



A Reference Dependence Approach to Enhancing Early Prediction of Session Behavior and Satisfaction

Tyler Brown
Tyler.M.Brown-1@ou.edu
The University of Oklahoma
Norman, Oklahoma

Jiqun Liu
jiqunliu@ou.edu
The University of Oklahoma
Norman, Oklahoma

ABSTRACT

There is substantial evidence from behavioral economics and decision sciences demonstrating that in the context of decision-making under uncertainty, the carriers of value behind actions are *gains and losses defined relative to a reference point* (e.g. pre-action expectations), rather than the absolute final outcomes. Also, the capability of early predicting session-level search decisions and user experience is essential for developing reactive and proactive search recommendations. To address these research gaps, our study aims to 1) develop reference dependence features based on a series of simulated user expectations or reference points in *first query segments* of sessions, and 2) examine the extent to which we can enhance the performance of early predicting session behavior and user satisfaction by constructing and employing reference dependence features. Based on the experimental results on three datasets of varying types, we found that incorporating reference dependent features developed in first query segments into prediction models achieves better performance than using baseline cost-benefit features only in early predicting three key session metrics (user satisfaction score, session clicks, and session dwell time). Also, when running simulations by varying the search time expectation and rate of user satisfaction decay, the results demonstrate that users tended to expect to complete their search within a minute and showed a rapid rate of satisfaction decay in a logarithmic fashion once surpassing the estimated expectation points. By factoring in a user's search time expectation and measuring their behavioral response once the expectation is not met, we can further improve the performance of early prediction models and enhance our understanding of users' behavioral patterns.

CCS CONCEPTS

• **Information systems** → **Users and interactive retrieval.**

KEYWORDS

Reference dependence, information retrieval, user satisfaction

ACM Reference Format:

Tyler Brown and Jiqun Liu. 2022. A Reference Dependence Approach to Enhancing Early Prediction of Session Behavior and Satisfaction. In *The*

ACM/IEEE Joint Conference on Digital Libraries in 2022 (JCDL '22), June 20–24, 2022, Cologne, Germany. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3529372.3533294>

1 INTRODUCTION

Understanding how users behave and evaluate their experience in search sessions is a central topic for interactive information retrieval (IR) research. A large body of IR research have developed and tested a series of process-oriented behavioral metrics and outcome-oriented offline metrics, aiming to predict users' search decisions and levels of satisfaction and evaluate the performance of interactive search systems [5, 19]. There is substantial evidence from behavioral economics and decision sciences demonstrating that in the context of decision-making under uncertainty, the carriers of value behind actions are *gains and losses defined relative to a reference point* (e.g. pre-action expectations), rather than the absolute final outcomes [8, 25]. Although this *Reference Dependence Effect* has been tested and confirmed by a series of behavioral experiments [25], it has been largely ignored by most formal IR models built upon oversimplified assumptions on user rationality [23]. This mismatch between ideal assumptions and bounded rationality causes difficulties for understanding and predicting human behavior and post-interaction experience in many domains and scenarios [24], including prolonged search sessions [15].

In addition to advancing science in understanding users, it is also critical to leverage the knowledge about users in providing adaptive and ideally proactive supports for users at early stages of search sessions [3, 11, 18, 22]. [21] found that users' search behaviors in first query of a session can achieve the same level of accuracy in search task prediction as using whole session search data. This finding suggests the possibility of early predicting different aspects of search interactions and providing proactive supports and interventions for struggling users. Leveraging the insights from user behaviors associated with the first query will allow us to go beyond traditional whole session approach [13, 14, 16] and connects user modeling with intelligent search recommendation development.

To address the above research gaps in IR user modeling and improving the practical value of prediction models, our study employs three datasets of varying types (see Table 1) which jointly provide 1,840 search sessions, simulates a variety of empirically confirmed *reference points* based on dwell time and click features, and develops *reference dependence features* to support session behavior and user satisfaction prediction. To support early prediction (as a preparation for proactive search recommendation), this study extracts search behavior and reference dependence features *only from the first query of each search session*. These first query measures and estimated reference dependent values are used to develop

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.

