

# Investigating the Impacts of Expectation Disconfirmation on Web Search

Jiqun Liu

School of Communication & Information  
Rutgers University  
New Brunswick, NJ, USA, 08901  
jl2033@scarletmail.rutgers.edu

## ABSTRACT

Expectation disconfirmation refers to the situation where a user's perceived performance of a system disconfirms her original expectation. Previous information systems studies have demonstrated that expectation disconfirmation can significantly affect a system user's behavior and experience. Inspired by this finding, we go beyond the traditional approach that focuses on the final post-search perception and study the expectation disconfirmation problem in Web search. Our study investigates *task difficulty expectation disconfirmation* and demonstrates that: (1) unexpectedly difficult task can significantly decrease a user's perceived level of search success and increase the perceived time pressure; (2) the size and direction of task difficulty expectation disconfirmation are significantly associated with Web search behavior; (3) it is possible to predict the state of expectation disconfirmation (especially the negative, unexpectedly difficult cases) based on search behavioral features. This study demonstrates the value of integrating expectation disconfirmation approach with interactive IR research and thus may encourage future researchers to further explore the effects of other aspects of users' expectations and post-search perceptions.

## CCS CONCEPTS

• Information systems → Users and interactive retrieval.

## KEYWORDS

Expectation disconfirmation; task difficulty; interactive IR

### ACM Reference Format:

Jiqun Liu and Chirag Shah. 2019. Investigating the Impacts of Expectation Disconfirmation on Web Search. In *2019 Conference on Human Information Interaction and Retrieval (CHIIR '19), March 10–14, 2019, Glasgow, United Kingdom*. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3295750.3298959>

## 1 INTRODUCTION

When interacting with information retrieval (IR) systems, users often engage in a series of information searching activities with various goals and expectations which emerge from characteristics of

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHIIR '19, March 10–14, 2019, Glasgow, United Kingdom

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6025-8/19/03...\$15.00

<https://doi.org/10.1145/3295750.3298959>

Chirag Shah

School of Communication & Information  
Rutgers University  
New Brunswick, NJ, USA, 08901  
chirags@rutgers.edu

their search tasks at hand. As a result, users' search behaviors often vary significantly across tasks of different types [11]. In particular, users tend to work harder (e.g., formulate more queries, visit more web pages) and seek available help(s) when they encounter an unexpected difficult task [2, 13]. To embrace a deeper understanding of task-based information searching, we need to investigate both users' expectations (prior to search) and perceptions (after search) of search task features (cf. [10]), especially task difficulty.

The size and direction of the discrepancy between pre-use expectation and post-use perception has been demonstrated to be predictive of information systems users' satisfaction and continuance intention [3, 6, 15]. Specifically, according to the Expectation Confirmation Theory (ECT) [15], when a system outperforms the user's original expectations (which means the expectation disconfirmation is *positive*), the user is more likely to be satisfied with the performance of the system. In contrast, when the expectation disconfirmation is *negative*, the user is more likely to be dissatisfied and thus may discontinue his or her usage of the system or affordance. In the context of Web search, the framework of ECT may serve as a new approach towards explaining Web search behavior.

There are several task-related features that are closely related to users' expectations and perceptions (e.g., task difficulty, topic familiarity, task familiarity). Among these subjective task features, task difficulty is arguably one of the most widely studied features in task-based information seeking and retrieval research (e.g. [1, 2, 8, 12]). It is worth noting that in the context of search, the difficulty of a task is closely associated with *the difficulty of using an IR system(s) to complete the task*. Specifically, for example, finding useful knowledge for writing an economics research paper could be a difficult intellectual task if the person only have a regular search engine. However, this task may not be difficult anymore if the person could get supports from an advanced retrieval system dedicated to economics literature. Thus, both the expectation and perception of task difficulty are partially determined by the system that can be used in performing the task. In this sense, employing the ECT approach can help reveal the connections among task difficulty expectation, task difficulty perception, and search interactions.

