095
LastSat	0.012	−0.088	−0.150
In-situ search experience - search result quality of the previous query			
LastDCG	0.122	0.097 (End:**0.303***)	0.053 (Peak:**0.302***)
LastRBP	0.108	0.110 (End:**0.322***)	0.052 (Peak:**0.320***)
LastERR	0.085	0.031 (End:**0.302***)	0.002
LastMaxUse	0.055	0.037	0.002
LastAvgUse	0.012	−0.072	−0.128
LastUseNum	0.155	**0.218****	0.112

Results in parentheses are separated by task types or query positions;
Values in **boldface** indicate the statistical significance: * corrected $p < 0.05$, ** corrected $p < 0.01$.

and LastMaxUse/AvgUse/UseNum) to represent their in-situ search experience. We calculated the correlation coefficients of these factors and expectations using Spearman's r, and the results are listed in Table 12. According to the results, only QueryOrder has negative relationships with all the expectations of # Useful pages, # Clicks, and spending time, indicating that users issuing more queries might get more knowledge about the task and expect to find fewer but specific information with less time. For LastGain and LastEffort, we expected they would impact users' expectations, but the results show no significant relationships in general. Besides users' feedback on their experience, we then investigated how the search result quality of the previous query affect users' expectations. Based on the correlation results, for four task types and all query positions in general, only LastUseNum has a positive impact on the expectation of # Useful pages and # Clicks. Other search result quality features have impacts on expectations at specific query positions. For example, the high-quality results in the previous query could increase users' expectation of # Clicks at the end query. In addition, the higher quality results in the previous query might also prompt users to have higher peak expectations of spending time in a session. However, those features have no significant correlations with expectations in general and might not directly impact users' expectations. Nevertheless, the significant relationships at specific query positions indicate that the in-situ search experience might have impacts varying in query positions in a session. These impacts suggest new evaluation metrics and user models reflecting search expectations in different query conditions.

As the query order has significant relationships with all three expectations, we further investigated how the expectations change in the search session. Because of the varied session length, we chose four query positions (i.e., First/Peak/Bottom/End) queries and visualizes the expectation distributions in Fig. 5. As there are 65 short sessions (sessions with fewer than 3 queries), and we did not filter out those short sessions for data integrity, the four query positions might be overlapped in some sessions, and the peak/bottom queries could also be the first/end queries. Generally, the different expectations at the four query positions indicate that users' expectations are dynamic in the ongoing search session. Although the query order has negative correlations with the three expectations, the expectation trends are not monotonic decreasing. Instead, users' expectations fluctuate with peak or bottom values in a session probably because they expect to find more useful pages, have more clicks, or spend more time during the exploration process or decrease their expectation to search for specific information.

5.4. RQ3: In-situ expectation and search behaviors

After investigating the expectation features, and contextual factors for RQ1 and RQ2, we investigate the impact of users' expectations on their search behaviors to answer RQ3. Table 13 lists the behaviors affected by expectations. In addition, during the qualitative analysis, we found users' behaviors in the current query could also be affected if the results confirmed or disconfirmed their expectations in the previous query, so we involved the expectation-confirmation features (i.e., LastGain, LastEffort, and confirmation in the current query) in the analysis and results.

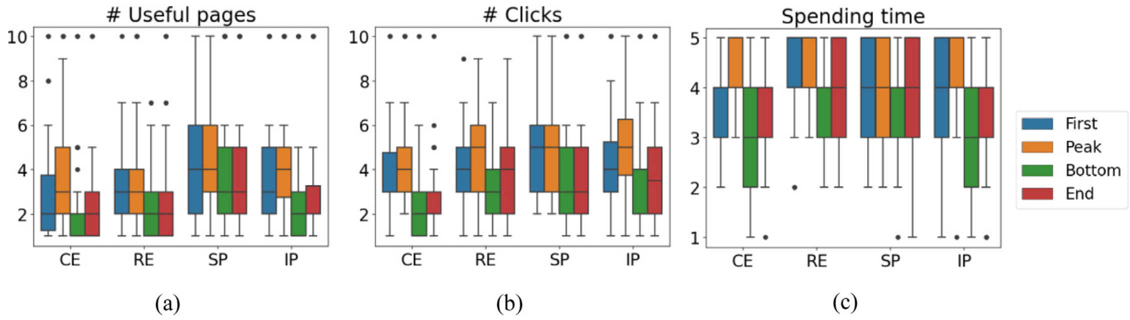


Fig. 5. Dynamic expectation states in the search session.

