Proceedings of the Tenth International Workshop on Evaluating Information Access (EVIA 2023), a Satellite Workshop of the NTCIR-17 Conference, December 12-15, 2023, Tokyo, Japan
DOI: https://doi.org/10.20736/0002001351

Decoy Effect in Search Interaction: A Pilot Study

Nuo Chen
pleviumtan@toki.waseda.jp
Waseda University
Tokyo, Japan

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

Tetsuya Sakai
tetsuyasakai@acm.org
Waseda University
Tokyo, Japan

Xiao-Ming Wu
xiao-ming.wu@polyu.edu.hk
The Hong Kong Polytechnic University
Hong Kong, China

ABSTRACT

In recent years, the influence of cognitive effects and biases on users' thinking, behaving, and decision-making has garnered increasing attention in the field of interactive information retrieval. The decoy effect, one of the main empirically confirmed cognitive biases, refers to the shift in preference between two choices when a third option (the decoy) which is inferior to one of the initial choices is introduced. However, it is not clear how the decoy effect influences user interactions with and evaluations on Search Engine Result Pages (SERPs). To bridge this gap, our study seeks to understand how the decoy effect at the document level influences users' interaction behaviors on SERPs, such as clicks, dwell time, and usefulness perceptions. We conducted experiments on two publicly available user behavior datasets and the findings reveal that, compared to cases where no decoy is present, the probability of a document being clicked could be improved and its usefulness score could be higher, should there be a decoy associated with the document.

CCS CONCEPTS

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

KEYWORDS

decoy effect, cognitive bias, interaction information retrieval

1 INTRODUCTION

Understanding how users think, behave, and make decisions during interactions with search systems represents a foundational research interest in the field of interactive Information Retrieval (IR). A cognitive bias is a systematic pattern of deviations in thinking which may lead to irrational judgements and problematic decision-making [24, 26]. In recent years, the exploration of cognitive biases and their impact on the information seeking and retrieval behaviors and outcomes has garnered increasing attention [1, 10, 15].

The *decoy effect*, which is one kind of cognitive bias, describes a situation in which individuals alter their preference between two initial choices when introduced to a third (*i.e.* the decoy), which is asymmetrically inferior to one of the initial choices [8]. Figure 1 illustrates an example of the decoy effect in shopping decision-making. In a shop, a customer who wants to buy a drink might waver between a 500ml bottle of water (for $1.19) and a bottle of soda with a similar size (for $1.49). The 500ml water is cheap, but the soda tastes better, so it might be hard to make a decision and

Figure 1: An example of the decoy effect. The image is generated by the authors using Midjourney.

which one the cosumer will choose might depend on the assessment of their relative utility. Yet, with a 250ml bottle of water for $1.09 presenting beside the 500ml water, the customer might lean towards the 500ml water, as they perceive a substantial relative gain from the comparison of the 500ml water and the 250ml water: spending an additional $0.10 to purchase a 500ml bottle of water, compared to the 250ml one, evidently presents a highly economical deal. In the above example, the 250ml water serves as the *decoy* to the *target* 500ml water.

In the field of information retrieval, Eickhoff [6] examined the impact of a decoy document on thresholds and strategies in crowdsourcing relevance judgments, showing that assessors could increase the relevance rating of target document when it is shown with the decoy document. Nevertheless, Eickhoff [6] only focuses on crowdsourcing assessors operating within the annotation interface and few research currently addresses how the decoy effect influences user interactions on Search Engine Result Pages (SERPs).

To address this gap, our study seeks to understand how the decoy effect at the document level influences users' interaction behaviors on SERPs, such as clicks, browsing dwell time, and usefulness perceptions. We conducted experiments on two publicly available user behavior datasets and the experimental results reveal that, compared to cases where no decoy is present, the probability of a document being clicked could be improved and its usefulness score could be higher, should there be a decoy associated with the document.

This study, through the perspective of behavioral economics, delves deeper into the behavioral patterns and decision-making processes of users interacting with search engines. The findings stand to encourage IR researchers to increase the explanatory power of formal search models from a realistic behavioral and psychological foundation.