Most of the existing research on the behavioral effect of task difficulty mainly focuses on the final state of perceived task difficulty (i.e. post-search measurement of difficulty) in their analyses (e.g., [11–13]). In the light of the ECT framework and the associated findings from empirical studies, we take a step forward by exploring the behavioral impacts of users' expectation-perception discrepancies with respect to task difficulty (i.e.,  $\Delta$  task difficulty) in Web search. Based on the existing measurements employed in

ECT studies [3], in this study, we measured task difficulty expectation disconfirmation using the difference between post-search measurement of perceived task difficulty and pre-search measurement of expected task difficulty (given the task description) (i.e.  $\Delta$  task difficulty = post-search task difficulty - pre-search task difficulty).

In addition to the evidences from ECT-based studies, there are two reasons that motivated us to examine  $\Delta$  task difficulty in the context of Web search: (1)  $\Delta$  task difficulty may serve as a (simplified) representation of the extent to which a user is mentally prepared for a potentially difficult task. Specifically, if  $\Delta$  task difficulty  $< 0$ , it may suggest that the user is mentally well-prepared and may perform well in search. However, if  $\Delta$  task difficulty  $\geq 0$ , then the user may be barely prepared or not mentally prepared at all for a potentially difficult task. As a result, the user may need to break his or her established search strategies, work harder on the task, and seek additional supports from the system. (2) For recommendations design, it is critical to predict the negative states of task difficulty expectation disconfirmation (in this case,  $\Delta$  task difficulty  $\geq 0$ ) based on users' search behaviors and provide appropriate recommendations to support users' information searching activities. In this study, we seek to answer three research questions:

- **RQ1:** What are the effects of task difficulty expectation disconfirmation ( $\Delta$  task difficulty) on users' perceived levels of search success and time pressure?
- **RQ2:** How is a user's search behavior associated with task difficulty expectation disconfirmation?
- **RQ3:** To what extent can we predict the state of task difficulty expectation disconfirmation from search behaviors?

The study designed to answer the RQs above has several contributions: (1) It demonstrates that task difficulty expectation disconfirmation can significantly affect a user's search experience and is closely associated with multiple aspects of Web search behavior; (2) It indicates the predictive power of search behavior models in predicting the state of task difficulty expectation disconfirmation and thus can help IR systems determine the appropriate timing and situation for offering recommendations and supports; (3) It highlights the value of integrating the ECT framework with interactive IR research and may encourage future researchers to further explore other aspects of users' expectations and perceptions of systems performance, such as perceived knowledge gain and ease of use. The study design is explained in the following section.

## 2 METHOD

To answer the aforementioned RQs, we collected data through a controlled lab study where we assigned two search tasks for each of the 40 participants and measured their perceptions of task difficulty and search experience in pre- and post-search questionnaires. The behavioral data were collected via a browser plugin.

**Search tasks and participants.** To effectively control the task context, we designed four work tasks within the domain of journalism: Copy editing (CPE), Story pitch (STP), Relationships (REL), and Interview preparation (INT). Each work task contextualized a corresponding search task. These tasks were constructed based on a modification of Li and Belkin [10]'s faceted classification scheme and the facet values of the four tasks were varied as indicated in Table 1. Also, we recruited 40 undergraduate students majoring

**Table 1: Task characteristics.**

Work task	Search task	Goal	Product
CPE	find authoritative webpages to confirm/disconfirm a given statement.	Specific	Factual
STP	find webpages that contain interesting facts about a given topic.	Amorphous	Factual
REL	find webpages that explain the relationships between two given facts.	Amorphous	Intellectual
INT	identify authoritative webpages and relevant facts about an interviewee.	Amorphous	Intellectual

in journalism from a major U.S. research university as study participants. To ensure that the participants have similar levels of familiarity with the assigned journalism tasks, we only selected upper-division undergraduate students who have taken either one journalism writing or reporting course. More detailed descriptions regarding the tasks and participants are reported in [16].