Table 13
Users' behaviors and satisfaction affected by expectation and/or expectation confirmation.

	Description	Example
Querying behavior	Users' behaviors in formulating the query, including their search intention, query reformation type, and query length.	"I found results that were better than I expected, so I only needed to use one (more) query to finish the task because it was more like finding information and then writing a paragraph from my head." [S04] "I was frustrated with the results, so that's why I switched up my (query) wording." [S06]
Browsing behavior	Users' behaviors on SERPs, including click and dwell time.	"I couldn't find enough information, which resulted in more clicks and more results, I'm trying to find what I'm looking for." [S22] "When I found out that there was better information than I expected, I spent more time looking through the search results that it gave me because I expected to find more useful information inside." [S04]
Satisfaction	Users' satisfaction with the query results.	"I did start to worry about there not being any information (about the fact). It was nice to find a very similar quotation in a very similar use of the fact, so I think it would be more satisfying than I would have previously thought." [S06] "I put unsatisfying because I searched really simple words, but I didn't find the information that I want, so I feel even if I put in more effort, nothing would pop up because I already searched so many (times)." [S11]

We categorized search behaviors into querying behaviors and browsing behaviors. These behaviors can be both affected by expectations of the current query and expectation confirmations of the previous query. In these results, we do not distinguish if the effects are from expectations or expectation confirmations because we do not investigate the order of the effects of expectation/confirmation and users' behaviors. For example, users' querying behaviors might be related to their expectation confirmation from the last query, and we collected users' expectations after they input the current query, but the cognitive process of forming the expectation might take place in advance and affect the querying behavior. In addition, we also extracted participants' satisfaction related to their expectations/confirmations and investigated the relationship.

We further investigated the impacts of expectation/confirmation features on users' behaviors. We extracted features of users' search behaviors, including querying behaviors (e.g., QueryLength, UniqueTerm), and browsing behaviors (e.g., ClickCount, ScrollDist, QueryDwellTime). We calculated the correlation coefficients of these behaviors and the expectations of the current query/confirmations of the previous query. We also investigate the relationship between satisfaction and expectations/confirmations of the current query. The results are listed in Table 14. For querying behaviors, QueryLength, UniqueTerm, and QuerySim are correlated with individual expectation features. For example, with a higher expectation of # Useful pages, users may submit shorter queries with fewer unique terms, and users may submit similar queries when they have a high expectation of # Clicks. Browsing behaviors, including mouse-based behaviors and time-based behaviors, are all positively correlated with the expectation features. Users with higher expectations will have more clicks, tend to click at lower ranks, and spend more time. In addition, they will choose the first click with more discretion (high TimeFirstClick) and browse the document thoroughly (high ScrollDist). These results indicate that users' expectations affect their most browsing behaviors and some specific querying behaviors. Users with a higher number of clicks and long dwell time are related to the expectation of higher useful pages. They might want to find more useful pages with a general query, more clicks, and more time. For expectations of higher # Click and spending time, users also spent more effort on clicking and reading during the query.

For the confirmations of the previous query, users submit queries with fewer unique terms when they have higher LastEffort. Although the confirmation features may not mainly influence users' search behaviors, the gain confirmation of the current query is highly correlated with users' query satisfaction. The effort confirmation of the current query also negatively affects users' query satisfaction. This is because users might be struggling and unsatisfied when they continue expecting to have a higher spending time and raising their efforts. These results also echo the findings regarding the effects of human expectation on the process and outcome of decision-making activities from previous studies (e.g. Ayaburi, Lee, & Maasberg, 2020; Cox & Fisher, 2004; Lankton, McKnight, & Thatcher, 2014; Liu & Shah, 2019).

Table 14
Correlation between user expectation and their behaviors and satisfaction.