2 RELATED WORK

Insights from cognitive psychology and behavioral economics suggest that, *cognitive biases* arise from one's limited cognitive ability when there are no enough resources to properly collect and process available information [9]. Due to cognitive biases, one's decisions under uncertainty can systematically deviate from what is expected given rational decision-making models [24–26].

In search and recommendation contexts, interactions between individuals and systems could lead to the incorporation of behavioral signals, influenced by *cognitive biases*, into datasets used for training machine learning algorithms, thereby potentially magnifying existing system biases [1, 10]. Cognitive biases might also result in significant deviations in users' behaviors and judgements from optimal or desired outcomes. Consequently, this could give rise to unfair decisions and outcomes between users who are more susceptible to certain biases and contextual triggers and those who are not [11]. Therefore, with an increasing number of individuals turning to search and recommendation systems to access and utilize information for life decisions, the influence of cognitive biases on the information retrieval process is drawing heightened attention from IR researchers. Thus far, a lot of studies have explored the influence of cognitive biases such as the anchoring effect [20], the priming effect [3, 19], the ordering effect [2], and the reference dependence effect [12, 27] on document examination, relevance judgment, and evaluation of whole-session search satisfaction. In recent years, with the growing knowledge about users' cognitive biases, some works began to introduce cognitive biases into the construction and meta-evaluation of evaluation metrics [4, 5, 13, 28], but this is out of the scope of this paper.

In this paper, we specifically shed light on one of the cognitive effects, the *decoy effect*. The *decoy effect*, which is one kind of cognitive biases, describes a situation in which individuals alter their preference between two initial choices when introduced to a third (*i.e.* the decoy), which is asymmetrically inferior to one of the initial choice [8]. In the fields of e-commerce and recommendation systems, there have been some studies exploring the impact of the decoy effect [18, 22, 23]. In the field of information retrieval, it is not clear how the decoy effect influences user interactions with and evaluations on Search Engine Result Pages. The work most closely related to ours in theme is that of Eickhoff [6], which shows that when a relevant item is presented alongside two non-relevant items, with one non-relevant item being distinctly inferior (i.e., the decoy), assessors tend to rate the superior non-relevant document as more relevant. Nevertheless, Eickhoff [6] only focuses on crowdsourcing assessors operating within the annotation interface and our study addresses how the decoy effect could influence user interactions on Search Engine Result Pages (SERPs), which is a broader scenario.

3 RESEARCH QUESTION

To address the research gap mentioned above, in this study, we seek to understand how the decoy effect at the document level influences users' interaction behaviors on SERPs, such as clicks, browsing dwell time, and usefulness perceptions. Specifically, our work sought to answer following **research question** (RQ):

How, and to what extent, the presence of a decoy influences the likelihood of a document being clicked, the browsing duration on it, and its perceived usefulness?

4 EXPERIMENT

In this section, we introduce the datasets employed and the methodology adopted in the experiment, as well as the experimental results.

4.1 Datasets

The THUIR2016 dataset [17] is collected under controlled laboratory setting, whereby participants were instructed to execute a given intricate search tasks utilizing commercial search engines. This dataset encompasses a total of 9 topics, 225 search sessions and 933 queries, along with the title and snippets on the SERP of each query. It also contains 4-level user self-rating usefulness scores for the items they clicked and 5-level graded relevance labels collected from external assessors.

The THU-KDD dataset [16] is also collected under controlled laboratory setting similar to the THUIR2016 dataset. This dataset encompasses a total of 9 topics, 450 search sessions and around 1100 queries, along with the title and snippets on the SERP of each query. It also contains 4-level user self-rating usefulness scores for the items they clicked and 4-level graded relevance labels collected from external assessors.

4.2 Data Processing Approach

To find out potential decoy instances from user logs, we first proffer a definition of a decoy instance. A pair of documents, composed of the *target document* and the *decoy document* (t, d), constitutes a decoy instance if and only if the following conditions are met: (1) t and d share certain degree of similarity in content, albeit not identical, i.e., $S_{\min} \leq \text{similarity}(t, d) \leq S_{\max}$, where $S_{\min}$ and $S_{\max}$ respectively represent the minimum and maximum similarity thresholds; (2) d is inferior in quality to t, i.e., $\text{quality}(t) > \text{quality}(d)$; (3) the position t and d within a SERP is close enough, i. e., $|rank(t) - rank(d)| \leq \Delta_{\text{rank}}$.