JCDL '22, June 20–24, 2022, Cologne, Germany

© 2022 Association for Computing Machinery.

ACM ISBN 978-1-4503-9345-4/22/06...\$15.00

<https://doi.org/10.1145/3529372.3533294>

early prediction models and to estimate users’ search decisions (e.g. session dwell time, total page clicks) and levels of satisfaction.

Our study makes threefold contributions: 1) it demonstrates that employing reference dependence features can improve the performance of predicting session-level behaviors and user satisfaction; 2) it indicates that the reference dependence models built upon simulated reference points can partially overcome cold-start problems (i.e. not knowing users’ actual pre-search expectation in terms of search gain and efforts) and facilitate early prediction of Web search interactions; 3) the early prediction models and the associated reference dependence features pave the path towards adaptive and proactive search recommendations, especially for struggling searchers under complex, intellectually challenging tasks.

2 RESEARCH QUESTIONS

To achieve the goals introduced above and develop a reference dependence approach to enhancing early search prediction, this study addresses two research questions:

- **RQ1:** To what extent can we improve the performance of early predicting session behavior and user satisfaction with reference dependence features?
- **RQ2:** To what extent can we further improve the predictive power of reference dependence models with varying types of dwell time expectation and rate of satisfaction decay?

To answer RQ1, we compared basic reference dependence model with the baseline *cost-benefit model* developed based on traditional outcome based features in predicting whole session behavior and satisfaction. RQ2 incorporated more advanced parameters characterizing different aspects of user expectations to the reference dependence model, aiming to better leverage the knowledge about reference dependence in understanding search interactions.

3 METHODOLOGY

In this analysis, we will be using three datasets: TREC session track 2013 (<https://trec.nist.gov/data/session2013.html>), TREC session track 2014 (<https://trec.nist.gov/data/session2014.html>), and KDD19 [17]. In the associated user studies where the three search datasets were collected, participants were asked to search for relevant information in order to complete assigned search tasks. The participants would perform several search queries sequentially until they had gathered enough information to satisfy their assigned task therefore completing a search session. During the searches, users’ search interaction signals and document features (e.g. title, snippet, relevance) were collected for facilitating evaluation experiments. In the KDD19 study, participants were also asked to give usefulness feedback on each clicked document and overall satisfaction rating for every individual query segment as well as each completed whole session. Participants were asked to record their answers with a 5-point scale, ranging from *unsatisfied* to *very satisfied*. The TREC session 13 dataset and TREC session 14 dataset consist of 133 and 1257 search sessions respectively. The KDD19 study resulted in 450 unique search sessions generated under 9 predefined search tasks. The descriptive statistics about the employed information retrieval datasets are presented in Table 1.

Table 1: Dataset characteristics

Datasets	Sessions	Queries	Tasks
TREC Session 2013	133	401	69
TREC Session 2014	1257	3276	60
KDD19 [20]	450	1548	9

Table 2: Cost Benefit metrics

Data set	CB Metric	Formula
TREC 2013	Relevant Doc Pct.	Total Relevant Docs/Total Docs
TREC 2013	Relevant Clicks Pct	Total Relevant Clicks/Total Clicks
TREC 2013	Relevant Doc Click Avg.	(Relevant Doc Pct + Relevant Click Pct)/ (Total Docs + Total Clicks)
TREC 2014	Relevant Doc Pct.	Total Relevant Docs/Total Docs
TREC 2014	Relevant Clicks Pct	Total Relevant Clicks/Total Clicks
TREC 2014	Relevant Doc Click Avg.	(Relevant Doc Pct + Relevant Click Pct)/ (Total Docs + Total Clicks)
KDD	Relevant Satisfaction	Avg. Query Satisfaction Score * Avg. Query Relevance
KDD	Relevant Clicks	Avg Query Clicks * Avg Query Relevance
KDD	Relevant Satisfaction Clicks	Relevant Satisfaction * Relevant Clicks

Note: Metrics were computed on a per query basis for the first query. Avg. Relevance is the avg. relevance score of the documents retrieved and Total docs is the total number of documents retrieved per query.