Spearman's r	# Useful pages	# Clicks	Spending time	Gain confirmation[1]	Effort confirmation[1]
Querying behaviors					
QueryLength	**−0.139***	0.002	−0.055	0.009	−0.155
UniqueTerm	**−0.198****	−0.109	−0.124	0.016	**−0.185***
NewTerm	−0.137	−0.084	0.059	0.089	−0.036
QuerySim	0.125	**0.166***	−0.020	−0.044	0.039
Browsing behaviors - Mouse based					
ClickCount	**0.242****	**0.303****	**0.248****	0.115	0.061
Clicks@3	**0.151***	**0.176****	0.137	0.026	0.049
Clicks@5	**0.176****	**0.216****	**0.174****	0.091	0.081
Clicks@5+	**0.184****	**0.238****	**0.227****	0.020	0.031
ClickDepth	**0.192****	**0.272****	**0.221****	0.071	0.049
AvgClickRank	**0.160***	**0.229****	**0.182****	0.048	0.031
UniquePage	**0.237****	**0.306****	**0.246****	0.115	0.059
ScrollDist	**0.183****	**0.226****	**0.240****	0.082	0.056
Browsing behaviors - Dwell time based					
QueryDwellTime	**0.191****	**0.266****	**0.333****	0.085	0.070
TimeFirstClick	**0.154***	**0.193****	0.136	0.165	0.011
TimeLastClick	**0.225****	**0.301****	**0.271****	0.135	0.078
Current experience					
Satisfaction	0.013	−0.032	−0.103	**0.793**** *(current query)*	**−0.224**** *(current query)*

[1] Gain/effort confirmations of the previous query (i.e., LastGain/LastEffort) unless specified;
Values in **boldface** indicate the statistical significance: * corrected $p < 0.05$, ** corrected $p < 0.01$.

Table 15
Accepted hypotheses with features.

Hypothesis	Features
H1	All expectations
H2	Task type
H3	Task experience, Perceived difficulty
H4	QueryOrder, LastDCG/RBP/ERR, LastUseNum
H5a	QueryLength, UniqueTerm, QuerySim, All browsing behaviors
H5b	UniqueTerm
H6b	All expectation confirmations

6. Discussion

6.1. Discussion of statistical hypotheses and research questions

Table 15 summarizes the statistical results of the relationships between expectation features and contextual factors of task characteristics, previous experience, search experience of the last query, search behaviors, and satisfaction. Overall, from the analysis results of the interview data and the survey data, as well as the search behaviors of 60 participants, this study demonstrates users' search expectations are different among task types and associated with their previous experience, search behaviors, and satisfaction. These results help us better understand how users form expectations in the search and how the expectation affects their search behaviors. Knowing the effects of search expectations will improve modeling users' cognitive process in information retrieval and reduce the bias caused by users' in-situ changing expectations and references.

Regarding the three proposed RQs, we have the following answers:

RQ1: Characteristics of users' search experience. We defined three expectation features based on the assumptions of information gain and search effort and left open-ended questions for the participants to discover other expectations. Based on the qualitative analysis and the quantitative results, participants could estimate their expectations after getting a basic knowledge about the topic in the first query. More than half of the participants agree with us regarding the expectations of useful pages and spending time as the information gain and search efforts. For other expectation features, some participants preferred to use efficiency to replace the expectation of spending time, and participants without expectations of exact numbers of useful pages or clicks would seek the specific answer to the task question. These findings support the assumptions in previous studies of user modeling involving expectations and time (Moffat et al., 2017; Smucker & Clarke, 2012).

Our findings extend previous assumptions regarding expectations made in formal models. Users who are unfamiliar with the topic may not have specific expectations of useful pages or clicks, especially at the first query. This result indicates that users may not expect exact numbers of useful pages or clicks as their perceptions of future gain or loss for the search when they have less knowledge about the topic. Instead of exact numbers, their expectations may be a range (e.g., expect one useful page or several) or a propensity (e.g., expect more useful pages or fewer). After iterating searching the topic, they could get more experience and have clear expectations in a search session. These possibilities might mitigate or further reinforce users' irrational perceptions of future gain and efforts and call for the exploration of more expectation features in different search stages. For example, when we do user modeling involving search expectations, we need to consider the stages of the sessions where users are searching for an unfamiliar topic. Their expectation for the first query might not be specific and could be a ratio, ordinal, or categorical variable rather than a numeric variable. Then, in the following queries in the same session, their expectation range could be more specific and predictable, offering more room for user expectation modeling and proactive search assistance.