In our experiments, given that the documents are all in Chinese, we first concatenate the title and snippet into a string. Subsequently, we employ a tool named `jieba`[1] to carry out word segmentation, obtaining a token list. We then calculated the cosine similarity with the vanilla definition [7] between each pair of documents under the same topic. We designated $S_{\min}$ as the 99th percentile of document similarity in each dataset, setting $S_{\max}$ to 0.95. In THUIR2016 dataset, $S_{\min}$ stands at 0.626 and in THU-KDD dataset, $S_{\min}$ stands at 0.594. For the second condition, we employ the relevance scores given by external assessors as the measurement of document quality, mandating that $\text{relevance}(t) - \text{relevance}(d) \geq 2$ to ensure that the decoy is substantially inferior to the target. For the third condition, we require the absolute value of the difference of the rank between t and d is smaller than or equal to 5 ($\Delta_{\text{rank}} = 5$). We processed the top 10 documents in each SERP on all datasets adhering to the aforementioned three conditions, and we identified 982 records of decoy pairs involving 318 distinct target documents in

[1] https://github.com/fxsjy/jieba

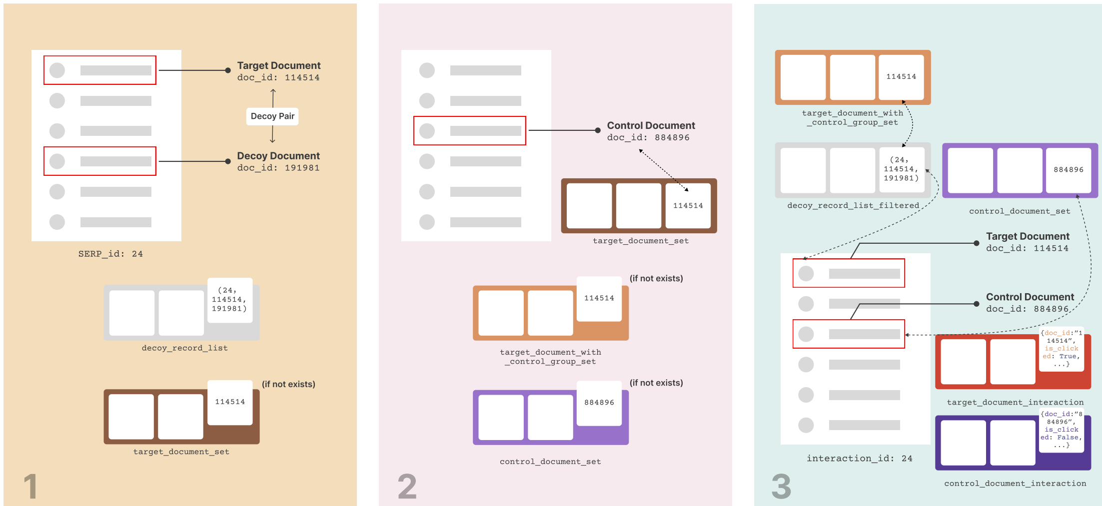

Figure 2: Data processing flow in the first experiment.

the THUIR2016 dataset; and 922 records of decoy pairs involving 376 distinct target documents. in the THU-KDD dataset. In the following discourse, we denote the set consisting of all target documents in a corpus as $\mathcal{T}$.

To investigate whether user interactions with documents are disparate when no decoy is present compared to situations with a decoy, we assign some documents not in $\mathcal{T}$ to the control group (*i.e., control documents*), adhering to the following condition: A document c which is not in the set of target documents (*i.e.,* $c \notin \mathcal{T}$) is considered a control document if and only if it matches a target document $t \in \mathcal{T}$ such that $\text{similarity}(c, t) \geq S_{\textbf{control}}$ and $|\text{relevance}(c) - \text{relevance}(t)| <= 2$. We denote the set of all such c as C, and the set of all t that can match with at least one c as $\mathcal{T}'$, where $\mathcal{T}' \subset \mathcal{T}$. In our experiments, we set $S_{\textbf{control}}$ to the 99.5th percentile of document similarity in each dataset. In THUIR2016 dataset, $S_{\textbf{control}}$ stands at 0.709 and in THU-KDD dataset, $S_{\textbf{control}}$ stands at 0.676. According to the aforementioned condition, we have identified 741 qualifying *control documents* in the THUIR2016 dataset and 1790 in the THU-KDD dataset.