3.1 Baselines: Cost Benefit Models

To facilitate early prediction of session behavior and satisfaction levels, we extracted a series of widely employed search interaction signals for further analysis, such as query length, clicking, and relevance levels of clicked documents. These basic features were used as raw features for developing baseline and reference dependent models. Note that we limit the predictor features to the first query of sessions in order to measure the feasibility, performance and generalizability of early prediction across varying datasets.

To evaluate our reference-based approach and models, we developed three cost-benefit (CB) measures as the baseline measures based on relevance labels, user satisfaction scores, and document clicking activities. Following traditional approaches, these CB measures covers both the search efforts or costs and the benefits aspects of searching and were built based on final outcomes and did not incorporate reference dependent features. Based on the actual relevant features available in each dataset, we developed a variety of CB metrics (presented in Table 2) that gave us a combination of the relevance and another key search metric across the data sets based on availability and were of a similar fashion as those used in prior research performed [17].

3.2 Reference Dependence Models

To answer the RQs, we simulated and evaluated five reference models in early predicting session behavior and satisfaction. Four of the reference models are time based where we assumed a user starts with a simulated satisfaction score of 10 and incorporates a specific time expectation the user expects to complete their session. As the search session exceeds the user’s expectation, the satisfaction score begins to drop at a simulated rate. For example, assuming a user expects to complete their query in 60 seconds and becomes dissatisfied at a rate of 15 seconds. As the user’s query time meets their expectation, 60 seconds in this case, their satisfaction score of 10 decrements by 1 based at the simulated rate of dissatisfaction of 15 seconds. This score decrements by 1 every 15 seconds until a score of zero is reached. We tested both a *linear* and *logarithmic* rate of decay to simulate different satisfaction decay patterns that

users may have when interacting with search systems. The linear rate of dissatisfaction would simulate a user becoming increasingly dissatisfied at a constant rate while the logarithmic model would model a user having a sharp decline in satisfaction once their expectation is exceeded and slowly leveling off. These time based reference dependent models were also generated off the user’s total dwell time on result pages using the linear and logarithmic decays for a total of four time based models.

Apart from general model fitting purposes, we employ the logarithmic model as it also echoes another user trait confirmed by behavioral research on reference dependence: people are more sensitive to the variations near their reference points and become less sensitive when the changes in perceived gains or efforts are far away from their reference points [2, 26, 27].

Our last reference dependence model was a similar metric where the users started at a neutral satisfaction score of zero and if they clicked on a document deemed relevant the score would increase by 1 and if it was a non-relevant document the score would decrease by 1. The final score was the summation on the clicks.

By incorporating these reference dependence features into the modeling, our analysis can take into consideration varying expectations and effort-based perceptual biases that user may have when engaging in the search session. Lastly, to better represent individual level differences in reference points, a sensitivity value was randomly generated from a normal distribution for each search session and multiplied by the CB and reference dependent features. This is because users may have varying levels of sensitivity and may have a higher tolerance for longer search sessions under complex tasks.

3.3 Analysis

For this analysis, we employed five mainstream prediction algorithms (Linear, Random Forest, Gradient Boosting, Support Vector Regression, and LASSO Regression) and implemented four evaluation metrics (R^2 , MAE, MSE, RMSE) to help us judge performance across the three datasets. The datasets were split into 90/10 training and testing sets for training the reference and baseline models.