RQ2: Contextual factors related to users' search expectations. We extracted the contextual factors of users' search expectations and divided them into three categories: task characteristics, previous experience, and in-situ search experience during the search process. After finding the differences in expectations among task types, we further analyzed these factors in individual task types and different query positions. The results indicate that we controlled the variation in the task topics, and users' expectations are related to the search task type, previous experience, and experience in the current search task. In addition, we used different measures to represent the search result quality, including effectiveness metrics and usefulness score variants. However, these search quality measures are significantly correlated with users' expectations only in specific task types or query positions, indicating that the impact of search result quality on users' expectations has boundaries and is moderated by contextual factors. Specially, we considered the query order part of the in-situ search experience because the query order can represent users' continuity intention and search state in a search session. Although the query order negatively correlates with users' expectations, their expectations could reach peak or bottom points in the search session, indicating users' dynamic expectation states in the ongoing search task.

Overall, these contextual factors reveal how users form their expectations in search tasks based on their previous and in-situ search experiences. Our findings regarding these factors show how to design search tasks or user studies with better control of users' expectations and analyze user expectations with complex variables in naturalistic datasets. Although our study did not involve different user groups (e.g., age, educational level) included in previous user study research (Chevalier, Dommes, & Marquié, 2015), we identified these contextual factors and investigated their effects in the user group of undergraduate students. Further study can employ these factors and broadly explore their effects in general user groups. Furthermore, our findings indicate that users' queries should not be analyzed by the same ad-hoc evaluation method in the same session. As the search result quality affects user expectations at specific query positions, there might be a key query in a session that leads users to different search experiences. Our findings suggest that involving user expectations could help develop adaptive user models to identify the key queries and evaluate the search session organically.

RQ3: Users' search behaviors are related to their expectations. We investigated the relationships between user search behaviors, satisfaction, and expectations/confirmation. The results reveal that users' expectations mainly affect their browsing behaviors, including mouse-based and dwell time-based behaviors. As their querying behaviors relate to specific expectation features (e.g., # Useful pages-QueryLength/UniqueTerm and # Clicks-QuerySim), these expectations can represent or quantify users' search intentions. The confirmation effects also influence their behaviors on specific task types. In addition, users' satisfaction is also highly correlated with their gain confirmation. Querying, browsing behaviors, and satisfaction are fundamental activities and experiences in user search interactions.

Our findings on the relationships between in-situ expectations and user behaviors indicate that more complex search behaviors (e.g., clicking at lower ranks, continuous clicking/leaving the query, and continuous querying/leaving the search session) are closely associated with users' expectations (Backus et al., 2017, 2022). Findings on these connections encourage researchers to further investigate behavior-expectation interactions (e.g., clicks at lower ranks, time of first/last click) and predict expectation features. In addition, the dominant relationship between gain confirmations and satisfaction identified in our study echoes the findings from previous studies on the expectation confirmation model (Cox & Fisher, 2004, 2009; Lankton & McKnight, 2012), suggesting the potential of ECT framework in predicting users' search satisfaction in task-based sessions.

6.2. Summary of contributions and implications

Inspired by the role of expectation in the human cognitive process, this research was conducted to investigate the role and impacts of in-situ search expectations. We did a controlled laboratory user study to collect direct evidence on how users form different expectations regarding information gain and cost and how those expectations affect their search behaviors. Taking a step forward from previous research on pre-search task-related anticipations, our study explored multiple contextual factors that affect users' in-situ expectations during sessions, and the results verify several of previous assumptions on expectation variations implicitly made in formal user models (Moffat et al., 2017). In addition, our study enhances the understanding of user expectation by moving beyond pre-search fixed perception measures and investigates the relationship between the variations in users' in-situ expectations and changes in their search behaviors and satisfaction (Liu & Shah, 2019). Findings on the changes and effects of in-situ expectation also extend our previous research on the transitions of task states and search tactics (cf. Liu & Shah, 2022) and characterize the hidden dynamics in complex search tasks over a new dimension.

Furthermore, this research examines the feasibility and importance of expanding the expectation confirmation theory to a new application context, the field of information retrieval (Cox & Fisher, 2009; Venkatesh & Goyal, 2010). At the theoretical level, findings in this research advance our knowledge about how users form their expectations *during* information search and enhance our understanding of the impact of in-situ user expectations on users' search decision-making activities in searching unfamiliar topics, at specific query positions, and in browsing results at lower ranks, and impacts on evaluating the search experience. At the practical level, those findings allow us to validate and improve existing user models in information seeking, design *expectation-aware* evaluation metrics, and incorporate user expectation as a key cognitive factor in personalizing user-centered IR algorithms (e.g., result ranking, query term recommendation, adaptive search interface layouts) and may allow systems to predict users' search tactics and interaction experience more accurately in real-time information search.