**Study procedure.** The controlled lab study consisted of three sessions: pre-search survey session, search session, and post-search survey and interview session. In the pre-search survey session, participants were asked to complete a demographic questionnaire and a task-based questionnaire which asks questions about their work task familiarity, topic familiarity, and expected task difficulty. After that, each participants were assigned two search tasks and they had up to 20 minutes to complete each search task. Participants' search actions (e.g., timestamp, query, URLs, page type, clickthrough) were recorded by a Firefox browser plugin and Morae. The task types were rotated for different participants. After completing the task-based search session, participants were asked to answer questions about their perceived task difficulty, level of search success, and time pressure. The questions and scales corresponding to different task-related features (from both pre- and post-search questionnaires) are provided in the Table 2. Also, we invited participants to participate in a post-search semi-structured individual interview where we asked them to further elaborate on their search strategies and experiences.

**Data analysis.** Given the three RQs proposed above, our data analysis can be segmented into three parts: To answer RQ1 (the effects of  $\Delta$  task difficulty on search experience), we regress perceived level of search success and time pressure (time condition) respectively on  $\Delta$  task difficulty. We also include task familiarity and topic familiarity in the regression models as controlled variables, aiming to clear out the potential mixed effects from these two contextual variables. Since both search success and time condition were measured using Likert scales (which means they are ordinal variables), we use *Ordered Probit Model* for regression analyses [4]. To answer RQ2 (the relationships between search behavior and  $\Delta$  task difficulty), we conduct *Kendall Rank Correlation Analysis* on each behavior- $\Delta$  task difficulty pairs. Specifically, we include three categories of search behavior features that have been applied in a variety of task-based IR studies (e.g., [5, 11, 14]), including:

- **Query behavior:** query length, query reformulation (generalization, specialization, repeat, new, word substitution).
- **Clicking and browsing:** # all actions (query+click+bookmark), # clicks, # unique URLs, # content pages visited (including repeatedly visited pages).

**Table 2: Task-related features**

Variables	Questions and Scales
Task familiarity	How much experience do you have with this kind of assignment? 1 (not at all)-7 (extremely) - pre-search
Topic familiarity	How familiar are you with the topic of this assignment? 1 (not at all) - 7 (extremely) - pre-search
Expected difficulty	(Given the task description) How difficult do you think it will be to find the information for this assignment? 1 (not at all) - 7 (extremely) - pre-search
Perceived difficulty	How difficult was it to find the information you need for this assignment 1 (not at all) - 7 (extremely) - post-search
Search success	How successful do you think you were in gathering the information to complete this assignment 1 (not at all) - 7 (extremely) - post-search
Time condition	Did you have enough time to complete the assignment successfully? 1 (far too little), 2 (too little), 3 (barely enough), 4 (enough), 5 (more than enough) - post-search

- **Dwell time and display time features:** dwell time on all content pages, mean dwell time on each content page, dwell time before saving the first useful web page, dwell time on search engine result page (SERP), display time on all content pages, mean display time on each content page.

Note that regarding dwell time and display time features, dwell time on a content page refers to the amount of time that a user actually spent on reading the page. In contrast, display time measures how long a page has been kept open during web search [5, 14].

To answer RQ3 (predicting the direction of  $\Delta$  task difficulty), we consider two different situations: positive state ( $\Delta$  task difficulty  $<0$ ) and negative state ( $\Delta$  task difficulty  $\geq 0$ ) of expectation disconfirmation. Then, we employ multiple classifiers (e.g., Logistic regression, Naive Bayes, Random forest, SVM) to predict and differentiate these two states based on the behavioral features identified above.

### 3 RESULTS

In the controlled lab study, we collected search behavior data from 693 query segments that were generated in 80 task/search sessions by 40 participants. The length of a session ranges from 2 to 29 queries, with an average of 8.66 queries. All of our analyses are conducted at session level and the results are explained as follows.

**RQ1: the effects of  $\Delta$  task difficulty on users' perceptions of search success and time pressure.** As it is shown in Table 3, when  $\Delta$  task difficulty (post-search task difficulty - pre-search task difficulty) increased, participants perceived a higher level of task difficulty which might significantly surpass their original expected level of difficulty. Consequently, their perceived level of search success decreased and the perceived time pressure significantly increased given the predefined time limit. It is worth noting that these impacts are still statistically significant when we include expected task difficulty, topic familiarity, task familiarity, and task type as controlled variables in the ordered probit regression model, indicating that the negative effects of  $\Delta$  task difficulty on the perceptions of search success and time pressure are robust and it is not very likely that these effects are caused by other task features.