We then extracted interaction records of all *control documents* in the THUIR2016 and THU-KDD datasets, obtaining 1384 and 2770 records respectively. Subsequently, from the records of decoy pairs in the THUIR2016 and THU-KDD datasets (with 982 and 922 records respectively), we filter out all records where $t \in \mathcal{T}'$, obtaining 739 and 828 records respectively. Note that, for decoy pairs from the same SERP interaction i, there could be situations where the same target document corresponds to multiple decoy documents. In our filtering process, we ensure that for a given SERP interaction i and a given target document t, only one record is eventually extracted. We concatenate the interaction records of target documents and control documents, ultimately obtaining document interaction record lists of lengths 2123 and 3598 in the two respective datasets. These two lists of interactions will be employed for the subsequent data analysis. In the subsequent analysis, we process the interaction signals as follows: for documents that have not been clicked, their usefulness score is assigned a value of 0, and their browsing duration is also set to 0. Figure 2 provides a brief outline of our data processing workflow.

4.3 Data Analysis and Experimental Results

From Figure 3, one can see that in both datasets, there are more instances receiving a score of 4 (the maximum) in *usefulness*. On the THU-KDD dataset, the target group exhibited fewer instances labeled as 0 and 1; the proportions of instances labeled as 2 and 3 were approximately similar between the target and control groups. On the THUIR2016 dataset, the target group exhibited fewer instances labeled as 1 and 2. Additionally, the proportions of instances labeled as 0 and 3 were approximately similar between the target and control groups. For *click probability*, there is no difference in the THUIR2016 dataset, while in the THU-KDD dataset, the click probability of target group is higher. Regarding *browsing duration*, on the THUIR2016 dataset, the target group exhibited a higher proportion of instances with browsing times ranging from 30 to 60 seconds. Apart from this observation, the distributions of browsing durations for both the target and control groups were approximately similar. On the THU-KDD dataset, the target group exhibited a higher proportion of instances with browsing times ranging between 0 to 5 seconds and 30 to 60 seconds, while a smaller proportion of instances had browsing times between 5 to 30 seconds.

From Figure 3, it is challenging to derive an intuitive conclusion, especially regarding the relationship between the decoy effect and variables such as click probability, browsing duration and usefulness. Nevertheless, Figure 4 shows that, across both datasets, the distribution of target documents and control documents over ranks diverges. Hence, it is necessary to factor out any latent effects stemming from position biases on our results in order to draw a