6.3. Limitations and future research

Our study has several limitations, including limited expectation features, controlled user groups, search tasks and environment, as well as limited search interaction signals, and thus calls for further research efforts in this direction. First, regarding expectation features, we can further investigate more features and dimensions of in-situ expectations, especially the ones that may be closer to users' actual perceptions (e.g., cognitive load, costs of browsing and mouse movement, eye movement and fixation analysis). It would also be useful to investigate the role of in-situ expectations under a broader range of task types or combinations of search task dimensions (cf. Liu, 2021), and further examine the possible associations between expectation variations and behavioral effects with certain search task features. Also, since *learning* is a key aspect and product of Web search, measuring users' in-situ expectations and progress in learning during search sessions can help further enhance our understanding of users' search decisions (e.g., query reformulation, click and examination, dwell/reading time, search stopping and query abandonment) (Liu & Jung, 2021; Rieh, Collins-Thompson, Hansen, & Lee, 2016; Urgo & Arguello, 2022). In addition, we can further conduct large-scale user studies or naturalistic studies to investigate the role of search expectations for broader, more general user groups in more search tasks and problematic scenarios. With respect to characterizing search interactions, some search interaction signals could not be captured due to the limitations of the study procedure and technical design. These behaviors are common in real-time search scenarios but restricted in this study, including submitting multiple queries and browsing concurrent SERPs at the same time, opening the result page in the background, and changing queries immediately. Besides, to collect direct evidence about in-situ user expectations, the participants' search processes were inevitably interrupted to a different extent by the pre/post-search questionnaires. We will investigate more search behaviors related to in-situ expectations and develop methods and metrics to quantify these expectations accurately for user modeling and expectation-aware search evaluation. Finally, we will further investigate the effects of expectation/confirmation on users' search behaviors and satisfaction at the whole session level. Based on our work and results, we aim to build and evaluate expectation-aware user models and leverage the knowledge about user expectations as well as other in-situ reference points (Chen et al., 2022; Liu & Han, 2020) in enhancing user-centered search systems and building adaptive search recommendations.

7. Conclusion

Our work empirically examined users' search expectations in Web search and collected direct evidence about in-situ expected gains and efforts as well as their effects. In addition, we systematically analyzed the contextual factors affecting user expectations and the effects of user expectations on different aspects of search interactions. Our findings advance the knowledge of how users form and change their in-situ search expectations and how their dynamic search expectations affect their search behavior and experience at different moments of search sessions. Knowledge learned through this study can help us better design user models and user-oriented IR systems. For example, we can incorporate expectation features into machine learning models to predict users' search tactics and interaction experience more accurately in real-time information searches. Furthermore, the scales and tools we developed for collecting in-situ user expectations could be further standardized and generalized for supporting user-centered search evaluation and the reuse of interactive information retrieval resources (e.g., questionnaires, search datasets, evaluation metrics). More broadly, our study empirically examined the role and impacts of individuals' in-situ expectations in a new context and demonstrates the value (highlighted in several existing perspective works, e.g. Azzopardi, 2021; Liu, 2022) of integrating the knowledge from cognitive psychology (especially regarding human expectations and biases) with the models, techniques, and measures in IR experiments. Our interdisciplinary approach not only confirms the classical behavioral theory regarding expectations in a novel and ubiquitous context but also opens new paths for investigating and modeling boundedly rational users in information-seeking and search episodes. Although we acknowledged some limitations exist in study design and data analysis, our findings shed light on the role of in-situ search expectation and pave the way towards further improving user studies and mixed method analysis concerning user expectations and related cognitive biases in online information seeking and retrieval.

CRediT authorship contribution statement

Ben Wang: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Review & editing. **Jiqun Liu:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Data availability

The raw data that has been used is confidential. Anonymized data might be available on request after project completion.

Acknowledgments

This work is supported by the National Science Foundation (NSF) grant IIS-2106152 and a grant from the Seed Funding Program of the Data Institute for Societal Challenges, the University of Oklahoma. We would like to express our sincere gratitude to the anonymous reviewers for their valuable comments and suggestions, which greatly improved the quality of this paper. We are also grateful to the editor for coordinating the review process. We also want to thank our undergraduate assistant, Zeming Hu, who helped us run some of the user study sessions and analyze data.