To answer RQ1, We define search success and time condition as dependent variables and run two separate ordered probit regression models (see Table 3). The result presented in Table 3 echoes the findings in previous ECT-based information systems studies (e.g., [3, 15]) as it empirically confirms the effects of task difficulty expectation disconfirmation on search experience in Web search. RQ2 takes a step forward by exploring the connections between  $\Delta$  task difficulty and Web search behaviors.

**RQ2: the relationships between Web search behaviors and  $\Delta$  task difficulty.** To understand why task difficulty expectation

**Table 3: Ordered probit regression model: odds ratios.**

	Search success	Time condition
$\Delta$ task difficulty	<b>.470***</b>	<b>.398***</b>
expected difficulty	<b>.463***</b>	<b>.482***</b>
goal: amorphous	2.046	2.907
product: intellectual	<b>.277*</b>	.387
topic familiarity	.960	.997
task familiarity	1.188	1.038

*Note:* \* $<.05$ , \*\* $<.01$ , \*\*\* $<.001$ . Statistically significant odds ratios are boldfaced. The first column presents all independent variables. Odds ratio $<1$  indicates negative effect; odds ratio $>1$  indicates positive effect.

**Table 4: Kendall rank correlation analysis: Kendall's tau.**

Behavioral features	Within-session/mean	Whole-session/sum
# all actions	<b>.248**</b>	<b>.194*</b>
# clicks	.110	<b>.203*</b>
# unique URLs	-.026	.134
# pages visited	.056	.144
dwell time on content page	.037	<b>.331***</b>
dwellcontent to first save	.039	<b>.271***</b>
dwell time on SERP	.022	.039
display time content page	<b>.218**</b>	<b>.271**</b>
query refo: generalization	<b>.234**</b>	<b>.236**</b>
query refo: specialization	.002	.149
query refo: new	-.043	-.010
query refo: repeat	-.004	-.077
query refo: wordSubstitute	-.011	-.025

*Note:* \* $<.05$ , \*\* $<.01$ , \*\*\* $<.001$ . Statistically significant correlations (measured by Kendall's tau coefficient) are boldfaced.

disconfirmation can significantly affect search experience, it is critical to investigate the relationship between Web search behavior and  $\Delta$  task difficulty and to understand how participants reacted to different levels of  $\Delta$  task difficulty at behavioral level.

When conducting behavioral analysis at task/session level, some existing studies only compute the average value (per query segment) for each behavioral feature, such as average dwell time on each query segment, average number of content pages visited (e.g. [14]). In this paper, query segment is defined as a *search session segment which starts from one issued query and ends at the next query*. To obtain a more comprehensive picture of participants' Web search behaviors, this study follows Liu et al [12]'s approach and calculates both whole-session (i.e. sum/total values for the entire task session) and within-session (i.e. average values per query segment) for each behavioral feature introduced in the previous Method section.

According to the results in Table 4, in general, when  $\Delta$  task difficulty increased and the expectation disconfirmation tended to

be negative, participants usually worked significantly harder on their search tasks by increasing their total actions, clicking more pages, and spending more time on reading content pages. This may be because when participants encountered an unexpected difficult task, they tended to increase their resource input (e.g., time, cognitive resource) in order to maintain their regular or average level of knowledge gain or search performance. Another possible reason is that participants were struggling to find satisfactory search strategies and results when the tasks were more difficult than they thought. Consequently, they had to be more active in searching information and trying various alternative search tactics.

In addition to the general trend discussed above, participants also formulated significantly more broad, generalized queries when  $\Delta$  task difficulty increased. This may be because when encountering an unexpected difficult task, participants often found it difficult to formulate longer, more specific queries to accurately locate useful information. In other words, unexpectedly difficult task might lead to the increase of perceived uncertainty in selecting query terms and determining directions for search [9]. As a result, participants tended to “take a step back” by formulating short and broad queries and constantly seeking new viable search paths. With respect to the types of behavioral features, our result suggests that compared to the within-session average values, the totalities of different types of search actions generated in the whole session can better represent  $\Delta$  task difficulty in Web search.