For RQ1, we ran the models with the three CB metrics as shown in Table 2 to generate a baseline on our evaluation metrics. Afterwards, we added the baseline features but added one of the reference dependence feature to test if we see an increase in the R^2 score and a subsequent decrease in the error metrics. For the query and dwell time reference base models, the starting search time expectation was set at 0 seconds as well as a rate of dissatisfaction of 15 seconds as a baseline. This was performed across only one set of time parameters to provide an initial understanding if we could see performance improvement by implementing a reference dependent feature and the parameters will be expanded in RQ2. This was performed for each target variable with all five models performance being recorded on all data sets where the data was available to determine if we see consistency. In addition, we also tested the Pearson correlation for the cost benefit metrics and the reference dependence features on each target variable.

For RQ2, we enriched our analysis by simulating various starting time expectations and rates of dissatisfaction or satisfaction decay to determine what mix of parameters returned the best results. Three features from the datasets, the cost benefit metric with the highest

Table 3: RQ1: CB and CB + Ref comparison

Data set	Target	CB vs Best Model	Best Model	R^2	MAE	MSE	RMSE
TREC 2013	Session Time	CB	SVR	0	4.74	37.58	6.13
	Session Time	CB + Query Lin Ref	Lasso	0.15	4.34	31.91	5.65
TREC 2013	Session Clicks	CB	Lasso	0	0.59	0.59	0.77
	Session Clicks	CB + Dwell Time Lin Ref	Lasso	0.04	0.56	0.56	0.75
TREC 2013	Session Dwell Time	CB	Lasso	0	4.95	38.77	6.23
	Session Dwell Time	CB + Dwell Time Lin Ref	SVR	0.15	4.63	29.04	5.39
TREC 2014	Session Time	CB	Linear	0	2.35	17.04	4.13
	Session Time	CB + Query Lin Ref	GB	0.27	2.22	12.52	3.53
TREC 2014	Session Clicks	CB	GB	0.17	0.62	0.49	0.7
	Session Clicks	CB + Click Ref Score	GB	0.46	0.49	0.32	0.57
TREC 2014	Session Dwell Time	CB	Linear	0.06	1.71	7.45	2.73
	Session Dwell Time	CB + Dwell Time Log Ref	RF	0.43	0.94	4.55	2.13
KDD	Session Time	CB	Linear	0.22	70.46	8227.62	90.71
	Session Time	CB + Query Log Ref	RF	0.44	58.78	5848.64	76.48
KDD	Session Clicks	CB	Linear	0.33	0.56	0.47	0.69
	Session Clicks	CB + Query Log Ref	Linear	0.34	0.56	0.47	0.68
KDD	Session SAT Score	CB	RF	0.25	0.72	0.66	0.82
	Session SAT Score	CB + Query Log Ref	RF	0.25	0.69	0.67	0.82

Note: SAT: satisfaction.

correlation to the target variable, and the reference dependent model was used for prediction of each target variable. Three of the previous models (Linear Regression, Gradient Boosting, and Random Forest) were run as they tended to perform the best under RQ1 in evaluation metrics. The simulation was run across each target variable using the linear and logarithmic model. For the starting search time expectation, we tested 0, 30, 60, and 90 seconds and tested rates of dissatisfaction at 15, 30, and 45 seconds.

4 RESULTS

For RQ1, Table 3 below shows the performance of cost benefit baseline evaluation metric and that of the best performing reference dependent model joined with the cost benefit metrics. This gives us a sense of the performance improvement by adding in a reference dependent feature compared with the traditional model.

Results show that in nearly every target variable and dataset, we saw an improvement in the R^2 score and a reduction in error metrics from the reference dependence model compared to the baseline models. We also found that the logarithmic model tended to be the better performing model over the linear model. This may be because the logarithmic model better captured the variation in users’ sensitivity to the changes in costs and benefits relative to the reference points. In addition, we also evaluated the Pearson correlation with the target variables. Table 4 shows the correlation results and how we tended to see higher correlations to the target variables with our reference dependent features when compared to the cost benefit metrics. This result indicates that users’ whole session behavior (e.g. total dwell time and clicks) and satisfaction level are highly associated with the reference dependent values in early search stages (i.e. first query segment), and that it is feasible and promising to conduct early prediction with the psychology-informed simulated reference points in multi-round search interactions. Note that in most cases, the reference dependent value in first query segments were negatively associated with whole session behavioral measures (e.g. session dwell time, total clicks), indicating that when a user spend more search efforts than expected in the first query segment, it may discourage them from actively interacting with search systems in the session.