References

Abeler, J., Falk, A., Goette, L., & Huffman, D. (2011). Reference points and effort provision. *American Economic Review*, *101*(2), 470–492.

Agosto, D. E. (2002). Bounded rationality and satisficing in young people's web-based decision making. *Journal of the American Society for Information Science and Technology*, *53*(1), 16–27.

Al-Maskari, A., & Sanderson, M. (2010). A review of factors influencing user satisfaction in information retrieval. *Journal of the American Society for Information Science and Technology*, *61*(5), 859–868.

Ayaburi, E. W., Lee, J., & Maasberg, M. (2020). Understanding crowdsourcing contest fitness strategic decision factors and performance: An expectation-confirmation theory perspective. *Information Systems Frontiers*, *22*(5), 1227–1240.

Azzopardi, L. (2011). The economics in interactive information retrieval. In *Proceedings of the 34th international ACM SIGIR conference on research and development in information retrieval* (pp. 15–24).

Azzopardi, L. (2014). Modelling interaction with economic models of search. In *Proceedings of the 37th international ACM SIGIR conference on research & development in information retrieval* (pp. 3–12).

Azzopardi, L. (2021). Cognitive biases in search: a review and reflection of cognitive biases in information retrieval. In *Proceedings of the 2021 conference on human information interaction and retrieval* (pp. 27–37).

Azzopardi, L., & Zuccon, G. (2016). An analysis of the cost and benefit of search interactions. In *Proceedings of the 2016 ACM international conference on the theory of information retrieval* (pp. 59–68).

Azzopardi, L., & Zuccon, G. (2019). Building economic models of human computer interaction. In *Extended abstracts of the 2019 CHI conference on human factors in computing systems* (pp. 1–4).

Backus, M., Blake, T., Masterov, D. V., & Tadelis, S. (2017). *Expectation, disappointment, and exit: Reference point formation in a marketplace*: *Technical report*, National Bureau of Economic Research.

Backus, M., Blake, T., Masterov, D., & Tadelis, S. (2022). Expectation, disappointment, and exit: Evidence on reference point formation from an online marketplace. *Journal of the European Economic Association*, *20*(1), 116–149.

Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review*.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *57*(1), 289–300.

Bhattacherjee, A. (2001). Understanding information systems continuance: An expectation-confirmation model. *MIS Quarterly*, 351–370.

Cancho, R. F. I., & Solé, R. V. (2003). Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences*, *100*(3), 788–791.

Caraban, A., Karapanos, E., Gonçalves, D., & Campos, P. (2019). 23 ways to nudge: A review of technology-mediated nudging in human-computer interaction. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1–15).

Chapelle, O., Metlzer, D., Zhang, Y., & Grinspan, P. (2009). Expected reciprocal rank for graded relevance. In *Proceedings of the 18th ACM conference on information and knowledge management* (pp. 621–630).

Chen, J., Mao, J., Liu, Y., Zhang, F., Zhang, M., & Ma, S. (2021). Towards a better understanding of query reformulation behavior in web search. In *Proceedings of the web conference 2021* (pp. 743–755).

Chen, S. C., Yen, D. C., & Hwang, M. I. (2012). Factors influencing the continuance intention to the usage of web 2.0: An empirical study. *Computers in Human Behavior*, *28*(3), 933–941.

Chen, N., Zhang, F., & Sakai, T. (2022). Constructing better evaluation metrics by incorporating the anchoring effect into the user model. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*.

Chevalier, A., Dommes, A., & Marquié, J. C. (2015). Strategy and accuracy during information search on the web: Effects of age and complexity of the search questions. *Computers in Human Behavior*, *53*, 305–315.

Cockburn, A., Quinn, P., Gutwin, C., Chen, Z., & Suwanaposee, P. (2022). Probability weighting in interactive decisions: Evidence for overuse of bad assistance, underuse of good assistance. In *CHI conference on human factors in computing systems* (pp. 1–12).

Cole, M. J., Hendahewa, C., Belkin, N. J., & Shah, C. (2015). User activity patterns during information search. *ACM Transactions on Information Systems (TOIS)*, *33*(1), 1–39.

Cox, A., & Fisher, M. (2004). Expectation as a mediator of user satisfaction. In *WWW conference workshop on web search effectiveness*. Citeseer.

Cox, A., & Fisher, M. (2009). An expectation-based model of web search behaviour. In *2009 second international conferences on advances in computer-human interactions* (pp. 49–56). IEEE.