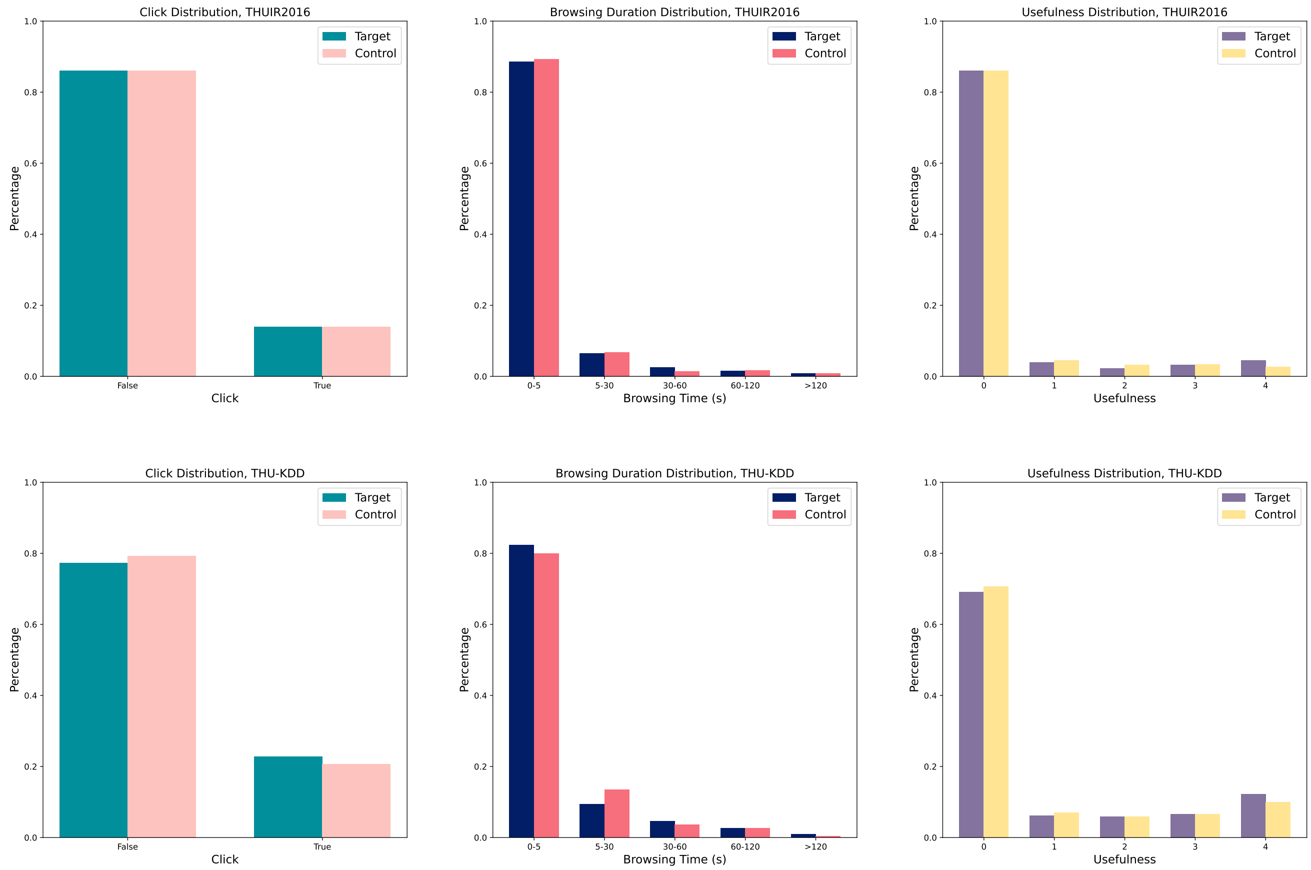

Figure 3: The distribution of click probability, browsing time and usefulness score on THUIR2016 dataset (top) and THU-KDD (bottom) dataset respectively.

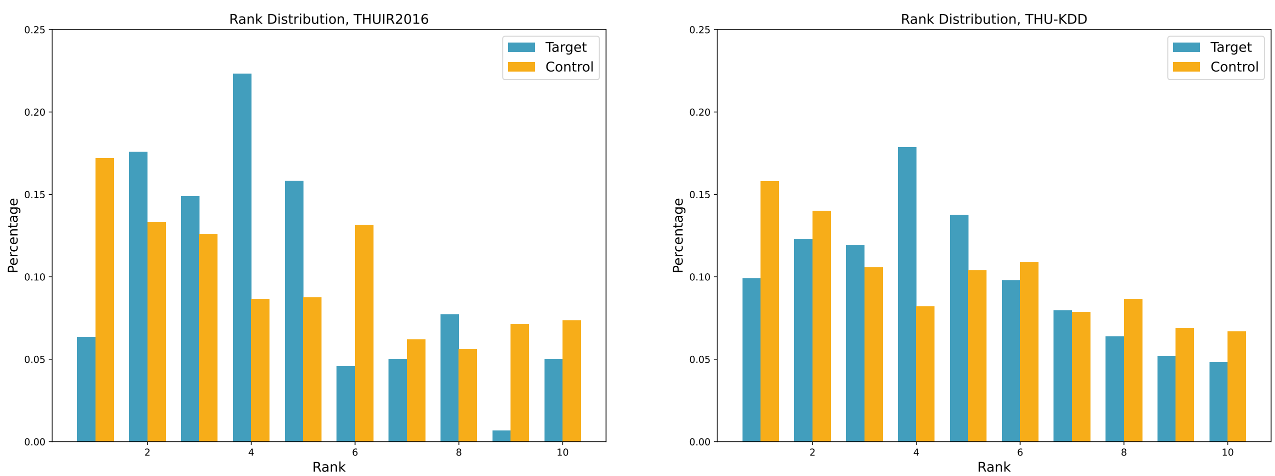

Figure 4: The distribution of rank on THUIR2016 (left) dataset and THU-KDD (right) dataset respectively.

conclusion on the relationship between the presence of decoy and users' behavior.