Regarding RQ2, Table 5 shows the best parameters (i.e. simulated starting time expectation and rate of satisfaction decay) for our

Table 4: RQ1: Pearson correlation

Target	Metric	TREC 2013	TREC 2014	KDD
Session Time	Baseline CB	-0.142	-0.001	0.034
Session Time	Lin Ref	-0.296	-0.381	-0.357
Session Time	Log Ref	-0.226	-0.34	-0.519
Session Time	Dwell Time Lin Ref	-0.37	-0.103	
Session Time	Dwell Time Log Ref	-0.095	-0.021	
Session Time	Clicks Ref Score	-0.106	-0.01	
Session Clicks	Baseline CB	0.019	0.097	0.369
Session Clicks	Lin Ref	-0.138	-0.02	-0.109
Session Clicks	Log Ref	-0.127	-0.076	-0.298
Session Clicks	Dwell Time Lin Ref	-0.332	-0.374	-0.208
Session Clicks	Dwell Time Log Ref	-0.278	-0.59	
Session Clicks	Clicks Ref Score	-0.098	-0.09	
Session Time	Baseline CB	0	0.077	
Session Time	Lin Ref	-0.298	-0.252	
Session Time	Log Ref	-0.254	-0.278	
Session Time	Dwell Time Lin Ref	-0.528	-0.405	
Session Time	Dwell Time Log Ref	-0.29	-0.376	
Session Time	Clicks Ref Score	-0.096	-0.09	
Session SAT Score	Baseline CB			-0.033
Session SAT Score	Lin Ref			0.117
Session SAT Score	Log Ref			0.247
Session SAT Score	Dwell Time Lin Ref			
Session SAT Score	Dwell Time Log Ref			
Session SAT Score	Clicks Ref Score			

Note: $p < 0.05$ is indicated by the boldfaced underlined numbers. SAT: satisfaction.

Table 5: RQ2: Reference parameter simulation

Data set	Target	Ref. Model	Starting Expectation (s)	Rate of Decay (s)	Prediction Algorithm
TREC 2013	Session Time	Logarithmic	60	30	Linear Regression
TREC 2013	Session Clicks	Logarithmic	90	15	Linear Regression
TREC 2013	Session Dwell Time	Logarithmic	90	15	Linear Regression
TREC 2014	Session Time	Logarithmic	60	30	Random Forest
TREC 2014	Session Clicks	Linear	0	45	Gradient Boosting
TREC 2014	Session Dwell Time	Linear	0	15	Random Forest
KDD	Session Time	Logarithmic	30	15	Gradient Boosting
KDD	Session Clicks	Logarithmic	30	15	Random Forest
KDD	Session Score	Logarithmic	60	15	Random Forest

session time and query dwell time reference based features. We observed best performance in session time prediction when we assume that the user expected to complete their session within 30-60 seconds with a dissatisfaction rate of 30 seconds. In most cases across the target variables this held true, showing the user often expected to find their results rather quickly. The best performing reference model whether being the linear or logarithmic model was also listed. The logarithmic model was overwhelmingly the better performing model across the target variables, indicating that the user tended to have a rather steep decline in satisfaction once their expectation is exceeded rather than a linear constant decline. This also suggests that once a user exceeds their expectation over time their dissatisfaction did not drop dramatically or in a strictly linear manner. For example, If the user exceeds their expectation by a minute their dissatisfaction is not that dissimilar from a minute and a half. These results echo the findings on reference dependence from behavioral economics experiments [9, 26] and demonstrate the value of incorporating a reference dependence approach.