**RQ3: predicting the state of task difficulty expectation disconfirmation.** The results in Table 4 empirically confirm the connections between explicit search behaviors and implicit  $\Delta$  task difficulty and also naturally lead to another question: to what extent can we predict the direction or state of task difficulty expectation disconfirmation based on search behavioral features? and more importantly, to what extent can we predict the negative state of expectation disconfirmation ( $\Delta$  task difficulty  $\geq 0$ ) and determine the right time to provide useful system recommendations?

To address these questions, based on the original definition of expectation disconfirmation [15], we first identify two states of task difficulty expectation disconfirmation for constructing prediction task: positive disconfirmation (i.e.  $\Delta$  task difficulty  $< 0$ ) and negative disconfirmation (i.e.  $\Delta$  task difficulty  $\geq 0$ ). Based on this classification, our dataset contains 47 positive cases and 33 negative cases. Given the nature of RQ3, we focus on two measures when comparing the performances of the classifiers: (1) *the overall accuracy of predicting the state of  $\Delta$  task difficulty*; (2) *the precision of predicting the negative disconfirmation cases* where timely system supports are badly needed. In prediction analysis, we adopt multiple classifiers (see Table 5) and compare them with the baseline (ZeroR). We use 80% data for training the models and 20% data for testing.

We use both whole-session and within-session behavioral features in building prediction models. The result in Table 5 demonstrates that the best classifiers/performers built on search behavioral features clearly outperform the baseline model in both overall accuracy and the precision of predicting negative cases. As our response to RQ3, this result illustrates the potential of search behavioral models in predicting the state of  $\Delta$  task difficulty and empirically proves that it is possible to predict the direction or state of expectation disconfirmation based on explicit search behaviors and thereby to determine the right time to provide apt supports for users.

**Table 5: Predicting the state of expectation disconfirmation.**

Classifier	Accuracy	Precision
Logistic Regression	.575	.462
Naive Bayes	.625	.552
K-nearest Neighbors	.600	.520
Support Vector Machine	.550	.412
Multilayer Perceptron	.513	.412
Random Forest	<b>.638</b>	.577
Decision Tree	.625	<b>.636</b>
ZeroR (baseline)	.587	.000

*Note:* ZeroR (baseline) model simply labels all cases with the most frequent label (in our dataset, positive label). The best performers in each column are boldfaced.

## 4 DISCUSSION AND CONCLUSION

This study investigates the expectation disconfirmation problem in Web search and tests its effects on behavior and experience. Specifically, in this work we focus on *task difficulty expectation disconfirmation* and address three logically connected research questions.

With respect to the RQ1, our results show that the increase in  $\Delta$  task difficulty can significantly decrease participants’ perceived level of search success and also increase their perceived time pressure. This results confirm the effects of expectation disconfirmation identified in previous classical research (e.g., [3, 15]) in the context of Web search. Regarding RQ2, our results show that  $\Delta$  task difficulty is positively correlated with several aspects of search behaviors, indicating that when participants encountered an unexpectedly difficult task, they tended to work harder and constantly try various alternative search tactics. Built upon our response to the RQ2, in RQ3, we test the predicative power of search behavioral models in predicting the state of  $\Delta$  task difficulty. The results indicate that the best classifiers/performers built upon search behavioral features clearly outperform the baseline model (ZeroR) in both the overall accuracy and the precision of predicting negative cases. The RQ2 and RQ3 jointly reveal the connections between task-related expectation, task perception, and search behavior. In particular, the answer to RQ3 demonstrates that the prediction models built on behavioral signals can to some extent predict the negative disconfirmation cases. Therefore, it is possible for IR systems to leverage this knowledge and offer proactive supports for the users who encounter unexpected difficult tasks and obstacles [17].

As always, there are limits to our study as well as needs for future efforts. This study only focused on task difficulty expectation disconfirmation and left out other aspects of users’ expectations, such as expected ease of use and knowledge gain [18]. However, this research illustrates a promising direction of applying ECT framework in understanding the effects of expectation disconfirmation and may encourage future research to explore other aspects of expectations and perceptions in users’ interactions with IR systems. In addition, the findings reported here need to be tested on other datasets collected from both lab and naturalistic settings. Based on larger scale datasets and additional fine-grained behavioral features (e.g., cursor movement features [7]), it seems realistic to believe that we will eventually be able to predict various aspects of expectation disconfirmation in search interaction and to proactively provide users with timely system recommendations and supports.