To control the impact caused by *rank position*, we employ regression analysis to investigate the relationships between the presence of a decoy and the probability of clicks, browsing duration, and usefulness scores. We constructed three regression models, taking whether the document is clicked (`is_clicked`), browsing duration (`duration`), and usefulness score (`usefulness`) as dependent variables respectively, and the presence of a decoy (`has_decoy`), the document's rank (`rank`), task ID (`task_id`), and user ID (`user_id`) as independent variables.

Note that contrary to computer science, in econometrics, regression models are predominantly employed for interpretation rather than for prediction. In a multiple regression model, each coefficient tells people the impact on the dependent variable of a one-unit change in that independent variable, holding all other independent variables constant [21]. In this study, we foucus on elucidating how, and to what extent, the presence of a *decoy* influences the likelihood of a document being clicked, the browsing duration on it, and its perceived usefulness. Hence, we do not partition the dataset into training and test subsets; instead, we perform regression on the entirety of the data. Including `rank`, `task_id`, and `user_id` as independent variables serves to use them as control variables to mitigate the potential influences from rank position, task type, and individual characteristics on the outcomes, thus better elucidating how variations in `has_decoy` would affect the values of `is_clicked`, `duration`, and `usefulness`. For `is_clicked`, we employ Logistic regression, and for `duration` and `usefulness`, we resort to Ordinary Least Squares (OLS) regression.

Table 1 shows the regression coefficient of the independent variable `has_decoy` with the dependent variables `is_clicked`, `duration` and `usefulness`. As previously stated, our focus in this research is to elucidate in what manner and to what extent the presence of a decoy impacts whether a document is clicked, the browsing duration, and the usefulness scores, and `rank`, `task_id`, and `user_id` are included merely to control for the effects brought by rank position, task, and individual characteristic respectively. Therefore, we opt to omit the reporting of the constant as well as the regression coefficients of `rank`, `task_id`, and `user_id` in the tables.

From Table 1, one can observe that: across the two datasets, the presence of a *decoy* could exert a positive influence on the likelihood of being clicked (coefficient = 0.363 and 0.217 respectively) and on the usefulness score (coefficient = 0.136 and 0.156 respectively), all with a statistical significance at the level of $p < 0.05$. The existence of a *decoy* also seems to render a positive impact on duration (coefficient = 1.916 and 1.913 respectively), however, the result on both datasets is not statistically significant.

More precisely, the regression results can be interpreted as follows: Given the document rank, type of task, and individual characteristics, when a *decoy* is present, in comparison to when it is absent (1) the likelihood of being clicked on the THUIR2016 and THU-KDD datasets would respectively increase by 36.3% ($p < 0.05$) and 21.7% ($p < 0.05$); (2) browsing time duration would rise by 1.92s and 1.91s respectively on the THUIR2016 and THU-KDD datasets; (3) usefulness score would escalate by 0.136 ($p < 0.01$) and 0.156 ($p < 0.01$) respectively on the THUIR2016 and THU-KDD datasets.

Dependent Variable	THUIR2016 Coefficient	THU-KDD Coefficient
`is_clicked`	0.363^{*}	0.217^{*}
`duration`	1.916	1.913
`usefulness`	0.136^{**}	0.156^{**}

Table 1: The regression coefficient of the independent variable `has_decoy` with the dependent variables `is_clicked`, `duration`, and `usefulness`. * and ** respectively indicate that the coefficients are significant at levels of $p < 0.05$, $p < 0.01$.

5 CONCLUSION

In this study, we seek to comprehend how the *decoy effect* at the document level impacts users' interaction behaviors on SERPs, such as clicks, dwell time, and usefulness perceptions. We conducted descriptive data analysis and regression analysis on two publicly available user behavior datasets. The experimental results indicate that, the likehood of a document being clicked could be improved and its usefulness score could be elevated, should there be a decoy associated with the document. From the experimental results, we observe that, given the document rank, type of task and individual characteristics, when a decoy is present, in comparison to when it is absent, there is a significant increment in the likehood of a document being clicked and its perceived usefulness.