5 DISCUSSION AND CONCLUSION

Users' search interaction and evaluation are affected by not only algorithmic biases [6], but also cognitive and perceptual biases [1]. This study seeks to leverage the insights about reference dependence bias from behavioral experiments in enhancing early prediction of session behavior and user satisfaction. To achieve this,

we simulated reference based behavioral models where users had a search time expectation and a rate of satisfaction decay as their search efforts exceeded their expectations in two different rates.

RQ1. In our first research question, we looked at whether by taking traditional cost-benefit metrics as a baseline and adding in a reference based feature with those traditional cost benefit features increased our predictive results. In nearly every data set and target variable, we observed significant improvements in our evaluation metrics when incorporating a reference based feature. This provides additional support that we can increase the predictive power of behavioral models by employing reference dependent approach and incorporating users' expectations when performing a search session and therefore increasing ability for early prediction. This aligns with prior research incorporating different reference points into whole session experience prediction [15].

RQ2. In our second research question, we examined the expected search time reference based feature and ran a simulation varying the parameters to help us find which parameter set achieved the highest results across our evaluation metrics. We varied this across both a linear and logarithmic reference dependent model simulating different behaviors the user may experience when engaging in a search session. We found that the parameters with the highest performance was when the user expected to find a result in 30 to 60 seconds and tended to use a much quicker rate of dissatisfaction (15-30 seconds). The logarithmic model performed better in nearly every case providing evidence that the user became dissatisfied very quickly after exceeding their search time expectation and leveled off rather than a constant rate of dissatisfaction. The results echo the finding [10, 17] that incorporating user expectations can increase our early predictive capabilities as well as help us understand that users expect to complete their searches quickly and become rapidly dissatisfied once their expectation is exceeded.

This paper represents an initial attempt and novel approach to incorporating the insights from reference dependence theory in addressing user modeling and IR evaluation problems. For future research, other parameters representing different user expectations and cognitive biases could be trialed as well as other adaptive models. This research has limitations as the reference dependent models were chosen based on time signals and further research could explore expanding these models based on other factors and features, such as knowledge states, cognitive loads, and other relevant neurophysiological measures [7, 29]. Also, task source (i.e. authentic task vs. simulated task) and search environment (e.g. controlled lab, naturalistic setting) may affect the way in which users' reference points and expectations affect their search decision-making [4, 12, 28]. With more empirical evidences on the role and impact of reference dependence effects in searching, researchers will be able to build computationally solid and behaviorally realistic user models for developing and evaluating user-oriented intelligent search systems.

6 ACKNOWLEDGEMENT

This work is partially supported by the National Science Foundation (NSF) grant IIS-2106152 and a Faculty Investment Program (FIP) grant from the Research Council of the University of Oklahoma Norman Campus.