## 5 ACKNOWLEDGEMENTS

This work is supported by the National Science Foundation (NSF) grant IIS-1717488. It also uses dataset generated as a part of the NSF grant IIS-1423239.

## REFERENCES

- [1] Jaime Arguello. 2014. Predicting search task difficulty. In *European Conference on Information Retrieval*. Springer, 88–99.
- [2] Anne Aula, Rehan M Khan, and Zhiwei Guan. 2010. How does search behavior change as search becomes more difficult?. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 35–44.
- [3] Anol Bhattacherjee. 2001. Understanding information systems continuance: an expectation-confirmation model. *MIS Quarterly* (2001), 351–370.
- [4] William H Greene. 2003. *Econometric analysis*. Pearson Education.
- [5] Daniel Hienert, Matthew Mitsui, Philipp Mayr, Chirag Shah, and Nicholas J Belkin. 2018. The Role of the Task Topic in Web Search of Different Task Types. In *Proceedings of the 2018 Conference on Human Information Interaction&Retrieval*. ACM, 72–81.
- [6] Mohammad Alamgir Hossain and Mohammed Quaddus. 2012. Expectation-confirmation theory in information system research: A review and analysis. In *Information systems theory*. Springer, 441–469.
- [7] Jeff Huang, Ryen W White, and Susan Dumais. 2011. No clicks, no problem: using cursor movements to understand and improve search. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 1225–1234.
- [8] Jeonghyun Kim. 2006. Task difficulty as a predictor and indicator of web searching interaction. In *CHI'06 Extended Abstracts on Human Factors in Computing Systems*. ACM, 959–964.
- [9] Carol C Kuhlthau. 1993. A principle of uncertainty for information seeking. *Journal of Documentation* 49, 4 (1993), 339–355.
- [10] Yuelin Li and Nicholas J Belkin. 2008. A faceted approach to conceptualizing tasks in information seeking. *Information Processing & Management* 44, 6 (2008), 1822–1837.
- [11] Jingjing Liu, Michael J Cole, Chang Liu, Ralf Bierig, Jacek Gwizdka, Nicholas J Belkin, Jun Zhang, and Xiangmin Zhang. 2010. Search behaviors in different task types. In *Proceedings of the 10th Annual Joint Conference on Digital Libraries*. ACM, 69–78.
- [12] Jingjing Liu, Jacek Gwizdka, Chang Liu, and Nicholas J Belkin. 2010. Predicting task difficulty for different task types. In *Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem-Volume 47*. American Society for Information Science, 16.
- [13] Jingjing Liu, Chang Liu, Michael Cole, Nicholas J Belkin, and Xiangmin Zhang. 2012. Exploring and predicting search task difficulty. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*. ACM, 1313–1322.
- [14] Matthew Mitsui, Jiqun Liu, Nicholas J Belkin, and Chirag Shah. 2017. Predicting Information Seeking Intentions from Search Behaviors. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 1121–1124.
- [15] Richard L Oliver. 1980. A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research* (1980), 460–469.
- [16] Eun Youp Rha, Matthew Mitsui, Nicholas J Belkin, and Chirag Shah. 2016. Exploring the relationships between search intentions and query reformulations. *Proceedings of the Association for Information Science and Technology* 53, 1 (2016), 1–9.
- [17] Chirag Shah. 2018. Information Fostering-Being Proactive with Information Seeking and Retrieval: Perspective Paper. In *Proceedings of the 2018 Conference on Human Information Interaction&Retrieval*. ACM, 62–71.
- [18] Ran Yu, Ujwal Gadiraju, Peter Holtz, Markus Rokicki, Philipp Kemkes, and Stefan Dietze. 2018. Predicting User Knowledge Gain in Informational Search Sessions. *arXiv preprint arXiv:1805.00823* (2018).