As far as we know, we are the first to addresses how the decoy effect influences user interactions on Search Engine Result Pages. Our work extends the endeavors of the IR community in exploring how cognitive biases impact user behaviors in document examining and relevance judgment, providing evidence from the perspective of the decoy effect.

However, our study is merely in a preliminary stage, with numerous aspects awaiting further exploration. For instance: (1) Does the impact of the decoy vary in size under different themes, topics, or search tasks? (2) Within the same search session, does the impact of decoy results vary across different cognitive states and information seeking intentions (e.g. exploring an unfamiliar domain, seeking for a known item, evaluating retrieved information) [14]? (2) The potential impact of the decoy effect on the benefit of users and information providers, as well as possible social and ethical issues it may bring about, such as fairness. Addressing these questions will necessitate subsequent experiments to provide more evidence.

ACKNOWLEDGEMENT

We thank Prof. Kana Shimizu for providing valuable feedback for this study. We also thank all reviewers of this paper for their valuable feedback and constructive criticism. Jiqun Liu's participation in this project is partially supported by the Bridge Fuding Investment Program (BFIP) Award from the University of Oklahoma Office of the Vice President for Research and Partnerships.

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* (Canberra ACT, Australia) *(CHIIR '21)*. Association for Computing Machinery, New York, NY, USA, 27–37. https://doi.org/10.1145/3406522.3446023

[2] Nick Bansback, Linda C. Li, Larry Lynd, and Stirling Bryan. 2014. Exploiting order effects to improve the quality of decisions. *Patient Education and Counseling* 96, 2 (2014), 197–203. https://doi.org/10.1016/j.pec.2014.05.021

[3] Praveen Chandar and Ben Carterette. 2012. Using Preference Judgments for Novel Document Retrieval. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Portland, Oregon, USA) *(SIGIR '12)*. Association for Computing Machinery, New York, NY, USA, 861–870. https://doi.org/10.1145/2348283.2348398

[4] Nuo Chen, Jiqun Liu, and Tetsuya Sakai. 2023. A Reference-Dependent Model for Web Search Evaluation: Understanding and Measuring the Experience of Boundedly Rational Users. In *Proceedings of the ACM Web Conference 2023* (Austin, TX, USA) *(WWW '23)*. Association for Computing Machinery, New York, NY, USA, 3396–3405. https://doi.org/10.1145/3543507.3583551

[5] Nuo Chen, Fan Zhang, and Tetsuya Sakai. 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* (Madrid, Spain) *(SIGIR '22)*. Association for Computing Machinery, New York, NY, USA, 2709–2714. https://doi.org/10.1145/3477495.3531953

[6] Carsten Eickhoff. 2018. Cognitive Biases in Crowdsourcing. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining* (Marina Del Rey, CA, USA) *(WSDM '18)*. Association for Computing Machinery, New York, NY, USA, 162–170. https://doi.org/10.1145/3159652.3159654

[7] Jiawei Han, Micheline Kamber, and Jian Pei. 2012. 2.4.7. Cosine Similarity. In *Data Mining (Third Edition)* (third edition ed.), Jiawei Han, Micheline Kamber, and Jian Pei (Eds.). Morgan Kaufmann, Boston, 77–78. https://doi.org/10.1016/B978-0-12-381479-1.00002-2

[8] Joel Huber, John W. Payne, and Christopher Puto. 1982. Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis. *Journal of Consumer Research* 9, 1 (1982), 90–98. http://www.jstor.org/stable/2488940

[9] Arie W. Kruglanski and Icek Ajzen. 1983. Bias and error in human judgment. *European Journal of Social Psychology* 13, 1 (1983), 1–44. https://doi.org/10.1002/ejsp.2420130102 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/ejsp.2420130102

[10] Jiqun Liu. 2023. *A Behavioral Economics Approach to Interactive Information Retrieval: Understanding and Supporting Boundedly Rational Users*. Vol. 48. Springer Nature.