Dupret, G. (2011). Discounted cumulative gain and user decision models. In *International symposium on string processing and information retrieval* (pp. 2–13). Springer.

Evans, J. S. B. (2020). Bounded rationality, reasoning and dual processing. In *Routledge handbook of bounded rationality* (pp. 185–195). Routledge.

Franque, F. B., Oliveira, T., Tam, C., & de Oliveira Santini, F. (2020). A meta-analysis of the quantitative studies in continuance intention to use an information system. *Internet Research*.

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, *40*(1), 35–42.

Gomroki, G., Behzadi, H., Fattahi, R., & Salehi Fadardi, J. (2021). Identifying effective cognitive biases in information retrieval. *Journal of Information Science*, Article 01655515211001777.

Hergovich, A., Schott, R., & Burger, C. (2010). Biased evaluation of abstracts depending on topic and conclusion: Further evidence of a confirmation bias within scientific psychology. *Current Psychology*, *29*(3), 188–209.

Hollander, M., Wolfe, D. A., & Chicken, E. (2013). *Nonparametric statistical methods, vol. 751*. John Wiley & Sons.

Hossain, M. A., & Quaddus, M. (2012). Expectation–confirmation theory in information system research: A review and analysis. *Information Systems Theory*, 441–469.

Huang, Y. M. (2019). Examining students' continued use of desktop services: Perspectives from expectation-confirmation and social influence. *Computers in Human Behavior*, *96*, 23–31.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, *93*(5), 1449–1475.

Kelly, D., et al. (2009). Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends® in Information Retrieval*, *3*(1–2), 1–224.

Lankton, N. K., & McKnight, H. D. (2012). Examining two expectation disconfirmation theory models: Assimilation and asymmetry effects. *Journal of the Association for Information Systems*, *13*(2), 1.

Lankton, N., McKnight, D. H., & Thatcher, J. B. (2014). Incorporating trust-in-technology into expectation disconfirmation theory. *The Journal of Strategic Information Systems*, *23*(2), 128–145.

Lau, A. Y. S., & Coiera, E. W. (2007). Do people experience cognitive biases while searching for information? *Journal of the American Medical Informatics Association*, *14*(5), 599–608.

Li, Y. (2009). Exploring the relationships between work task and search task in information search. *Journal of the American Society for Information Science and Technology*, *60*(2), 275–291.

Li, Y., & Belkin, N. J. (2008). A faceted approach to conceptualizing tasks in information seeking. *Information Processing & Management*, *44*(6), 1822–1837.