REFERENCES

- [1] Leif Azzopardi. 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*. 27–37.
- [2] Nicholas C Barberis. 2013. Thirty years of prospect theory in economics: A review and assessment. *Journal of Economic Perspectives* 27, 1 (2013), 173–96.
- [3] Sumit Bhatia, Debapriyo Majumdar, and Nitish Aggarwal. 2016. Proactive information retrieval: Anticipating users’ information need. In *European Conference on Information Retrieval*. Springer, 874–877.
- [4] Pia Borlund and Jesper W Schneider. 2010. Reconsideration of the simulated work task situation: A context instrument for evaluation of information retrieval interaction. In *Proceedings of the third symposium on Information interaction in context*. 155–164.
- [5] Ye Chen, Ke Zhou, Yiqun Liu, Min Zhang, and Shaoping Ma. 2017. Meta-evaluation of Online and Offline Web Search Evaluation Metrics. SIGIR.
- [6] Michael D Ekstrand, Robin Burke, and Fernando Diaz. 2019. Fairness and discrimination in recommendation and retrieval. In *Proceedings of the 13th ACM Conference on Recommender Systems*. 576–577.
- [7] Jacek Gwizdka. 2010. Distribution of cognitive load in web search. *Journal of the American Society for Information Science and Technology* 61, 11 (2010), 2167–2187.
- [8] Daniel Kahneman. 2003. Maps of bounded rationality: Psychology for behavioral economics. *American economic review* 93, 5 (2003), 1449–1475.
- [9] Daniel Kahneman and Richard H Thaler. 2006. Anomalies: Utility maximization and experienced utility. *Journal of economic perspectives* 20, 1 (2006), 221–234.
- [10] Diane Kelly, Jaime Arguello, Ashlee Edwards, and Wan ching Wu. 2015. Development and Evaluation of Search Tasks for IIR Experiments using Cognitive Complexity Framework. *ICTIR* 15, Article 101-110 (September 2015). <https://doi.org/10.1145/2808194.2809465>
- [11] Markus Koskela, Petri Luukkonen, Tuukka Ruotsalo, Mats Sjöberg, and Patrik Florén. 2018. Proactive information retrieval by capturing search intent from primary task context. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 8, 3 (2018), 1–25.
- [12] Jiqun Liu. 2021. Deconstructing search tasks in interactive information retrieval: A systematic review of task dimensions and predictors. *Information Processing & Management* 58, 3 (2021), 102522.
- [13] 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.
- [14] 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.
- [15] Jiqun Liu and Fangyuan Han. 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*. 1141–1150.
- [16] 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*. 1313–1322.
- [17] Jiqun Liu and Chirag Shah. 2019. Investigating the impacts of expectation disconfirmation on web search. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*. 319–323.
- [18] Jiqun Liu and Chirag Shah. 2019. Proactive identification of query failure. *Proceedings of the Association for Information Science and Technology* 56, 1 (2019), 176–185.
- [19] Mengyang Liu, Yiqun Liu, Jiaxin Mao, Cheng Luo, Min Zhang, and Shaoping Ma. 2018. Satisfaction with Failure or Unsatisfied Success: Investigating the Relationship between Search Success and User Satisfaction. In *Proceedings of the 2018 World Wide Web Conference*. International World Wide Web Conferences Steering Committee, 1533–1542.
- [20] Mengyang Liu, Jiaxin Mao, Yiqun Liu, Min Zhang, and Shaoping Ma. 2019. Investigating Cognitive Effects in Session-level Search User Satisfaction. In *Proceedings of the 25th ACM SIGKDD 2019, Anchorage, AK, USA, August 4-8, 2019*. 923–931. <https://doi.org/10.1145/3292500.3330981>
- [21] Matthew Mitsui, Jiqun Liu, and Chirag Shah. 2018. How Much is Too Much?: Whole Session vs. First Query Behaviors in Task Type Prediction. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM, 1141–1144.
- [22] Kai Puolamäki, Jarkko Salojärvi, Eerika Savia, Jaana Simola, and Samuel Kaski. 2005. Combining eye movements and collaborative filtering for proactive information retrieval. In *Proceedings of the 28th annual international ACM SIGIR Conference on Research and Development in Information Retrieval*. 146–153.
- [23] Herbert Simon. 1972. Theories of bounded rationality. *Decision and organization*. 161–176.
- [24] Herbert A Simon. 1955. A behavioral model of rational choice. *The quarterly journal of economics* 69, 1 (1955), 99–118.
- [25] Richard H Thaler. 2016. Behavioral economics: past, present, and future. *American Economic Review* 106, 7 (2016), 1577–1600.
- [26] Amos Tversky and Daniel Kahneman. 1991. Loss aversion in riskless choice: A reference-dependent model. *The quarterly journal of economics* 106, 4 (1991), 1039–1061.
- [27] Amos Tversky and Daniel Kahneman. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty* 5, 4 (1992), 297–323.
- [28] Yiwei Wang and Chirag Shah. 2022. Authentic versus synthetic: An investigation of the influences of study settings and task configurations on search behaviors. *Journal of the Association for Information Science and Technology* 73, 3 (2022), 362–375.
- [29] Yao Zhang and Chang Liu. 2020. Users’ Knowledge Use and Change during Information Searching Process: A Perspective of Vocabulary Usage. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*. 47–56.