[11] Jiqun Liu. 2023. Toward A Two-Sided Fairness Framework in Search and Recommendation. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval* (Austin, TX, USA) *(CHIIR '23)*. Association for Computing Machinery, New York, NY, USA, 236–246. https://doi.org/10.1145/3576840.3578332

[12] 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* (Virtual Event, China) *(SIGIR '20)*. Association for Computing Machinery, New York, NY, USA, 1141–1150. https://doi.org/10.1145/3397271.3401085

[13] Jiqun Liu and Fangyuan Han. 2022. Matching search result diversity with user diversity acceptance in Web search sessions. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2473–2477.

[14] Jiqun Liu, Shawon Sarkar, and Chirag Shah. 2020. Identifying and predicting the states of complex search tasks. In *Proceedings of the 2020 conference on human information interaction and retrieval*. 193–202.

[15] 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.

[16] 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 International Conference on Knowledge Discovery & Data Mining* (Anchorage, AK, USA) *(KDD '19)*. Association for Computing Machinery, New York, NY, USA, 923–931. https://doi.org/10.1145/3292500.3330981

[17] Jiaxin Mao, Yiqun Liu, Ke Zhou, Jian-Yun Nie, Jingtao Song, Min Zhang, Shaoping Ma, Jiashen Sun, and Hengliang Luo. 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* (Pisa, Italy) *(SIGIR '16)*. Association for Computing Machinery, New York, NY, USA, 463–472. https://doi.org/10.1145/2911451.2911507

[18] Fan Mo, Tsuneo Matsumoto, Nao Fukushima, Fuyuko Kido, and Hayato Yamana. 2022. Decoy Effect of Recommendation Systems on Real E-commerce Websites. In *CEUR Workshop Proceedings*, Vol. 3222. CEUR-WS, 151–163.

[19] Falk Scholer, Diane Kelly, Wan-Ching Wu, Hanseul S. Lee, and William Webber. 2013. The Effect of Threshold Priming and Need for Cognition on Relevance Calibration and Assessment. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Dublin, Ireland) *(SIGIR '13)*. Association for Computing Machinery, New York, NY, USA, 623–632. https://doi.org/10.1145/2484028.2484090

[20] Milad Shokouhi, Ryen White, and Emine Yilmaz. 2015. Anchoring and Adjustment in Relevance Estimation. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Santiago, Chile) *(SIGIR '15)*. Association for Computing Machinery, New York, NY, USA, 963–966. https://doi.org/10.1145/2766462.2767841

[21] Gary Smith. 2012. Chapter 10 - Multiple Regression. In *Essential Statistics, Regression, and Econometrics*, Gary Smith (Ed.). Academic Press, Boston, 297–331. https://doi.org/10.1016/B978-0-12-382221-5.00010-6

[22] Erich Christian Teppan, Gerhard Friedrich, and Alexander Felfernig. 2010. Impacts of Decoy Effects on the Decision Making Ability. In *2010 IEEE 12th Conference on Commerce and Enterprise Computing*. 112–119. https://doi.org/10.1109/CEC.2010.30

[23] Erich Christian Teppan and Markus Zanker. 2015. Decision Biases in Recommender Systems. *Journal of Internet Commerce* 14, 2 (2015), 255–275. https://doi.org/10.1080/15332861.2015.1018703 arXiv:https://doi.org/10.1080/15332861.2015.1018703

[24] Amos Tversky and Daniel Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science* 185, 4157 (1974), 1124–1131. https://doi.org/10.1126/science.185.4157.1124 arXiv:https://www.science.org/doi/pdf/10.1126/science.185.4157.1124

[25] Amos Tversky and Daniel Kahneman. 1991. Loss Aversion in Riskless Choice: A Reference-Dependent Model. *Quarterly Journal of Economics* 106 (1991), 1039–1061.

[26] Amos Tversky and Daniel Kahneman. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5 (1992), 297–323.

[27] Ben Wang and Jiqun Liu. 2023. Investigating the role of in-situ user expectations in Web search. *Information Processing & Management* 60, 3 (2023), 103300.

[28] Fan Zhang, Jiaxin Mao, Yiqun Liu, Weizhi Ma, Min Zhang, and Shaoping Ma. 2020. Cascade or Recency: Constructing Better Evaluation Metrics for Session Search. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Virtual Event, China) *(SIGIR '20)*. Association for Computing Machinery, New York, NY, USA, 389–398. https://doi.org/10.1145/3397271.3401163