Lieder, F., Griffiths, T. L., M Huys, Q. J., & Goodman, N. D. (2018). The anchoring bias reflects rational use of cognitive resources. *Psychonomic Bulletin & Review*, *25*(1), 322–349.
Liu, J. (2021). Deconstructing search tasks in interactive information retrieval: A systematic review of task dimensions and predictors. *Information Processing & Management*, *58*(3), Article 102522.
Liu, J. (2022). Toward cranfield-inspired reusability assessment in interactive information retrieval evaluation. *Information Processing & Management*, *59*(5), Article 103007.
Liu, J., Cole, M. J., Liu, C., Bierig, R., Gwizdka, J., Belkin, N. J., et al. (2010). Search behaviors in different task types. In *Proceedings of the 10th annual joint conference on digital libraries* (pp. 69–78).
Liu, J., & Han, F. (2020). Investigating reference dependence effects on user search interaction and satisfaction: A behavioral economics perspective. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval* (pp. 1141–1150).
Liu, J., & Jung, Y. J. (2021). Interest development, knowledge learning, and interactive IR: toward a state-based approach to search as learning. In *Proceedings of the 2021 conference on human information interaction and retrieval* (pp. 239–248).
Liu, J., Mitsui, M., Belkin, N. J., & Shah, C. (2019). Task, information seeking intentions, and user behavior: Toward a multi-level understanding of web search. In *Proceedings of the 2019 conference on human information interaction and retrieval* (pp. 123–132).
Liu, J., & Shah, C. (2019). Investigating the impacts of expectation disconfirmation on web search. In *Proceedings of the 2019 conference on human information interaction and retrieval* (pp. 319–323).
Liu, J., & Shah, C. (2022). Leveraging user interaction signals and task state information in adaptively optimizing usefulness-oriented search sessions. In *Proceedings of the 22nd ACM/IEEE joint conference on digital libraries* (pp. 1–11).
Mao, J., Liu, Y., Zhou, K., Nie, J. Y., Song, J., Zhang, M., et al. (2016). When does relevance mean usefulness and user satisfaction in web search? In *Proceedings of the 39th international ACM SIGIR conference on research and development in information retrieval* (pp. 463–472).
McGregor, M., Azzopardi, L., & Halvey, M. (2021). Untangling cost, effort, and load in information seeking and retrieval. In *Proceedings of the 2021 conference on human information interaction and retrieval* (pp. 151–161).
Mendel, R., Traut-Mattausch, E., Jonas, E., Leucht, S., Kane, J. M., Maino, K., et al. (2011). Confirmation bias: why psychiatrists stick to wrong preliminary diagnoses. *Psychological Medicine*, *41*(12), 2651–2659.
Moffat, A., Bailey, P., Scholer, F., & Thomas, P. (2017). Incorporating user expectations and behavior into the measurement of search effectiveness. *ACM Transactions on Information Systems (TOIS)*, *35*(3), 1–38.
Moffat, A., Thomas, P., & Scholer, F. (2013). Users versus models: What observation tells us about effectiveness metrics. In *Proceedings of the 22nd ACM international conference on information & knowledge management* (pp. 659–668).
Moffat, A., & Zobel, J. (2008). Rank-biased precision for measurement of retrieval effectiveness. *ACM Transactions on Information Systems (TOIS)*, *27*(1), 1–27.
Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review*, *106*(4), 643.
Piwowarski, B., Dupret, G., & Lalmas, M. (2012). *Beyond cumulated gain and average precision: Including willingness and expectation in the user model* (Ph.D. thesis), UPMC Univ Paris 06, CNRS, LIP6 UMR 7606: Sorbonne Universités.
Qazi, A., Tamjidyamcholo, A., Raj, R. G., Hardaker, G., & Standing, C. (2017). Assessing consumers' satisfaction and expectations through online opinions: Expectation and disconfirmation approach. *Computers in Human Behavior*, *75*, 450–460.
Rha, E. Y., Mitsui, M., Belkin, N. J., & Shah, C. (2016). Exploring the relationships between search intentions and query reformulations. *Proceedings of the Association for Information Science and Technology*, *53*(1), 1–9.
Rieh, S. Y., Collins-Thompson, K., Hansen, P., & Lee, H. J. (2016). Towards searching as a learning process: A review of current perspectives and future directions. *Journal of Information Science*, *42*(1), 19–34.
Sarkar, S., Mitsui, M., Liu, J., & Shah, C. (2020). Implicit information need as explicit problems, help, and behavioral signals. *Information Processing & Management*, *57*(2), Article 102069.
Simon, H. A. (2000). Bounded rationality in social science: Today and tomorrow. *Mind & Society*, *1*(1), 25–39.
Smucker, M. D., & Clarke, C. L. (2012). Time-based calibration of effectiveness measures. In *Proceedings of the 35th international ACM SIGIR conference on research and development in information retrieval* (pp. 95–104).
Steelman, Z. R., & Soror, A. A. (2017). Why do you keep doing that? The biasing effects of mental states on IT continued usage intentions. *Computers in Human Behavior*, *73*, 209–223.
Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics*, *106*(4), 1039–1061.
Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323.
Urgo, K., & Arguello, J. (2022). Learning assessments in search-as-learning: A survey of prior work and opportunities for future research. *Information Processing & Management*, *59*(2), Article 102821.
Venkatesh, V., & Goyal, S. (2010). Expectation disconfirmation and technology adoption: polynomial modeling and response surface analysis. *MIS Quarterly*, 281–303.
White, R. (2013). Beliefs and biases in web search. In *Proceedings of the 36th international ACM SIGIR conference on research and development in information retrieval* (pp. 3–12).
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.
Zhang, F., Liu, Y., Li, X., Zhang, M., Xu, Y., & Ma, S. (2017). Evaluating web search with a bejeweled player model. In *Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval* (pp. 425–434).
Zhang, F., Liu, Y., Mao, J., Zhang, M., & Ma, S. (2020). User behavior modeling for web search evaluation. *AI Open*, *1*, 40–56.
Zhang, F., Mao, J., Liu, Y., Xie, X., Ma, W., Zhang, M., et al. (2020). Models versus satisfaction: Towards a better understanding of evaluation metrics. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval* (pp. 379–388).